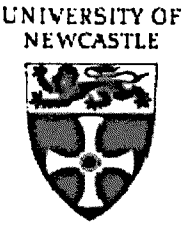


Models of Genetic and Non-Genetic Factors in Human Longevity

Fotios Drenos

Faculty of Medical Sciences

School of Clinical Medical Sciences

Gerontology

2004

Models of Genetic and Non-Genetic Factors in Human Longevity

Fotios Drenos

Thesis submitted to the University of Newcastle for the degree of PhD in the Faculty of Medical Sciences

2004

Candles

The days of our future stand in front of us
like a row of little lit candles --
golden, warm, and lively little candles.

The days past remain behind us,
a mournful line of extinguished candles;
the ones nearest are still smoking,
cold candles, melted, and bent.

I do not want to look at them; their form saddens me,
and it saddens me to recall their first light.
I look ahead at my lit candles.

I do not want to turn back, lest I see and shudder
at how fast the dark line lengthens,
at how fast the extinguished candles multiply.

Constantine P. Cavafy (1899)

Abstract

There is little doubt today that ageing is a partially inherited characteristic with the environment playing an equally important role. In this project our aim was to elucidate the gene-gene and gene-environment interactions relevant to ageing through the use of theoretical models and the evolutionary theories of ageing. With our first model we establish for the first time the plausibility of an immunogenetic trade-off between reproduction and survival under infection pressure from the environment while making detailed predictions for the expected point of balance in two different countries, one developing and the other developed, together with predictions about the surprising speed with which an evolutionary transition between the two states can occur. In our second model we develop a detailed simulation program based on epidemiological studies to account for the action of the apolipoprotein gene in western populations, its association with lifestyle parameters, and its evolution over the last 2 million years. We suggest a two-stage history for evolution of Apo E where the establishment of the ε3 allele took place during the shift of humanoids to a meat-based diet and the ε2 allele started to appear slowly as a rare mutation. Later, with the spread of agriculture and the increasing longevity of humans, the alleles began to be selected more and more towards their current frequencies. Finally, we show how a combination of socioeconomic factors and the stochasticity of mortality can be the driving forces behind the heterogeneity seen in human populations today and reveal the key factors generating this heterogeneity.

List of Contents

List of Figures, Tables, Boxes and Appendices

Acknowledgments

Although one should carry the burden uphill alone, the belief to the cause and the tenacity of heart and mind to reach the end are instilled before the journey begins. I would like to thank my family for all their support and sacrifice so that I could follow my dreams (Dimitrios, Ioanna, Foteini, Maria, Annita). I would also like to thank Georgia for always being there while I was chasing my demons. Of course I have to thank the Dr Hadween Trust for the financial support and Professor Tom Kirkwood who offered his help and friendship beyond his role as my supervisor in this project.

1. Introduction

1.1 Prologue

Ageing and the finite nature of life have always been among the greater anxieties that people have to face. But the cruel face of ageing it is not death itself but the physiological decline that it brings to a once active and strong youth. In ancient Greek mythology ageing was known as one of the gods called Geras who was the ugly son of Nyx (Night) and Erebus (Darkness of the Underworld).

Actually to define ageing is more difficult than expected since different organisms exhibit different age-related changes in a variety of basic processes. Even in a single species there is no one measurable quantity that characterizes the stage and rate of ageing. Instead we are forced to use general descriptions, which have value only when applied to populations, although they relate to properties of individuals. Ageing is often defined as the progressive loss of function accompanied by decreasing fertility and increasing mortality with advancing age (Kirkwood and Austad, 2000). This definition, although useful in general circumstances, refers to "progressive loss of function" which could mean anything from poor eyesight to poor thermoregulation. "Decreasing fertility and survival" are measurements best determined for large populations, since they exhibit a great variability in individuals and small populations.

Despite the fact that an adequate definition for ageing at the organism level remains problematic, the existing definitions are widely accepted and provide a sound basis for gerontological science.

1.1.1 Demography of ageing

Human progress has produced a profound demographic change in favour of the old[1]. As biomedical science manages to fight more and more diseases, the probability that someone will reach old age keeps increasing. In the UK, life expectancy at birth at the start of the 20th century was 50.4 and 53.9 years for males and females respectively, while at the end of the century the corresponding numbers had risen to 74.6 and 79.6 years. Mean life span in 2021 is expected to reach 78.6 years for men and 82.7 years for women (Table 1.1). Although declining child mortality played a significant role in the increase of life expectancy at birth, a corresponding increase has also affected life expectancy at older ages. For example, in the last 100 years expectation of life for men aged 80 rose by 2 years and for those aged 60 by 5 years, while it is expected to increase another 1.5 and 3 years respectively by the year 2021. These differences may not be big in absolute terms but an 80 year old man today has gained a 40% increase in life expectancy relative to 1911, whereas the corresponding increase in life expectancy at birth has been 30%. Moreover the fraction of the population reaching age 100 in 1911 was just 1 in 10,000 people, of whom 75% were woman. In 1950 the fraction increased to 2.5 in 10,000, while in 1996 it was 55 in 10,000 with 49 of them being women and 6 being men, i.e. 55 times more than in 1911 (Table 1.2, and Figure 1.1) (Source: National Statistics, Social Trends 30 2000).

[1] The term old can be used to describe both a centenarian and a 70 year old person, but in most countries the term old is used for the ages of 65+ years as the age at which one can begin to receive pension benefits (World Health Organisation).

	Year	1911	1931	1951	1971	1991	1997	2011	2021
Males	At birth	50.4	58.0	66.1	68.8	73.2	74.6	77.4	78.6
	At age 20	44.0	46.5	49.4	50.9	54.2	55.5	58.0	59.1
	At age 40	27.5	29.5	30.8	31.8	35.2	36.4	39.0	40.0
	At age 60	13.7	14.5	14.8	15.3	17.7	18.8	21.0	22.0
	At age 80	4.9	4.8	5.0	5.5	6.4	6.7	7.7	8.3
Females	At birth	53.9	62.0	70.9	75.0	78.8	79.6	81.6	82.7
	At age 20	46.4	49.4	53.6	56.7	59.6	60.3	62.0	63.1
	At age 40	29.8	32.2	34.9	37.3	40.0	40.8	42.5	43.5
	At age 60	15.3	16.4	17.8	19.8	21.9	22.6	24.1	25.1
	At age 80	5.6	5.6	5.9	6.9	8.4	8.5	9.1	9.9

Table 1.1. Expectation of life (years) at selected ages by gender, 1911 to 2021, within the United Kingdom (Amaranayake et al., 2000). Data are three-year averages centred on the year shown. For 2001 the data are from the 1998-based national population projections. The expectation of life, shown in this table, is the average number of years which a person of that age could be expected to live, if the rates of mortality at each age were those experienced in that year. The mortality rates that underlie the expectation of life figures are based, up to 1996, on total deaths occurring in each year.

	Males	Females	All
1911	0.026	0.076	0.102
1915	0.034	0.088	0.122
1920	0.032	0.086	0.118
1925	0.038	0.113	0.151
1930	0.027	0.115	0.142
1935	0.037	0.127	0.164
1940	0.029	0.147	0.176
1945	0.035	0.143	0.178
1950	0.038	0.227	0.265
1955	0.046	0.308	0.354
1960	0.072	0.459	0.531
1965	0.095	0.654	0.749
1970	0.134	0.935	1.069
1975	0.193	1.361	1.554
1980	0.257	1.884	2.141
1985	0.351	2.647	2.998
1990	0.449	3.417	3.866
1995	0.549	4.610	5.159
1996	0.580	4.943	5.523

Table 1.2. Number of centenarians per thousand of population in England and Wales, 1911 to 1996 (Amaranayake et al., 2000).

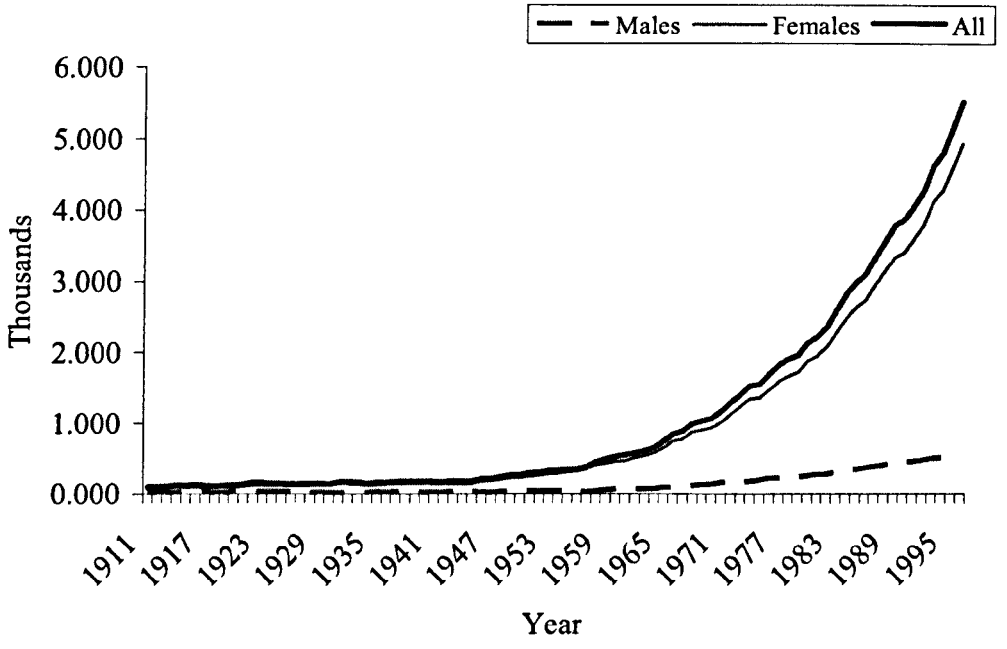


Figure 1.1. Increasing numbers (thousands) of centenarians in England and Wales between 1911 and 1995. Note the marked difference between males and females, due mainly to the greater longevity of women than men but partly also to the effect of wars (Amaranayake et al., 2000).

1.1.2 Gompertz-Makeham equation

Despite the difficulties in defining ageing, a useful quantitative measure is based upon what is known as the Gompertz Law of Mortality, named after the British actuary Benjamin Gompertz. Gompertz (1825) observed that the mortality rate in adult humans increases in a geometric progression described by the equation:

$$\mu_x = \alpha e^{\beta x} \qquad (1.1)$$

where β is the "actuarial ageing rate" expressing how fast the force of mortality increases with age. The coefficient α is a measure of the "basal vulnerability". Basal vulnerability is the force of mortality extrapolated to age 0, although actual infant mortality is higher. What is missing from the basic Gompertz equation (1.1) is the distinction between intrinsic and extrinsic mortality, the latter representing the hazards of the environment that are independent of adult age. This is overcome by adding a new term γ for the extrinsic mortality to give:

$$\mu_x = \alpha e^{\beta x} + \gamma \qquad (1.2)$$

known as the Gompertz-Makeham equation. Equation (1.2) is flexible enough to account both for populations living under secure conditions (humans and captive populations) where the death rate is based mainly on the effects of ageing and also for populations in the wild where individuals living long enough to experience senescence are rare.

Promislow et al. (1996) have described how the Gompertz-Makeham equation fails to fit changes in mortality rate in certain invertebrate species (*Caenorhabditis elegans*, *Drosophila melanogaster*, *Ceratitis capitata*). The truth is that the Gompertz-Makeham model is too simple to fully describe the extremely complex phenomenon of ageing. Despite this, it is still the best approximation, in many

species, for the observed results of senescence without the need for more detailed mathematical models with complex and limiting assumptions.

In Figure 1.2 the restrictions of the Gompertz model and its fit to the human death rates are easily observed when compared with mortality data for Britain in 1991. One of the characteristics of a geometric progression is that it can be plotted as a straight line when the values of μ are plotted on a logarithmic scale. The gradient of the line is equal to the rate of ageing β in the population. Figure 1.3 compares the observed rate with that predicted by equation 1.1. As can be seen the Gompertz model is a good description of the actual age-related changes in adult mortality but it shows a poor fit during very early deaths. In the rest of our work we will use the Gompertz-Makeham equation to account only for effects later in life. In cases where extrapolation to younger ages is required the results cannot be considered reliable (for a detailed description of patterns in early mortality see Lee, 2003).

1.2. The Evolutionary Theories of Ageing

The evolutionary theories of ageing try to explain in evolutionary terms the diversity of survival patterns between species and especially the universality, origin and shaping of senescence in higher animals (Kirkwood, 1985), in contrast to the view that ageing is just an inevitable characteristic of complex metazoans and requires no evolutionary explanation.

1.2.1 Adaptive and non-adaptive theories of ageing

The first distinction among the various evolutionary theories of ageing is between "adaptive" and "non-adaptive" theories (Kirkwood and Cremer, 1982).

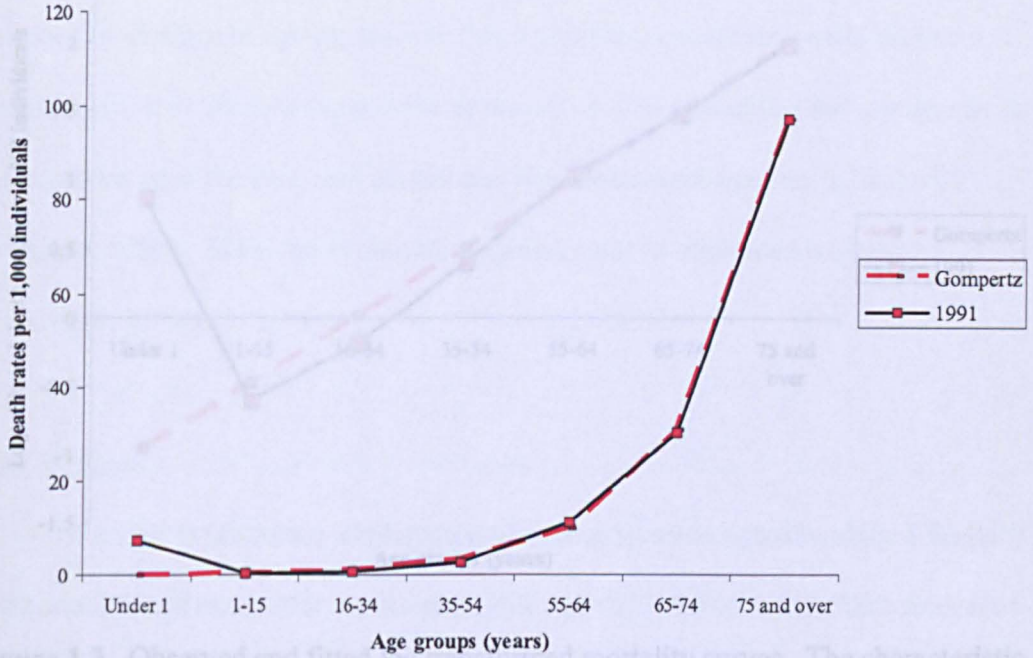


Figure 1.2. Observed and fitted curves of mortality in relation to age groups. Note how the shape of the curves for age classes older than 15 years is close to the expected exponential curve predicted by the Gompertz model (Office for National Statistics).

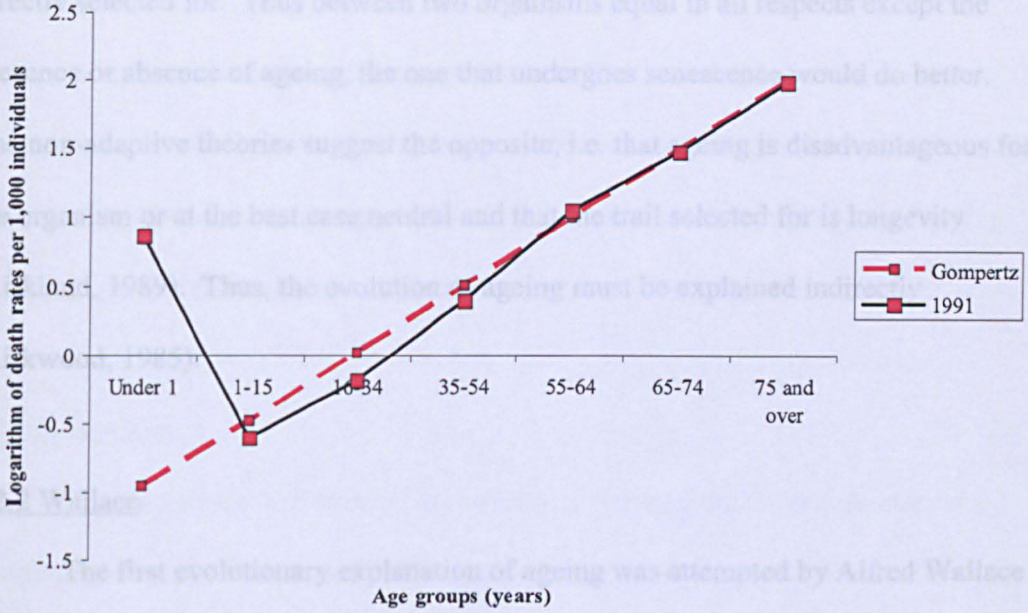


Figure 1.3. Observed and fitted log transformed mortality curves. The characteristic of all exponential curves is that when they are plotted on a logarithmic axis they are transformed into a straight line. The gradient of the line represents the "actuarial ageing rate" of the population. The fit of the curve should be considered with caution since the age groups are not uniform and may slightly disturb the slope of the observed line especially for the lowest and highest age groups (Office for National Statistics).

Adaptive theories suggest that ageing itself is advantageous to the organism and is directly selected for. Thus between two organisms equal in all respects except the presence or absence of ageing, the one that undergoes senescence would do better. The non-adaptive theories suggest the opposite, i.e. that ageing is disadvantageous for the organism or at the best case neutral and that the trait selected for is longevity (Kirkland, 1989). Thus, the evolution of ageing must be explained indirectly (Kirkwood, 1985).

1.2.2 Wallace

The first evolutionary explanation of ageing was attempted by Alfred Wallace in an unpublished note, written between 1865 and 1870, with the title "The Action of Natural Selection in Producing Old Age, Decay, and Death". This was eventually published as a footnote in the English translation of Weismann's 1881 essay "The Duration of Life" (1881).

> *"The deficiency of nourishment would lead to parts of the organism not being renewed; they would become fixed, and liable to more or less slow decomposition as dead parts within a living body. The smaller organisms would have a better chance of finding food, the larger ones less chance. That one that gave off several small portions to form each a new organism would have a better chance of leaving descendants like itself than one which divided equally or gave off a large part of itself. Hence it would happen that those which gave off very small portions would probably soon after cease to maintain their own existence while they would leave numerous offspring."* and *"...for it is evident that when one or more individuals have provided a sufficient number of successors they*

themselves, as consumers of nourishment in a constantly increasing degree, are an injury to those successors."

Wallace's theory was not developed in sufficient detail to be clear and it seems to combine adaptive and non-adaptive elements in its logic (Rose, 1991). Despite that, the first step towards an evolutionary account for the origins of ageing had been made and new and more challenging theories followed soon after.

1.2.3 Weismann

The first significant attempt to understand ageing from an evolutionary perspective was by Weismann (1881) who proposed that senescence is the result of an evolved limitation to the division potential of somatic cells. Weismann reasoned that this limitation evolved because it is beneficial to eliminate old and decrepit individuals, not because they are harmful but because they are a luxury. This first version of Weismann's hypothesis, which is the one for which he is mostly known today, is a pure adaptive theory with a programmed death mechanism. In this, old members of the population are assumed to be decrepit even in the absence of ageing leading to an obvious circular argument.

However in a later essay, Weismann went on to propose a second non-adaptive theory that considered old individuals as neutral and the evolution of ageing as a case of what he called "panmixia", where neutral characters deteriorate during evolution in a process similar to the loss of eyesight of cave-dwelling animals (for a full review of Weismann's ideas see Kirkwood and Cremer, 1982).

1.2.4 Mutation accumulation theory

For many years after Weismann, evolution of ageing was neglected until Medawar (1952) put forward the basic principle for all evolutionary theories to follow. Medawar pointed out that in an age-structured population (i.e. a population in which individuals can be segregated by age) the genetic contribution of the older members of the population to the production of new individuals is reduced with time independently of the presence of senescence. To follow Medawar's argument we have only to imagine a set of 1000 test tubes. If 10% of them are broken accidentally and in a random manner every month, we will have to buy 100 more each month to replace the breakages. If we do the same each month, the number of tubes from the original set will decrease irrespective of the existence of ageing or not. So the older the test tubes are, the smaller their number will be in the total just because they are exposed for longer to the hazard of being broken, consequently limiting their contribution to the future sets. In living organisms a similar process occurs through the random loss (by accidents, predation and disease) of individuals which affects more the old than the young members of the population even if there is no intrinsic biological ageing process, simply due to the fact that the old are exposed to danger for longer. Thus natural selection that maintains survival and fertility becomes weaker through the life history. Medawar (1952) proposed that this decline of the force of natural selection means that mutant alleles with deleterious effects will reach a higher frequency in a mutation-selection balance the later the age at which they reduce fitness (Medawar, 1952; Partridge and Barton, 1993). This concept is now known as the "mutation accumulation theory" (Kirkwood and Austad, 2000).

1.2.5 Antagonistic pleiotropy theory

Williams (1957) proposed an alternative to the mutation accumulation theory based on the same decline of the force of natural selection with age. The initial assumption is that senescence is an unfavourable character and its development is opposed by selection. To explain its presence, Williams suggest that senescence arises as the indirect outcome of selection acting on genes with different effects of fitness in different ages. This theory, now called the "pleiotropy" or "antagonistic pleiotropy theory" (Kirkwood and Austad, 2000) assumes that pleiotropic genes exist with opposite fitness effects at different ages or different somatic environments, being beneficial in earlier life but detrimental later. It must be noted that although William considers senescence as a result of adaptation his theory is a non-adaptive one since no selective advantage is recognised in ageing itself (Kirkwood, 1985).

1.2.6 Disposable soma theory

The newest and more unitary theory for the evolution of ageing is the disposable soma theory (Kirkwood, 1985; Kirkwood and Rose, 1991). In this, the organism is considered as a physical unit, transforming energy from its environment into progeny. In order to reach the point of progeny production though, it must use energy for growth, foraging, defence and repair. The theory then predicts that the energy invested in somatic survival and maintenance under the optimum allocation of energy is less than required to achieve intrinsic immortality. That optimum strategy for energy expenditure is based on the expected survival of the organism in the wild. Too high an investment in survival would be wasted when the organism dies, while too low an investment would result in premature death. So the disposable soma

theory is based on the underlying processes of somatic repair and maintenance closing the gap between mechanistic and evolutionary theories of ageing (Zwaan, 1999).

1.2.7 Evolution of ageing and longevity

Before we present the evidence for the different theories for the evolution of ageing we should briefly discuss the differences between longevity and ageing in terms of natural selection. As we saw earlier, evolutionary theory recognises that ageing originated and is maintained in life history as the result of a decrease in the force of selection with age. According to this, ageing itself, or rather the rate of ageing, cannot be selected for since it is considered as the result of "un-selection" or the by-product of selection for another character. In both mutation accumulation and optimality theories the parameter that is actually selected for is longevity. In the case of mutation accumulation, natural selection postpones the action of deleterious genes towards older ages where the representatives of the age class can have little impact in the population, thus the proportion of survivors, i.e. the longevity of the individual, is the relevant quantity and not the rate of deterioration which is the product of the changing selection force. When the optimality theories are considered, external mortality and its balance with reproduction are the two major contributing factors affecting the equilibrium of the trade-offs and the level of damage that can be tolerated at each age class, thus a shift in one of the two basic attributes will be selected through its effects on expected longevity and fecundity and not due to the amount of damage accumulated. Thus, although the term 'evolution of ageing' is often employed, it is 'evolution of longevity' that is often more appropriate except when one is specifically considering the origin of the processes of senescence. Within this thesis we shall follow customary practice and sometimes use the term 'evolution

of ageing' as a shorthand to refer to the evolution of the ageing process including longevity, although where greater precision is appropriate the term will be specified more precisely. This usage should not be taken to indicate any support for the idea that ageing, as a trait, has evolved for adaptive reasons.

1.2.8 Evidence for the evolutionary theories of ageing

Theoretical biology, as all other theoretical sciences, is based on the use of mathematics to provide explanations of empirical phenomena (for further discussion see Sober, 1993). Mathematical theories are based on assumptions of processes and they have requirements and predictions based on real processes or simple approximations of them. In order to verify a theory we need first to test its requirements and predictions logically-mathematically and then experimentally. In this section the attempts to investigate the modern evolutionary theories of ageing using model organisms will be described. The theories that are more or less widely accepted today are the mutation accumulation, the antagonistic pleiotropy and the disposable soma theory of ageing, the last two commonly being referred to as the optimality theories, since they are concerned with the optimal evolutionary strategy of the organism necessary to optimize its fitness, and they both require the existence of trade-offs.

1.2.8.1 Optimality theories

The optimality theories have in common the idea of trade-offs, in the case of antagonistic pleiotropy, between early and late-life fitness traits, while for the disposable soma theory between maintenance and reproduction. We will now review the evidence for the existence of trade-offs in the life history of organisms.

One of the first methods used to test evolutionary theories experimentally was the separation of an initial laboratory *Drosophila* population into two groups, upon which different selection forces were exerted. One group was selected for early reproduction, termed the "young" line, while the other was selected for late reproduction, the "old" line. In one of the first experiments of this type Rose and Charlesworth (1981) found a negative correlation between early and late fertility and a significant negative correlation between longevity and early fecundity in the old line. Further experimental work using the same methodology of artificial selection (Luckinbill et al., 1984) gave additional support to the trade-off between early fecundity and lifespan. Using similar principles Service et al. (1988) exerted what they described as "reverse selection" for early-life fecundity in a population of *Drosophila melanogaster* that had previously been selected for late fecundity. They used the original population as a measure of comparison for the ability of the selected flies to return to their pre-selection state. Service et al. (1988) reported that during the course of reverse selection the early fecundity of the population increased significantly while their longevity decreased to the control levels as expected by the optimality theories of ageing. Partridge et al. (1999) also used the method of 'young' and 'old' lines to study trade-offs taking care to minimize gene-environment interaction and reported an increase in survival at the same time as a decrease in early life fertility, but contrary to most previous studies there was no increase in late-life fertility for old lines. It is possible that the increase of fecundity later in life, seen by others, was either an artefact of inadvertent selection for early breeding in the young lines or due to uncontrolled larval densities (Zwaan, 1993)

Although the use of selection for late-life survival has provided valuable insight into the evolution of ageing and its mechanisms, these kinds of experiments

suffer from various problems (see Promislow et al., 1996). Zwaan (1993) used an experimental design in which selection was applied directly for longevity. His results clearly showed that the long-lived females had lower fecundity than the short-lived ones at all ages. The disagreement with previous studies which reported higher late-life fecundity for the old lines was attributed to the inability of experiments selecting on age at reproduction to distinguish between late-life fecundity selection and selection for longer lifespan. This raises the possibility of independent genetic variance between early- and late-life fecundity, or of interference from other correlated responses to selection, such as body size or developmental time. These considerations led Zwaan (1993) to propose that the observed increase in life span was due to the pleiotropic allocation of resources between maintenance and reproduction, as explained in the disposable soma theory, with lipid metabolism and starvation resistance playing an important role.

Another approach to test the evolutionary theories of ageing and the existence of trade-offs is by the use of genetic correlation between fitness components of early and late life. If trade-offs exist between fecundity and age-dependent survivorship, a negative correlation is expected between the two traits, or a positive correlation between fecundity and Gompertz parameters (rates of ageing). On this basis, Hughes (1995) tried to test the validity of the mutation accumulation and pleiotropy theories using males of many different genotypes, but constructed to be identically heterozygous for the third chromosome which represents about 40% of the entire *Drosophila* genome. No correlations between fitness characters of early and late life were seen, except a weak relationship between male mating ability and Gompertz parameters of a not significant level. This does not necessarily mean that trade-offs were not present in the life history of the male *Drosophila*. Hughes (1995) recognised

that the lack of significant negative correlations could be attributed to the lack of reproductive activity in the flies used for the survivorship assay and to the fact that only genes present on the third chromosome could be accounted for. Thus if the genes with the pleiotropic effects were on the other 60% of the genome, the method used would fail to detect their effects. In a similar experiment Tatar et al. (1996) used a very large population of flies, heterozygotic for the second chromosome, and reported a negative correlation between early fecundity (at age 3 days) and late-age mortality, while for older ages the correlation was positive. The first was explained as the result of the contribution of pleiotropy between early- and late-life fitness traits in ageing, while the second was suggested to be the result of accumulated mutations.

Partridge and co-workers produced a series of experiments concentrating on the trade-offs between survival and fertility and their underlying mechanisms. Prowse and Partridge (1997) examined the effect of reproduction on future survival and fecundity of male *Drosophila melanogaster* by creating three groups of males: never, partly and always exposed to females. By reversing their mating status during the lifespan, they showed that the reproductive costs to fecundity far outweighed the cost to survival leading to the sterility of always-exposed males by an age when 70-90% of them were still alive. It was concluded that sterility made a much more important contribution to the evolution of ageing, defined as the decrease in reproductive value, than did death. On the same theme Sgro and Partridge (1999) went on to study the increase in death rates of female *Drosophila* after mating, and demonstrated the existence of a delayed wave of death following reproduction. They used females from young and old lines which they made sterile, either by irradiation or by the use of the dominant allele ovo^{D1} which halts oogenesis in stage 4. In their conclusions, Sgro and Partridge (1999) suggested that this delayed wave of death might reflect

resource allocation between fecundity and survival as suggested by the disposable soma theory. Finally Sgro et al. (2000) studied the effect of selection for age at reproduction on female behaviour to test for correlation to selection. Their result suggested that the response of the flies with respect to age at reproduction is in part due to changes in the scheduling of female reproductive behaviour, with the young lines mating at higher frequencies early in life and old lines mating later, thus giving an underlying cause for the differences in reproduction that have been observed for the trade-off between fecundity and lifespan.

The fitness traits of *Drosophila* and the trade-offs between them have recently been studied using new methods that are much more sensitive to detect changes in variance and correlations between the traits, or even target the specific genes involved. Stearns et al. (2000) used a new method that could solve many of the problems seen in experiments of selection for specific traits. They started by introducing an environmental challenge to the population leaving the system to find the right adaptation to solve it. They did this by manipulating the external mortality, artificially imposing high adult death rates on one line and low death rates on another. The results of the experiment confirmed the prediction of the optimality theories of ageing by exhibiting that there is no trade-off between early- and late-life fecundity but rather the trade-off is between early fecundity and late survivorship. To identify the specific genes responsible for this negative correlation between fecundity and survival, a promising study has been that of Silbermann and Tatar (2000) involving the molecular chaperone *Hsp70*, a heat shock protein whose main role is to provide protection of cells from denaturing stress. Using transgenic female flies engineered to overexpress Hsp70, they concluded that induced expression of Hsp70 improves thermotolerance and age-specific survival in the long term by inhibiting

differentiation and growth. Although the number of eggs produced by the female does not decrease, the adult progeny are affected by the protein's interference with the proportion of eggs that successfully hatch. Silbermann and Tatar (2000) proposed that Hsp70 imposed reproductive costs either by direct cost upon fertility, disrupting specific cellular processes during development, or by drawing resources for its own production from the pool of substrates available to the eggs.

Although *Drosophila* has been the main organism for these studies, other organisms have also been used in selection experiments including the flour beetle, *Tribolium castaneum*, the bean weevil, *Acanthoscedelis obtectus* and the melon fly, *Bactro cucurbitae* where similar trends were found as in *Drosophila melanogaster* (for review see Zwaan, 1999).

Another extremely important organism for research into ageing is the small nematode worm *Caenorhabditis elegans*, the organism in which specific genes significantly extending lifespan were first identified (Friedman and Johnson, 1988). These long-lived mutants have proved to be equally important in the elucidation of the mechanisms and evolution of ageing as the *Drosophila* selection studies seen earlier. In this case the focus is to find if the longer lived worms have a lower reproductive rate, and what are the genes responsible for the balance between them. Tissenbaum and Ruvkun (1998) studying the way that an insulin-like signal affects survival and reproduction in *C elegans*, found that weak *daf-2* alleles and maternally rescued *age-1* alleles extending lifespan without arresting in dauer stage, also have the ability to reduce the fertility of the organism (see also part 1.3 for more on long lived *C elegans* mutants and dauer formation). The authors though, dismissed the idea that this reduced fertility was responsible for the extension of the lifespan (Tissenbaum and Ruvkun, 1998). Furthermore, evidence pointed to life shortening effects of some

modes of mating[2] in *C elegans*, similar to those observed in *Drosophila*, providing further evidence for the existence of trade-offs (Gems and Riddle, 1996). In a study of 15 temperature-sensitive and 1 non-conditional *daf-2* mutants, Gems et al (1998) found no decrease of brood size for the long living forms. Intriguingly most organisms carrying *daf-2* variants showed a significant reduction of unfertilized oocytes after the depletion of self sperm, thus pointing towards a possible trade-off between lifespan and number of oocytes produced. However puzzling results started to emerge for the causative nature of the trade-off. When the entire gonad of the nematode was removed there was no change in its lifespan, but if the germline precursor cells were removed, with a laser microbeam, while the somatic gonad was left intact, a 60% increase was observed in the nematode lifespan (Hsin and Kenyon, 1999; Riddle, 1999). This has led to the shift of our view for life-history trade-offs from a limited resource allocation mechanism to one of signalling, where a germ line originated signal represses the longevity of the organism, something that as we will explain later is not incompatible with the idea of a negative relationship between early and late life fitness (Leroi, 2001; Barnes and Partridge, 2003).

In the case of vertebrates, selection experiments are much more difficult, while mutations that significantly extend lifespan are rarer. Nevertheless, natural populations of guppies (*Poecilia reticulata*) have been used to elucidate the ecological and evolutionary aspects of senescence. By manipulating exposure to predators, different populations of guppies were established, either with high or low predation. It was found that higher mortality rates led to earlier maturity and higher investment in early reproduction (Reznick et al., 2001; Bryant and Reznick, 2004). In the case of birds the trade-off seems to operate between clutch size and adult survival

[2] C elegans can reproduce either as a hermaphrodite or by crossing with males

(Ghalambor and Martin, 2001; Nager et al., 2001) where adult mortality can explain a significant amount of clutch size variation even when nest predation is controlled (Martin, 2002). Mice, the most common mammalian model, have also been used extensively in biogerontological research. Usually, experiments with mice are focused on the effect of dietary restriction on longevity, however in the last few years a number of mutations that extend lifespan have been found. Although these mutations are still not fully characterized, and the precise ways in which lifespan is affected are largely unknown, it has been found that at least for one of them, the dwarf mouse mutation, reproduction is greatly affected too, with most of the mice carrying the mutation being infertile (Liang et al., 2003). Again, in the case of mice, as in *C elegans*, we should be cautious not to be misled into believing that the link between reproduction and longevity necessary signifies a causative relationship between the two.

Lately there has been interest in the existence of similar trade-offs in humans, a highly controversial issue. These studies are mainly based on historical populations for which records for dates of birth, death and number of offspring exist. In the forefront of the debate is the work of Westendorp and Kirkwood (1998) which examined the relationship between longevity and reproduction using a historical data set from the British aristocracy. From the data it could be seen that when the study was restricted to women that reached menopause, female longevity was negatively correlated with number of children and positively correlated with age at first childbirth, confirming the existence of trade-offs. Westendorp and Kirkwood's conclusion was criticised by Gavrilov and Gavrilova (1999) for failing to adjust the data for age at marriage, for overlooking confounding factors, such as the husband's fertility, and for using incomplete data. However in a reply to Gavrilov and

Gavrilova, Westendorp and Kirkwood (1999a) showed that neither of the first two concerns invalidated the finding, while they argued that although the data were inevitably incomplete, there was no reason why the missing data should have biased the comparison between women who dies at different ages in such a way as to produce an artifactual association. In a subsequent study on the genealogy of the British peerage which went some way towards resolving the problem of incompleteness in the dataset, Doblhammer and Oeppen (2003) confirmed a strong correlation between parity and late-life mortality, even after correction for differences in health and mortality before the age of 50. Lycett et al. (2000) examined the same relationship using records for a historical population that inhabited the Krummhorn region of north-west Germany between 1720 and 1870. The general data obtained from this analysis seemed not to support the existence of trade-offs, but when social status was taken into account it was seen that for the poorest social group there was a negative correlation between fecundity and survival. Similarly, Korpelainen (2000) found that an increase in longevity, at least in women, was linked with a decrease in the number of children, but also, with a small increase in the number of surviving children. Finally, Thomas et al (2000), using data from 153 countries, also reported a negative trend between female life span and fecundity, even after statistical correction for confounding factors. On the other hand, there are historical population data that do not support the existence of such a trade-offs. These are both older (Henry 1956, Gautier and Henry 1958, reviewed in Doblhammer and Oeppen, 2003) and more recent studies (Le Bourg et al., 1993; Wrigley et al., 1997), but more interestingly direct long term observation of young women (20-30 years old) with bilateral salpingo-oophorectomy and no post operative hormonal replacement showed no influence in longevity (Nilsson et al., 2003b). The later can be attributed to an earlier

set of the trade-off in such a way that the actual investment in reproduction has already taken place in the level of precursor cells during early development, as seen earlier in the case of *C. elegans*, and the completion or not of reproduction during adulthood will have only a small effect under the specific environmental parameters.

Despite the importance of studies on historical populations for our understanding of human ageing, they must be approached with great caution as their validity is based on many different factors, such as the detail of the kept records, duration of marriage, specific causes of death etc. Furthermore, the nature of the trade-off in humans, and in other animals, can have very complex dynamics, where the decrease in reproduction with increasing life span can be attributed to behavioural, economic or social reasons instead of physiological mechanisms (Mulder, 2000; Gavrilov and Gavrilova, 2002).

1.2.8.2 Mutation accumulation

Specific evolutionary predictions, that can be tested experimentally, can be made by mathematical models based on the mutation accumulation theory. Common simplifying assumptions are that mutations have an additive effect on age-specific survival and that they have the same probability to act at any age. These predictions are: a) the variance of fitness traits should increase with age; b) reverse selection experiments for early fitness, on lines produced from selection for late-life fitness, should only return to initial state slowly; c) the controlled introduction of spontaneous or directed mutations should accelerate the rate of senescence; d) inbreeding depression should increase with age (Promislow and Tatar, 1998).

The change in fitness variance with age has been the most tested prediction of the mutation accumulation theory and has provided the most contradictory results.

According to the theory, the equilibrium frequencies of mutant alleles will increase with age due to the inability of natural selection to remove them from the population. Thus a decline in mean performance and an increase in the genetic variance of the population are expected with age. The requirement for an increase of genetic variance is fundamental, although it is not exclusive to the mutation accumulation theory. To test for an increase of genetic variance with increasing age, genetic variation for fecundity, mortality and male mating ability have been used, mainly in populations of *Drosophila melanogaster*. The first among many experiments to test the prediction for increased variance was that of Rose and Charlesworth (1981) which found no change in additive genetic variance with age. Engstrom et al. (1989) found an increased variance for fecundity, although questions have been raised about the mathematical manipulation of the results (Promislow and Tatar, 1998). Tatar et al. (1996) estimated the genetic variance for age-specific fecundity and reported a bimodal pattern, with peaks of the variance at 3 and 17-31 days, which they explained as the effect of mutation accumulation for the latter and the probable effect of selection maintaining variation in larval developmental rate for the former. Their analysis was therefore equivocal suggesting mutation accumulation as a major source of standing genetic variance for senescence without excluding the role of antagonistic pleiotropy. The genetic variance for male life history has been tested by Hughes (1995) who assayed male flies for age-specific competitive mating ability, age-specific survivorship, body mass and fertility and obtained results in accordance with the mutation accumulation theory.

Mortality is probably the most important feature of the life cycle in ageing research, and its genetic variation was first investigated by Hughes and Charlesworth (1994), who reasoned that if genes with deleterious effects are confined to older ages,

as the mutation accumulation theory states, then their frequency will increase with age and their mean value of fitness will decrease, increasing the genetic variability in the population. The results showed that genetic variability of mortality in the male *D. melanogaster* increased in later ages exhibiting substantial genetic and environmental variability. Promislow and Tatar (1996), using much larger sample sizes, observed an increase in the variance components for mortality but only during the first half of life, after which there was a marked decline. This unexpected decline in variance may have been due to a variety of reasons, such as temporal changes in laboratory conditions, variation in reproduction, demographic heterogeneity, insufficient sample size in very old age classes and finally, equally poor gene regulation in all genotypes. Promislow and Tatar (1996) rejected neither mutation accumulation nor antagonistic pleiotropy, but proposed that the assumptions made for the effects of the age specificity of rates and effects of polygenic mutations should be updated to account for the deceleration of mortality rates in older ages and heterogeneity. Pletcher et al. (1998) also found evidence for the increase of variance in early ages and its decline afterwards, suggesting that the absence of significant effects on mutation at older ages might have been responsible for this decline.

The slow return to pre-selection phenotypes, in *Drosophila* experiments designed to test the optimality theory, can distinguish between mutation accumulation and antagonistic pleiotropy. In the first case the return is expected to be slow, since it relies on mutation-selection balance and drift, while for the second it will be rapid, since there is a positive selection force acting upon the trade-off between longevity and early-life fitness. Service et al. (1988) applied reverse selection for early-life fitness to a laboratory population of *Drosophila melanogaster* that had been previously selected for postponed senescence. The characteristics tested were early

fecundity resistance to stress of starvation, desiccation and ethanol. The results showed that early fecundity increased to control population levels while starvation resistance declined in accordance with a negative genetic correlation as predicted by the antagonistic pleiotropy theory. However desiccation and ethanol resistance did not decline, suggesting a lack of trade-offs with early-life fitness components, providing possible evidence for the operation of mutation accumulation. Promislow and Tatar (1998) considered these results of Service et al. (1988) and found both methodological and theoretical problems. They concluded that the reverse selection response might be due to epistasis, combined with the different genetic background of the long-lived lines.

Instead of testing the role of mutation accumulation in the evolution of senescence, the effect of controlled mutations on age-specific mortality can be directly assessed. In order to estimate the effect of mutations in ageing, Houle et al. (1994) created a set of mutation accumulation lines[3] of *Drosophila* with the use of the SM1 balancer chromosome and analysed 48 lines for the following traits: early and late fecundity, male and female longevity, male mating ability, early and late, productivity and body weight. The results showed that mutational effects were positively correlated among the young and old classes, although one of the requirements of the mutation accumulation theory is that mutations must have independent effects on early and late-life performance, otherwise changing the age where selection acts more strongly will have no effect on early-life fitness traits. Houle et al. (1994) concluded that the mutation accumulation theory is not enough to

[3] A mutation accumulation line is a *Drosophila* line that can accumulate mutations in the absence of selection against them. This has the effect that new mutations are fixed in the populations at almost the same rate as they appear.

explain the evolution of senescence, at least in *D. melanogaster*; and that mutation-selection balance is sufficient to explain the additive genetic variance for fecundity and longevity.

Finally, testing inbreeding depression can distinguish between mutation accumulation and antagonistic pleiotropy effects. Under mutation accumulation, if deleterious mutations have a higher frequency in older than in younger ages and if these mutations are recessive, as most mutations are, then inbreeding depression would be less in early than in later life. Charlesworth and Hughes (1996) developed a model to predict the age-related patterns of inbreeding depression and genetic variance. The theoretical predictions were later compared to experimental data of virgin male *D. melanogaster* and they found that in accordance with the mutation accumulation theory, the inbreeding load increased with age. Further work by Hughes and her team (Hughes et al., 2002) using 10 isogenic lines and all the possible crossings between them, increasing the ability to detect dominance variance, found a dramatic increase of inbreeding effects with age, confirming their previous results.

1.2.8.4 Conclusions

As can be seen, the results obtained in different studies have not yet reached a definite and safe conclusion about the role of mutation accumulation theory and its complex relation to the optimality theories. The contradictory results obtained, even from very similar experiments, cast doubts over the validity of some of the approaches used. The truth is that selection experiments using laboratory-adapted stocks may give rise to false conclusions, both for mutation accumulation and antagonistic pleiotropy theories (Promislow and Tatar, 1998). Moreover the theoretical predictions of the model for the mutation accumulation theory assume that

fitness traits are normally distributed, that mutations affect only one age class, have an additive effect on survival and they all have equal effects, but none of the above is biologically realistic as can be seen from various studies (for more details see Promislow and Tatar, 1998).

As for the optimality theories, from the experiments discussed, it is obvious that trade-offs between fecundity, especially early in life, and lifespan do exist. Despite that, the controversy continues about what kind of trade-offs there are, what are the mechanisms or the genes responsible, whether they are intrinsic or extrinsic, in which organisms they exist, and how important they are in the evolution of ageing. These are questions that are very difficult to answer based on our current knowledge and further investigation is needed both in model organisms and in humans with new methodologies that will overcome the limitations of the approaches used up to now. We can summarize our current ideas by saying that a number of different mechanisms of trade-offs in different levels of organization seem to exist. As we saw in part 1.2.7.1, the classical view of life history theory for a direct resource allocation trade-off has been enriched with behavioural, developmental, signalling, and immunogenetic mechanisms affecting the negative relationship between reproduction and survival. The trade-off of reproduction and survival is most likely controlled both by genetic mechanisms that carry the evolutionary history of the genus and through the plasticity of behaviour and early development responding to current experience in such a way that maximum fitness under the given conditions can be achieved.

Probably, as many studies have shown (Snoke and Promislow, 2003), neither of the two theories can account for all the complex changes taking place during ageing. A more holistic view for the evolution of senescence is required that will be able also to account for recent findings such as deceleration of the mortality rate in the

very old, the decrease of genetic variance in the later part of life and the convergence and cross-over of mortality curves among cohorts in older ages.

1.3 Genetics of Ageing

So far we have considered general aspects of ageing and the theories put forward to explain its presence and evolutionary maintenance. Now we are going to consider the role of genes involved in ageing and longevity. As we saw earlier, senescence is detrimental to the organism, so we do not expect to see genes promoting ageing but rather genes evolved to confer longevity.

The most straightforward way to search for longevity genes is generally held to involve the use of model organisms. Since the complexity of the interactions of genes within the genome and with the external environment increases with increasing level of organisation, the existence of lower organisms with a smaller number of genes and a clearer set of gene-environment interaction has proved a valuable tool in our efforts to elucidate the genetic mechanisms involved in ageing and longevity. However it is also apparent that a clear theoretical framework is essential to generate predictions that can make such studies much more efficient. We now consider the major model organisms used for studies in the genetics of ageing

1.3.1 *Saccharomyces cerevisiae*

The first organism to be considered will be *Saccharomyces cerevisiae*, better known as baker's yeast. *S. cerevisiae* is a unicellular organism with its lifespan measured usually by the number of times it divides, i.e. by the number of daughter cells produced (Jazwinski, 1999a). Various genes have been implicated in the ageing

process in yeast; the first to be studied were the *RAS* genes involved in metabolic control and stress resistance. In particular *RAS2*, which encodes a homologue of the mammalian signal transduction protein *c-H-ras*, operates as a part of a nutrient sensor having a central role in cell growth and division through several signal transduction pathways. These lifespan extending effects of the gene are exerted through at least two different pathways of metabolic and proliferation control; in one by the stimulation of adenylate cyclase, while in the other through the Mitogen Activated Protein (MAP) kinase pathway, both of them also being involved in cell growth and differentiation of multicellular organisms (Jazwinski, 1998). In addition, *RAS2* plays an important role in stress responses through the *Hog1* kinase, which is a member of the MAP pathway. *Hog1* is a homologue of p38 kinase (also belonging to the MAP pathway) responsible for stress response in mammals (Jazwinski, 1999a). *RAS2* may operate in different ways to affect the ageing process, either as an adaptation gene compensating for a decline in homeostasis with age, or as a regulator of maintenance processes preserving homeostasis (Jazwinski, 1999c).

Other important genes for postponed senescence in yeast are the *CDC7*, *CDC25*, *SIR4* and *SIR2*, *SGS1*, *PHB1/2* and the retrograde regulation genes (*RTG1*, *RTG2*, *RTG3*). *CDC7* is a protein kinase, regulating cell cycle control and transcriptional silencing. *CDC25* is involved in the glucose response signalling pathway, its mutants (*cdc25-10*) mimicing calorie restriction and having an increased lifespan. The *SIR* complex is a transcription silencer of telomeric genes and of genes required for sexual mating (*HML* and *HMR*) (Finch and Tanzi, 1997). The *SGS1* gene is a DNA helicase with a role in DNA recombination. An interesting feature of *SGS1* is its homology to the human gene for Werner's syndrome, showing a 24% identity and a 34% similarity (*Saccharomyces* Genome Database). The retrograde regulation

genes were first identified in studies with "petite" yeast (cells lacking fully functioning mitochondria) and they are responsible for an intracellular signalling pathway between nucleus and mitochondria, the retrograde pathway, which leads to the induction of expression of nuclear genes (Jazwinski, 1999b). Other genes found in "petite" yeast cells are the *PHB1* and *PHB2* coding for mitochondrial proteins, homologous to the human prohibitin gene, whose function is to interact with *RAS2* and adjust the mitochondrial activity (Jazwinski, 1999b).

Lately, even more genes affecting yeast lifespan have been identified, like *Rpd3*, a histone deacetylase required for proper timing of replication origin firing and repression of transcription of other genes; and the *HXT* and *HXK* genes and the *Snf1* kinase complex involved in glucose sensing and calorie restriction (for a recent review of the genetics of yeast ageing see Bitterman et al., 2003). An emerging key factor in yeast ageing is the *SIR2* gene, which suppresses rDNA recombination. The rDNA locus, comprising 10% of the yeast genome, consists of 100-200 tandemly arrayed 9kb repeats coding for the rRNA. Only about half of rDNA repeats are transcriptionally active at any time. Homologous recombination between adjacent repeats leads to the formation of extra-chromosomal circular forms of rDNA known as ERCs, which probably titrate the replication and transcription machinery of cells and result in its death (Sinclair and Guarente, 1997). Since the *SIR2* gene is responsible for suppressing this recombination, its importance in the longevity of the yeast is evident.

1.3.2 *Caenorhabditis elegans*

The nematode worm *Caenorhabditis elegans,* a member of the Rhabditidae, is the first organism in which "ageing" genes were identified. This roundworm, of

around 1mm in length, is a mainly self-fertilizing hermaphrodite and its life cycle takes about 15-20 days at 20^0 C, making it ideal for ageing research. The first gene to be identified was *age-1*, also known as *daf-23*, a member of the *daf* family of genes (Morris et al., 1996). These are involved in entry into the dauer larval stage, a long-lived stress-resistant dispersal form that the organism uses under adverse conditions. Known *daf* genes with effects on longevity are the *daf-2*, *age-1* and *daf-16*, the first two being dauer-constitutive (enter dauer stage even without the otherwise essential pheromone signal) while the third is dauer-defective (not able to enter dauer stage). Cloning of *daf-2* and *age-1* suggested that they participate in a signal transduction cascade similar to what is found in vertebrates (Hekimi et al., 1998). *age-1* encodes a PI3-kinase while *daf-2* encodes an insulin receptor transmembrane tyrosine kinase. All three of the genes discussed are involved in the insulin/insulin-like growth factor 1 (IGF-1) endocrine system, which seems to be evolutionarily conserved from the nematode to humans. Two tissues, the sensory system and the reproductive system are believed to regulate lifespan by influencing the components of the DAF-2 insulin/IGF-1 pathway (Guarente and Kenyon, 2000). The DAF-2 pathway, in addition to ageing, affects reproduction, lipid metabolism and dauer formation independently of one another. Its main effect in ageing is exerted through *daf-16*, a transcription factor that regulates a variety of genes including those involved in stress response, antimicrobial defence and metabolism, while at the same time repressing specific life-shortening ones (Murphy et al., 2003).

Other genetic mechanisms extending the lifespan of *C elegans* are also considered. Mutants of the clock gene (*clk*) also show extended life span and exhibit slow development, lengthened cell cycle, and changes in behaviour such as slower oesophageal pumping and swimming rates. *clk-1* encodes a protein which is highly

conserved in all eukaryotes, its yeast homolog COQ7 being a nuclear-encoded regulator of mitochondrial function, which indirectly regulates the transcription of genes responsible for growth on non-fermentable carbon sources (Finch and Tanzi, 1997; Hekimi et al., 1998; Jazwinski, 1999b). It is believed that the longevity effect of *clk-1* is, at least partially, achieved by the redaction of extramitochondrially produced ROS (reactive oxygen species). Yeast, worms and mice seem to accumulate demethoxyubiquinone instead of ubiquinone in the transport chain in the absence of clk-1, thus reducing the generation of ROS during respiration (Hekimi and Guarente, 2003).

1.3.3 *Drosophila melanogaster*

A more complex model organism than those considered up to now, commonly used for genetic studies, is the fruit fly *Drosophila melanogaster*. The fruit fly has proved to be a very useful and powerful model in ageing research, its role spanning from the development of early ageing hypotheses, such as the rate of living hypothesis, to recent evolutionary experiments and from demographic studies of patterns of mortality to the plasticity of life span (reviewed in Helfand and Rogina, 2003). In addition to its significance in demographic studies, *Drosophila* has also been used to elucidate the genetic and physiological mechanisms of longevity. These studies have shown an inverse relationship between enhanced stress resistance and life span. Different lines exhibit resistance to different stresses but the most common are starvation, desiccation, heat, ethanol and most importantly oxidative stress associated with increased levels of antioxidant enzymes (Jazwinski, 1998). The oxidative stress hypothesis states that oxygen radicals generated during metabolism can damage proteins, lipids and DNA in the cells of the organism; mutations in genes

that reduce the activity of anti-oxidant enzymes such as superoxide dismutase (SOD), Mn superoxide dismutase (SOD2), catalase and thioredoxin reductase (TxrR) also reduce the lifespan of the fruit fly. On the other hand, experiments trying to increase the activity of the genes responsible for these enzymes throughout life (Orr and Sohal, 1994) were unable to give consistent results, probably due to the significance of early development and maternal effects in *Drosophila* (Helfand and Rogina, 2003). Studies focused only in adults have managed to see a consistent life span increase of 10-20% (Sun and Tower, 1999). Recently it was shown that the effects of SOD on longevity are genotype- and sex-specific, with gene x genotype epistasis interactions playing an important role (Spencer et al., 2003). In the case of starvation resistance, another important determinant for *Drosophila* longevity under stress, phosphoglucomutase, an enzyme controlling the utilization of glucose, may play an important role, since it is part of an elaborate mechanism affecting the total stored calories in the fly (Rose, 1999; Montooth et al., 2003).

A promising case of a longevity mutation is the P-element insertion mutation of the *methuselah* gene conferring approximately 35% increase in average life-span and enhanced resistance to stresses such as starvation, high temperature, and free-radicals. The *methuselah* gene codes for a protein of the G-protein receptor family, a group of receptors involved in a wide range of biological activities including neurotransmission, hormone physiology, drug response and transduction of external stimuli (Lin et al., 1998). Specifically the *mth* gene may play a role in maintaining homeostasis and metabolism modulating cellular functions in response to stress. Recently, a second longevity gene was found in Drosophila named *Indy* (for *I'm not dead yet*) resulting in a doubling of average life-span and a 50% increase in maximal life-span (Rogina et al., 2000). Although specific details for the function of *Indy* are

not known, it is closely related to a mammalian sodium dicarboxylate co-transporter, a membrane protein that transports Krebs cycle intermediates, and it is expressed in the fat body, gut and oenocytes suggesting that *Indy* probably has a role in the absorption of metabolites and in intermediary metabolism.

Finally, of great interest and importance is the effect of insulin/IGF signalling in *Drosophila* ageing. Earlier we saw how *daf-2* and *daf-16* can more than double the life span of *C elegans*. Screening of a number of mutations in each of the known components of the Insulin/IGF signalling pathway in the fruit fly, led to the discovery of three genes extending the *Drosophila* lifespan, the *InR* and *Dp110*, equivalents of the nematodes *daf-2* and *age-1* respectively, and *chico*, an insulin receptor substrate with no nematode equivalent (Gems and Partridge, 2001). *InR* affects neurosecrotory tissue responsible for juvenile hormone. Homozygous *InR* mutants are lethal while heteroallelic flies are usually viable, dwarf and infertile adults, with females having ovaries similar to the diapaused wild-type. From all the different *InR* mutants, females of InR^{p5545}/InR^{E19} have a reduced age-specific mortality and an 85% extension in longevity. The fact that not all *InR* alleles have the ability to increase lifespan has been attributed to complications in early development that continue later in life (Tatar et al., 2001). In contrast to *InR*, *chico* homozygous female mutants were viable, and exhibited an increase of 48% median and 41% maximum life span, while homozygous males were slightly short lived. The null allele $chico^{1}$ heterozygotes, showed a 36% and a 13% increase in median life span for females and males respectively. Additionally, heterozygote females are subfertile while the homozygotes are almost sterile. In this case, as in *InR*, the gene leads to a reduced body size but in a recessive fashion, in contrast to its life extending effects that are semidominant and independently regulated from size (Clancy et al., 2001). The

existence of similar, evolutionarily conserved mechanisms between *Drosophila* and *C elegans* for the extension of life is of great importance in the biology of ageing and our understanding of its mechanisms.

1.3.4 Mammals

Among mammals, the organism most frequently used in laboratories for genetic research is the mouse. Genetic studies in mice concentrated mainly on the association of the immune system, and especially the Major Histocompatibility Complex (MHC) to ageing. The results show a positive correlation between immune responsiveness and longevity (Jazwinski, 1996). Certainly, by far the most studied aspect of mice longevity is the effect of caloric restriction. Caloric restriction usually refers to diets providing an energy intake 30%-40% less than that obtained by animals fed *ad libitum*, or to the maintenance of the animals at a reduced stable weight. A 40% reduction of calories, or a 50% weight compared to freely fed mice shows a 20-40% increase in longevity[4]. Caloric restricted mice are smaller, with smaller major organs and less fat, while their fertility decreases or stops completely, with puberty and menopause occurring later. The effects of restricted energy intake are dependent on sex, stage of development, duration of the treatment and level of restriction (reviewed in Shanley and Kirkwood, 2000). The precise changes caused by caloric restriction in the organism, or their specific mechanisms, are still debated but generally there is an up-regulation of somatic maintenance and repair. One hypothesis for the underlying mechanisms is that by decreasing the energy intake, the oxidative damage is also decreased, due to lower energy flux and metabolism, thus slowing ageing. Unfortunately despite the great number of experiments conducted,

[4] Similar extension of lifespan due to caloric restriction has been reported in rodents, yeast, worms and lately in primates.

the results obtained are largely unreliable, especially for the *in vivo* confirmation of the idea (Merry, 2004). If indeed caloric restriction can affect the production of ROS, then similar genetic pathways to those discussed for the other model organisms and their resistance to stress should be considered in mammals. The only mutant known to have increased resistance to oxidative stress and prolonged lifespan is the $p66^{Shc-/-}$ mouse. The $p66^{Shc}$ is a splice variant of $p52^{Shc}/p46^{Sch}$ situated on the mammalian *Shc* locus and it is a cytoplasmic signal transducer, transmitting mitogenic signals from activated receptors to *Ras* (a homologue of *RAS2*, seen earlier in yeast). Mutation of the $p66^{Shc}$ increases oxidative stress resistance in the intracellular and organism level, while its normal function is to be activated by ROS, and through p53, to induce apoptosis (Migliaccio et al., 1999; Napoli et al., 2003).

In contrast with the organisms considered up to now, where a single insulin/IGF receptor exists, mammals have several homologous receptors including the insulin receptor and the IGF-I receptor (Gems and Partridge, 2001). Most of the long lived mutant mice show disruptions in the regulation of their hormone balance. Hereditary dwarfism in mice has led to a number of different models for extended life span. The Ames dwarf mice are small and deficient in growth hormone (GH), prolactin and thyroid hormone. Although they are born with a normal size, their postnatal growth is retarded, reaching only one third of the size of their wild type siblings. It has been found that they can outlive the wild type mouse by around 350 days for the males and more than 470 days for the females. (Brown-Borg et al., 1996). A point mutation in the *Prop-1* gene which sits upstream of the pituitary specific transcription factor-1 (*Pit-1*) gene and is required for its activation, is responsible for the effect (Liang et al., 2003). Very similar are the Snell dwarf mice, with the point mutation directly in the *Pit-1* gene preventing the DNA binding of its product, thus

leading again to pituitary hormone deficiency and a 50% reduced adult size. The effects of the homozygous mutation on longevity are highly controversial but recent studies showed a 40-50% extension for females, while in the case of males husbandry conditions could lead to either a reduction (when housed with normal males) or increase (when housed with normal females) of life span (Flurkey et al., 2002). Since both Ames and Snell dwarf mice have a defect in exactly the same pathway, they share a lot of their characteristics, including reduced metabolism, body core temperature, infertility and most importantly very low levels of plasma insulin, IGF-1 and glucose. Growth hormone stimulates IGF-1 production, leading us to believe that the life extending effect of the mutations are due to a reduction of the insulin/IGF-1 signalling pathway in a similar way to that seen in *C elegans* and *Drosophila* (Gems and Partridge, 2001). Further evidence for that can be found in the "little mouse" and GHR/BP knockout mice model. Having a missense mutation in the *Ghrhr* locus the "little mouse" is defective in the response to the hypothalamic peptide GH releasing hormone, having 1% of the normal GH level and two thirds of the normal adult size. The homozygous genotype increases the life span of the mice by around 25% ,for both sexes, compared with the normal heterozygous for the $Ghrhr^{lit}$ gene (Flurkey et al., 2001; Liang et al., 2003). *GHR/BP* gene encodes two proteins, the growth hormone receptor (GHR), a membrane-bound receptor with a major role in the signal transaction process, and the growth hormone binding protein (GHBP) that probably modulates the level of circulating GH. The specific locus is also responsible for the Laron syndrome in humans, a form of hereditary dwarfism. GHR/BP knockout mice are also dwarfs with a weight of 40% relative to the wild type and increased fat accumulation, while they exhibit an increase in longevity of around 50%. The low IGF-1 level confirms the idea that the lack of GH signalling reduction is responsible

for alteration of the insulin/IGF-1 pathways, thus decreasing the ageing rate of the organism (Coschigano et al., 2000; Coschigano et al., 2003; Liang et al., 2003).

Studies to examine the association between IGF-1 and survival in mammals directly, found that mice null for the insulin-like growth factor-1 receptor (IGF-1R) were dying in birth, probably due to respiration problems. On the other hand, the heterozygous knockout mice were viable, without developing dwarfism, and with normal metabolism, nutrient uptake, physical activity, fertility and reproduction. The level of IGF-1 was halved, as was the IGF-1R signalling response, while importantly there was a 26% increase in longevity (Holzenberger et al., 2003; Liang et al., 2003). Finally, mice with a fat-specific insulin receptor knockout (FIRKO) also had an increase in mean life span of 18%. These animals had normal levels of the insulin receptor in all tissues except adipose tissue, where it was only 10-15% of the wild type. When food intake was measured as per gram body weight, the FIRCO mice seemed to consume more food despite the fact that they had a 50-70% decrease in body fat. Interestingly these knockout mice were able to maintain a normal glucose tolerance at 10 months of age in contrast to the wild type (Bluher et al., 2003; Liang et al., 2003). The latter leads to the interesting idea that caloric restriction and IGF-1 signalling in mammals might work in similar or even linked ways to extend the life span of the organism. Work to validate this idea has proved that if a relationship exists it is certainly not a simple one and further work is needed (reviewed in Gems and Partridge, 2001).

1.3.5 Other model organisms

Other organisms are also used as models for research into ageing. One of the emerging models is the zebrafish *Danio rerio* which combines vertebrate biology with

ease of large-scale mutational experiments. Although the zebrafish has been extensively used in other branches of biology, especially developmental biology, its potential significance for gerontology has only recently become evident (see Gerhard, 2003). Another underrepresented class of animals in longevity studies is the Aves. Birds are long-lived homeotherms living up to three times more than mammals of equivalent size (Holmes and Ottinger, 2003). Their ability to live much more than mammals, while having a twofold metabolic rate compared to them, suggests that birds have specific adaptations against free-radical damage, a key parameter for cellular senescence. Moreover, various bird species have been used as models for a variety of ageing associated processes such as reproduction, neurodegeneration and respiration (reviewed in Holmes and Ottinger, 2003). Comparable to caloric restriction in other model organisms, many studies have tested the effects of dietary intake on breeding and overall health in birds (Walsh and Brake, 1999). The most popular avian laboratory models that can find further use into ageing research are the Japanese quail (*Coturnix japonica*), the Budgerigar (*Melopsittacus undulatus*) and the canary (*Serinus canaria*) which are relatively short-lived species although wild species could also be used,as the white-crowned sparrow (*Zonotrichia leucophrys*), a model that has been extensively studied in the fields of neurobiology, nutrition and physiology (reviewed in Holmes and Ottinger, 2003). Nonhuman primate models have been used for a long time in biogerontological studies, especially to elucidate the physiological changes with ageing in various organs, as the brain and the heart, and to study age related diseases such as cardiovascular disease and diabetes. Caloric restriction studies suggest that similar mechanisms operate in primates as in other mammals with late maturity, increased insulin sensitivity and lower triglyceride levels and cholesterol (Lane, 2000). Genetic studies though, are very rare, or non-existent,

in nonhuman primates, mainly due to the extremely complex relationships in gene-gene and gene-environment interactions and the lack of any known long-living mutations.

1.3.6 Humans

We have discussed the use of different model organisms to elucidate the genetic mechanisms behind ageing and longevity. Although these are very useful for the description of basic mechanisms of senescence in the molecular level their relevance to human ageing should be approached with great care. Despite the obvious common ancestry of humans and model organisms, the precise evolutionary pressure experienced by different phylogenetic groups and developmental schedules in the history of life might have significantly altered the importance and precise associations of pathways, even if initially they seem as evolutionary conserved in terms of basic function. Moreover, the main focus of our anthropocentric science is to improve human life expectancy and health. Ethical and practical problems make direct genetic studies and interventions in humans impossible. Nevertheless alternative ways have been found, mainly through the use of association studies and linkage analysis. Association studies, which are the most popular way to check for longevity genes, test for non-random associations of candidate genes with longevity in unrelated individuals (De Benedictis et al., 2001). Linkage analysis is designed to test for non-random segregation at a marker locus in a pedigree and is much more difficult due to its requirements for two or more very old individuals in more than one generation (Lio et al., 2003). Earlier, we saw the importance of the immune system in the longevity of mice, but immune responses have also been studied in humans, particularly the effects of the *HLA-DR* (human leukocyte antigen) locus on longevity.

HLA-DR is associated with autoimmune diseases, diabetes and probably apoptosis. In a case-control study of human centenarians three alleles, *DR7*, *DR11* and *DR13*, showed differences in frequency between centenarians and controls (Ivanova et al., 1998). *DR7* is considered a risk factor for CMV and HBV infections, celiac disease, and is also associated with Crohn's disease, a disease of the digestive track, and B-cell lympocytosis. A more recent study in a Sardinian population, though, has shown that the effect of HLA on longevity might be population specific (Lio et al., 2003), something which might have been expected, considering that even in lower organisms genetic background plays a significant role in the genetics of longevity. Later we will return to the effect of the immune reaction on longevity with a model of trade-offs between reproduction and survival.

The best known example of a longevity gene is *Apo E* one which codes for apolipoprotein E, a major part of lipid metabolism. *ApoE* and *ACE* (Angiotensin converting enzyme) loci were investigated due to their association with cardiovascular disease, atherosclerosis, LDL-cholesterol and Alzheimer's disease. Two alleles of apolipoprotein E and a variant of angiotensin converting enzyme showed a significant difference in frequency between centenarians and a younger control population (Schachter et al., 1994). The two *ApoE* alleles were the *ε4* which promotes premature atherosclerosis and is a risk factor for ischaemic heart and Alzheimer's disease (*ApoE* also plays a role in nerve development and repair), showing a decreased frequency in centenarians, and the *ε2* which is protective against high levels of cholesterol but also has an hypertriglyceridaemic effect and shows a higher frequency in the centenarian population. *ACE*, coding for a dipeptidylcarboxy-peptidase of the rennin-angiotensin system that regulates blood pressure and the production and degradation of angiotensin II and bradykinin respectively to affect cardiac growth, shows a deletion

(D) - insertion (I) polymorphism responsible for a significant effect of serum ACE activity (Schunkert et al., 1994; Ramaraj et al., 1998). The increased frequency of *ACE/DD* found among centenarians during the Schachter et al (1994) study is surprising since the specific allele is linked with myocardial infarction, ventricular hypertrophy, and increased production of reactive oxygen species. Furthermore it is believed to have neuroendocrine and immunomodulation functions and has a strong linkage with the gene for human growth hormone (Schachter et al., 1994). It is also believed, that under some circumstances the *ACE* gene can be protective against serious cardiovascular diseases such as coronary artery disease (Pepine, 1997). Further studies in an enlarged cohort of French centenarians was able to confirm the association of longevity with *Apo E* but not with the *ACE* locus, which was considered as a false positive association (Blanche et al., 2001). In our second model we will examine the effect of *Apo E* on longevity in greater detail. We construct a virtual population to test the way it acts on the population to produce changes in mortality and we propose a new theory for its evolution. Other members of the apolipoproteins family, that have not been studied as extensively as *Apo E* but may play a role in human longevity are the *Apolipoprotein B VNTR* and *Apolipoprotein J* (Gonos, 2000).

Recently, a genome-wide scan using 308 individuals belonging to 137 sibships of exceptional longevity, found a significant linkage near microsatellite D4S1564 on chromosome 4 and longevity (Puca et al., 2001). Further work identified that the microsomal transfer protein (*MTP*) gene was the gene responsible for the linkage. MTP is thought to be the limiting factor of lipoprotein assembly. People with two non-functional copies of the gene suffer from abetalipoproteinemia and have a near absence of ApoB particles in serum. *MTP* was also associated with lipoprotein

profiles, insulin resistance and fat distribution, though unfortunately no life extending effects have been found for the gene as yet (Geesaman et al., 2003). Generally a wide array of genes linked with age related disease has been considered as affecting longevity but with most of the time little experimental support (see De Benedictis et al., 2001).

Although, life extending mutations are not known in humans, the opposite, mutations that produce an ageing-like phenotype have been observed. The most famous of these is Werner's progeria, a rare autosomal recessive disorder with symptoms resembling ageing, such as greying hair, skin thinning, cataracts, atherosclerosis, osteoporosis and high risk for tumour development (Hegele, 2003). The first symptoms begin to appear around the age of puberty, when the typical growth spurt is absent resulting in small stature, while most patients die before the age of 50. A mutation of a single gene (*WRN*), belonging to the RecQ family of DNA helicases, is responsible for the disease. The gene product of *WRN* is involved in a number of different enzymatic activities including strand displacement as a DNA-dependent ATPase, a 3`-5` DNA helicase for DNA unwinding and a 3`-5` exonuclease activity (Bohr et al., 2002). Additionally, participation of human helicases *WRN* and *BLM*, responsible for the Bloom syndrome, in a protein complex functioning in DNA metabolic processes at telomeric ends, has also been found (Opresko et al., 2002). Recently, it was proposed that cases of an atypical form of Werner's syndrome are due to a mutation of a nuclear lamin A/C gene, a hotly debated subject (Chen et al., 2003). A known laminopathty which also causes a segmented[5] ageing phenotype, very close to that seen in Werner's syndrome, is the Hutchinson-Gilford progeria, a very rare autosomal dominant disease (<1/1,000,000)

[5] A phenotype that exhibits some ageing related characteristics but not others.

leading to death usually around age 14 (Hegele, 2003). In this case too, a single mutation in the *LMNA* gene encoding for the nuclear lamin A/C, comprising the nuclear lamina, is responsible. Other mutations in single genes are also known to cause progeria, such as the Cockayne syndrome, linked to mutations of the *CSA* gene in chromosome 5, showing profound postnatal growth failure and systemic signs of premature ageing, probably due to defects in DNA repair (Rapin et al., 2000); and the Trichothiodystrophy syndrome, due to mutation in either of two helicases *ERCC2/XPD* or *ERCC3/XPB*, which encode the 2 helicase subunits of transcription/repair vector TFIIH, leading to a compromised transcription due to accumulation of DNA damage and inactivation of critical genes and enhanced apoptosis (de Boer et al., 2002; Hasty and Vijg, 2002). Finally, ataxia telangiectasia is caused by a mutation in the *ATM* gene, encoding a homologue of phosphatidylinositol kinases which plays a fundamental role in signal transaction, while the patients show cerebellar dysfunction, immunodeficiencies, graying of hair and cancer predisposition, among other effects (Jazwinski, 1996; Martin and Oshima, 2000). These rare mutations, as we saw, are usually linked with the failure of the cell to maintain genomic integrity. It is surprising to see how single gene mutations associated with repair and maintenance of the DNA can have such large effects in the survival of the organism and the production of an ageing-like phenotype (Hasty et al., 2003). Nevertheless, evolutionary theories of ageing, and especially the disposable soma theory, long before the molecular genetics of these mutations were known, hypothesized the existence of such basic mechanisms that will regulate the rate of ageing. On the other hand, no examples are known for the ability of these pathways to increase longevity, limiting their potential to intervene in the process of senescence.

Despite the very promising genetic pathways influencing ageing discussed, such as oxidative stress resistance, insulin pathways and DNA repair and maintenance, we should not forget that only around 25-30% of longevity is actually in our genes (Ljungquist et al., 1998; Cournil and Kirkwood, 2001; Mitchell et al., 2001). As we shall see later, environmental conditions in adulthood, early developmental plasticity and maternal effects and probably most importantly, stochastic events, are of paramount importance for the quality and length of our lives.

1.4 Aims and scope of this project

The aim of this thesis is to elucidate the gene-gene and gene-environment interactions that affect ageing and longevity in humans. That is achieved through the construction of theoretical models of population and evolutionary genetics, based on epidemiological and demographical studies in contemporary populations. The models developed can help identify critical genes involved in ageing and account for some of the current data while our predictions can be further tested with the use of population studies. Furthermore, the project also aims to provide a theoretical approach to the intrinsic variability of the ageing process and account for the recently observed deceleration of the mortality in later ages.

As a first step towards modelling life history theory and the effects of environmental or genetic parameters on longevity we re-address the original model of the disposable soma theory (Kirkwood and Rose, 1991). Using the Mathematica programming language we analyse the parameters of the model and see how the fundamental predictions of the theory are derived from the equation used to describe the trade-off between longevity and reproduction.

In our first model we describe a trade-off between reproductive success and survival based on the hypothesis that women with an increased resistance to potentially fatal infections are at greater risk of rejecting fetal implantation, and we examine its effects in populations with high and low rates of infection-induced mortality. We use the model to predict the rate of change of the responsible gene frequencies during rapid change of infection rate, such as has occurred in developed countries within the last century. The results resolve the paradox of genes that impair fertility in present-day populations by showing how they may be a legacy of a powerful trade-off that probably acted until very recently in human evolution in developed countries and is likely still to be operating in developing ones.

The second model addresses the action of longevity assurance genes and their interactions with a number of environmental parameters. Apolipoprotein E is used as an example of such a gene and its interaction with a number of lifestyle parameters affecting the onset of cardiovascular disease is examined. Furthermore, the variability of gene-environment interactions and the stochasticity of statistical distributions are considered as the possible key factors responsible for the heterogeneity of the ageing process. We also use the model to explore the force of natural selection on late acting genes, such as Apo E, and their future and past evolution.

Finally we consider the effects of social class on longevity and ageing, showing how, physiological and psychological stress, attitude towards health, lifestyle choices, access to health services and early developmental effects can increase the intrinsic variability of ageing. Insights from our previous work are examined in the light of social variability and the possible effects of such variability are explored. Our analysis suggests that evolutionary differentiation due to social effects will have had

little impact on the genetics of human longevity, although social factors are very important in influencing gene-environment interactions.

2. Basic Aspects of Life History Theory and the Disposable Soma Theory of Ageing

As we saw, the purpose of this project is to develop theoretical models of gene and gene-environment interactions affecting human ageing and longevity that can be tested through epidemiological and genomic studies directly in human populations. These aims are addressed with the use of mathematical models of the selection forces that act on a population in order to predict the extent of polymorphism expected with respect to factors influencing the ageing process. The first model that will be discussed is of the disposable soma theory as described by Kirkwood and Rose (1991), which we will subsequently develop further to account for other kinds of trade-off, especially those that work through the immune system during pregnancy.

Before we describe the models some introduction to the basic terms, methods and programs used in this work will be given.

2.1 The use of mathematical models

Human conceptual abilities are limited when compared to the complexity that surrounds us and which we may wish to explain. The fundamental factor in how we choose to describe complexity is the way we use abstraction to group observations and focus only on the particular problem that we want to solve. So, during scientific inquiry we take the problem and reposition it in a conceptual universe, reflecting our theories about its behaviour, which are based on the different parts of reality we observe. This is the model used to test our predictions, which may sometimes be a simple conceptual model and at other times a formal mathematical model. Thus, the

use of models is common in all branches of science, whether this is based on physical observation, manipulation of a natural system or a series of differential equations. Even when we use a model organism, as seen earlier, to obtain information on the evolutionary processes and mechanisms relevant to ageing we treat the model as a realization of life in general, based on the assumption that its intrinsic properties are isometric with other living organisms. Since our basic assumption of isometry is only partially true, our model is only a metaphorical realization of the problem in an isolated conceptual reality (Lewontin, 1968).

More specifically, mathematical models of biology provide us with an even wider capacity to abstract from the problem and test hypotheses that are either complex or non-testable by experimental methods. This provides us with a very powerful tool for the extension of our logic and the chance to reason based on the properties of numbers. According to Nietzsche, mathematics exists due to our ability to discount the facts that no two real objects are exactly the same to be counted, and that no exactly straight line, perfect circle, or absolute magnitude exists in nature (Nietzsche, 1996). This specific ability, to reduce the problem only to its relevant characteristics, is what makes the use of mathematics central in the study of complex systems in biology and science in general. It enables us to construct a frame of equations and mathematical principles that describe a testable logical argument. In contrast to physical models, mathematical models have no innate properties; instead they are based on a number of assumptions. The delicate balance between assumptions and reality has been a hotly debated issue, especially between theoreticians and experimentalists where the latter accuse the former for oversimplification. Undoubtedly, instances do occur where too simple or too complex assumptions have led to misrepresentation of actual processes leading to erroneous

conclusions (May, 2004). Thus, assumptions should be made carefully in order to be complex enough to represent reality but simple and general enough not to restrict the model by imposing a very limited set of parameter values and conditions. A second point that has commonly attracted criticism in mathematical modelling is how a disjunctive statement can have any value in predicting patterns that were not included in the model (Sober, 2002). Not all models though need to be predictive. In some cases their main requirement is to fit and perhaps explain the available data. In this case we consult the existing information and find the parameter values that maximise the probability of the data (Akaike, 1973). In the very simple case of the comparison between two data sets, an extremely common practice in biology, we use our expectation that the null hypothesis is true to infer a prediction for future observations by assessing the probability that the null hypothesis holds or not (in the case of a Bayesian analysis) or that the data could have arisen with reasonable probability (in the case of a frequentist analysis). Even models that are historically proven to be predictive, such as Einstein's general theory of relativity or Poisson's prediction of optical diffraction, are all based on the condition that they fit previous knowledge (Hitchcock and Sober, 2004). Theoretical models should contain enough information to be able to make predictions but not so much as to overfit the data on an erroneous hypothesis (Hitchcock and Sober, 2004).

The different levels of balance between simplicity and realism in mathematical models make them challenging to use. Great care needs to be taken both in the assumptions and hypotheses used for their formulation. The different branches of science have long benefited from a theoretical approach to complex problems. In biology, mathematics were introduced in the area of quantitative genetics in the early 20th century and they have very quickly became an integral part of the study of life.

Like all new techniques, though, theoretical models have been the subject of criticism and misuse. This does not mean that they are of limited or uncertain use but rather that they require time to develop as a methodology and they should, like all scientific research, be treated with a critical eye. We will continue section 2 with the description of well tested mathematical models that have been the standard tool for the evolution of life history. In this case a theoretical approach is the only possible way to account for changes through natural selection since the time required for those changes to take place make any direct experimentation, at least in the case of humans, forbidding.

2.2 Measuring fitness

The theoretical scope of the project requires the introduction of ways to measure the fitness of an organism, and since we are interested in its genetic properties, the fitness of its genotype or individual genes. Although a universally accepted definition for fitness does not exist we could say that in asexual organisms the fitness of the genotype is defined by its success, or not, in producing clones of itself, whereas in organisms with a sexual mode of reproduction, we can only refer to the fitness of genes in terms of their success in passing as many copies of themselves as possible into the next generation.

In the models to follow, fitness is measured as the instantaneous rate of natural increase r, as this is given by the Euler-Lotka equation.

$$\int_0^\infty e^{-rx} l(x) m(x)\, dx = 1$$

This is the basic equation for life history evolution and it was first postulated by the mathematician Leonhard Euler in 1760, and it was independently re-discovered by

Alfred Lotka, a mathematical demographer, in 1907. The model links the probability of survival to a certain age class *l(x)*, to the number of offspring expected in that age class *m(x)*, and the rate of growth of the population *r*. The requirements for the Euler-Lotka equation to hold are that the population is in stable age distribution[6], and that the mortality and fecundity remain constant in time. Although the above do not usually hold in natural populations, minor deviations have a very small impact on the quantitative predictions of the model (Hamilton, 1966; Stearns, 1992; Charlesworth, 1994).

It is easy to imagine how ***r*** can also be applied to measure the fitness of a single genotype, as when reproduction is through clonal expansion, since we can consider each genotype as a separate population increasing in a different rate (Charlesworth, 2000). In the case of diploid organisms with sexual reproduction though, ***r*** is only an approximation of the rate of change of an allele frequency when selection is weak. Sex differences in survival, non-random mating, density-dependent changes in mortality and fecundity, and finally, spatially and temporally fluctuations in environmental conditions can have important effects on the level of approximation obtained by ***r***, although these parameters can usually be ignored in large populations. Other, more direct methods of change in allele frequencies can also be used as a measure of genetic fitness, as we will see in our *Apo E* model.

Next we will see how the Euler-Lotka equation can be used to predict the optimal strategy that an organism has to follow, in relation to ageing, to reach an evolutionary stable strategy under the disposable soma theory.

---

[6]The proportion of organisms in each age class remains constant in time

2.3 Programming with Mathematica version 4.0

Before we explain and test the model of the disposable soma theory of ageing, we should also consider the programming aspect of it. Due to the large amount of complex calculations required, computers have proved to be essential tools for theoretical biology. In our analysis of the disposable soma theory model a mathematical programming language, called Mathematica was used. In more detail, Mathematica is a fully integrated mathematical computing system combining interactive calculations, visualization tools and a complete programming environment (Wolfram, 1999). Its relative simplicity and its capacity to be programmed give it a flexibility that makes Mathematica a very successful tool for biological modelling. The model that follows is written in Mathematica and it is presented in its original output form of a Mathematica notebook. A notebook is an interactive document format that can combine specific commands in the syntax of Mathematica-encoded mathematical expressions, free text which can include standard mathematical expressions, and display items of output such as graphs.

2.4 Analysis of the disposable soma theory model.

Although the first appearance of the ideas behind disposable soma theory was in a *Nature* paper of 1977 (Kirkwood, 1977), the first description of the underlying mathematical principles were presented much later (Kirkwood and Holliday, 1986; Kirkwood and Rose, 1991). Here we will replicate the model using Mathematica and we will discuss and analyse the conclusions and implications of the theory as those are derived from the model.

As we saw earlier, the disposable soma theory is based on the framework of physiological ecology and its main target is to describe the optimal balance between reproduction and maintenance in a given organism under given conditions. As the problem of the evolution of senescence is considered under the light of life history theory the Euler-Lotka equation is used to measure fitness

$$\int e^{-rx} l(x,s) m(x,s) dx = 1 \quad (2.1)$$

where survival $l(x,s)$ and fecundity $m(x,s)$ are expressed as functions depending on age x and investment in maintenance s. From equation 2.1 we can see that a complex relationship exists between the Malthusian parameter r and investment in maintenance s.

The equations for the schedules of mortality and reproduction can be derived from the Gompertz-Makeham model seen in section 1.1.2 (equation 1.2). The change of survivorship $\delta l(x)$ during a infinitely small change of age δx can be calculated by multiplying the rate of mortality $\mu(x)$ from the Gompertz-Makeham equation at age x with the number of survivors $l(x)$ at age x

$$\delta l(x) = -\mu(x) l(x) \delta x \quad (2.2)$$

Solving equation 2.2 for $l(x)$ gives us the analytical formula linking survival with age x, and the parameters basal vulnerability α, actuarial ageing rate β and age independent mortality γ of the Gompertz-Makeham model. Assuming that the age of maturation is a and the ratio of the population reaching this age is $(1-V)$ where V is the total juvenile mortality we can express survival to age x as:

$$l(x) = (1-V) \exp[\frac{-\alpha(e^{\beta x} - e^{\beta a})}{\beta} - \gamma(x-a)] \quad (2.3)$$

The decline of reproduction can be obtained if we consider that a similar underlying deteriorating mechanism operates as in the case of survival. Considering

that at age of maturation a reproduction $m(a)$ starts with a maximum value h and that no age independent decline exists, we can express reproduction at age $x \geq a$ as:

$$m(x) = h \exp[\frac{-\alpha(e^{\beta x} - e^{\beta a})}{\beta}] \quad (2.4)$$

We should note that by using the same value of β in both cases of age dependent survival and fecundity we assume that the mechanisms underlying the two processes have the same magnitude. One consequence of using a single β is that the model can not account for menopause and post reproductive lifespan. This has little effect on the original analysis since the life history of mice was used to fit the model but when data from human populations are used the discrepancy is important. In this instant we will use a fixed upper limit of reproductive lifespan of 50 years of age after which we will consider that the value of $m(x)$ is zero.

The central concept of the disposable soma theory is the existence of an energy budget that is invested in maintenance and reproduction dependent on the ecological expectations of the organism. Kirkwood and Rose (1991) included investment in maintenance s (on a scale from 0 to 1) in three parameters of the model, the actuarial ageing rate β, age of maturation a_0, and maximum reproduction h_0. Supposing that a value s_0 exists for the investment of maintenance for which the organisms does not undergo senescence we can express the rate of ageing for $s < s_o$ as:

$$\beta = \beta_0(\frac{s_0}{s} - 1) \quad (2.5)$$

where β_0 is the actuarial rate of ageing when only 50% of s_0 is invested in maintenance. For values of $s = s_0$ and above β is assumed to be zero and the organism does not suffer from age dependent mortality. It was also assumed that the age of maturity decreases as more and more energy is invested in reproduction to the point of

a_0, the age of maturity when all of the available energy is directed towards reproduction.

$$a = \frac{a_0}{1-s} \quad (2.6)$$

Again, in the case of reproductive rate at the age of maturity h it was assumed that there is a maximum reproductive rate h_0 which is achieved when all the energy is invested in the reproduction of the organism so that:

$$h = h_0(1-s) \quad (2.7)$$

Giving values to all the parameters of the model, mirroring what is seen in primitive human populations, we can plot the change in survival $l(x)$ and fecundity $m(x)$ with age x and investment on maintenance s. We can see in fig 2.1 that increasing investment in maintenance significantly increases the lifespan of the organism (a) while at the same time age at maturity increases, maximum rate of reproduction decreases and the decline of reproduction later in life is less steep (b). As we saw earlier in section 1.1.2 the Gompertz model cannot account for very early life mortality, and since the model is based on the Gompertz equation it inherits the same limitations which can be seen in figure 2.1.a as the straight line in the left part of the graph representing juvenile mortality. To be able to fit human data into the model we included an upper limit for reproduction, which can be seen as the point where the right most curve is cut off in fig 2.1.b. The effect of a sudden end to reproduction becomes more apparent as the investment in maintenance increases. With the model to be described in chapter 3 we will address again the problem of menopause in the disposable soma theory model and we will adopt a more realistic strategy to account for the end of reproduction.

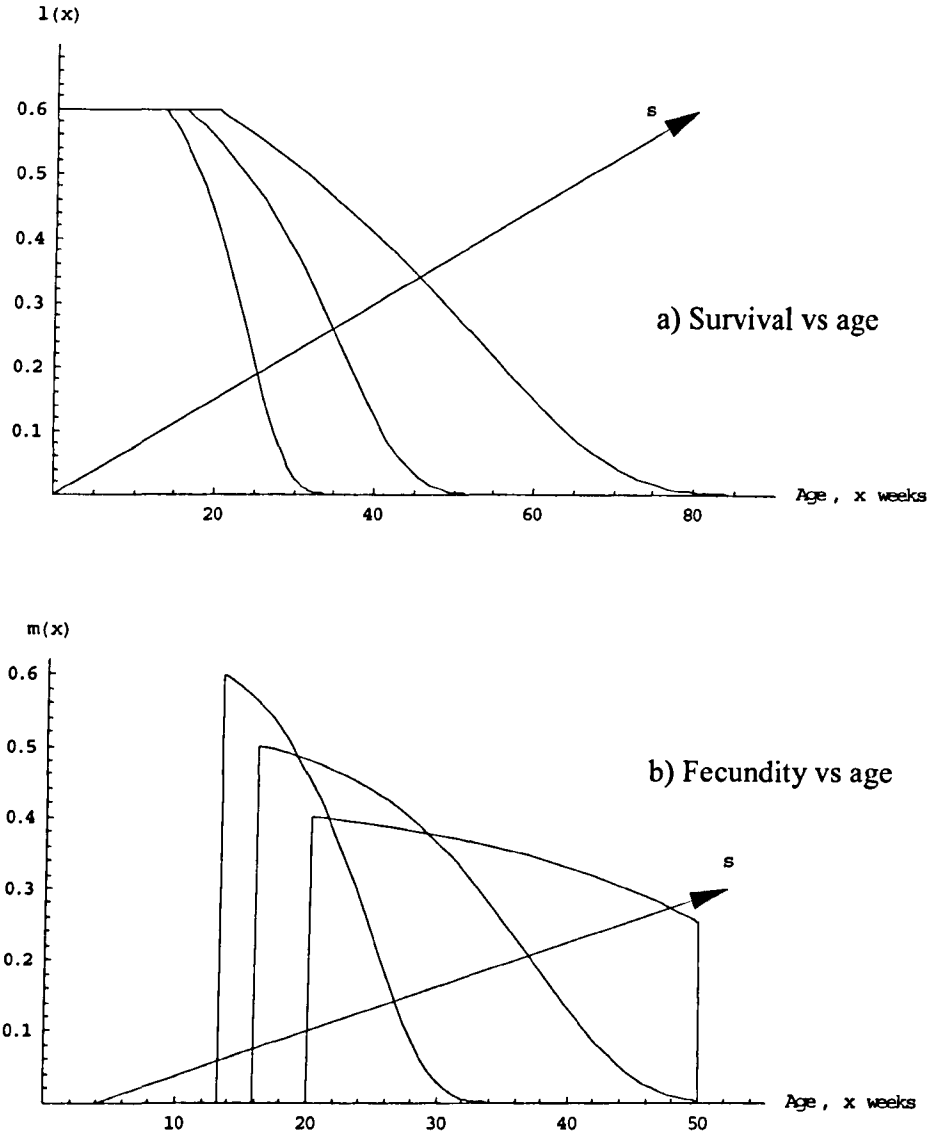


Figure 2.1. **a)** Graph representing the change of survival with age for three values of investment in maintenance, 0.4, 0.5, and 0.6 from left to right **b)** Graph representing the change of fecundity with age for the same changes of investment in maintenance as in (a).

The opposite effects that increasing or decreasing investment in maintenance have on reproduction and survival are among the main predictions of the disposable soma theory, where high early development is expected to be followed with an increased rate of mortality, while a long lifespan is linked with a lower reproductive potential. Earlier, in section 1.2.7.1, we saw a number of experiments in *Drosophila* and other model organisms that seem to support this idea of trade-offs between longevity and reproduction (Luckinbill et al., 1984; Partridge et al., 1999).

Having calculated the survival and fecundity curves we can now proceed to find the optimal level of investment in maintenance for the organism under the given conditions. Using equation 2.1 we can construct a curve between the instantaneous rate of natural increase r and the investment in maintenance s as can be seen in figure 2.2. The value of s, in this case 0.5, corresponding to the highest value of r, 0.05 in fig 2.2, is the optimum strategy for a human population living under the conditions described by the parameters used. The relatively flat area around the maximum permits strategies with small differences to coexist in the population.

We will continue our analysis of the disposable soma theory by testing the effects of changing the parameters of the model and describing the resulting changes of the optimal balance between investment in reproduction and survival and their evolutionary consequences. The first parameter to be considered is total juvenile mortality. It is easy to understand that by lowering the ratio of the population reaching maturity we also decrease the rate of population increase as can be seen in figure 2.3.a. Furthermore, by increasing juvenile mortality we observe that the point of the optimal balance between survival and fecundity is shifted towards higher values

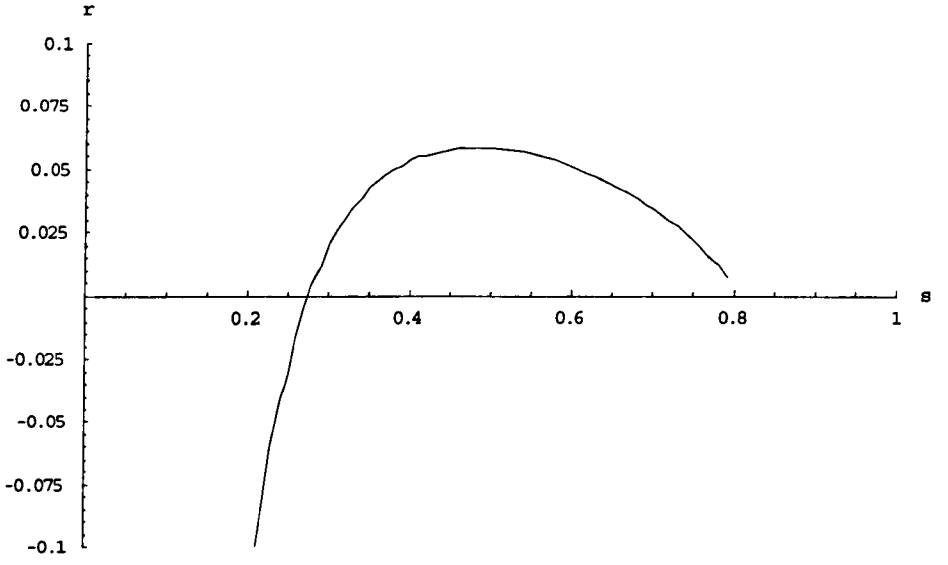


Figure 2.2. Curve describing the association between the malthusian parameter r and investment in maintenance s. The maximum point of r is 5% for a 0.5 value of s. The parameters used are: $\alpha = 0.001$, $\beta_0 = 0.2175$, $V = 0.4$, $\gamma = 0.01$, $a_0 = 8$ years, $s_0 = 0.8$, $h_0 = 1$ child/year.

of investment in maintenance. Thus the model predicts that when a small ratio of offspring reaches maturity, adult organisms invest less in reproduction and more towards survival, prolonging both their longevity and reproductive lifespan. Similarly, when the basal vulnerability α or the actuarial rate of ageing β increase, the ability of the population to grow is lowered, while the optimal investment in maintenance required increases (figure 2.3.b and c). So according to the disposable soma theory model, under conditions of high intrinsic mortality the organism should invest more towards maintenance of its body in order to maximise its fitness since the accumulated damage lowers the lifespan of the organism to a level far lower than that expected under the specific environmental parameters. In contrast, when external mortality is increased, we see in figure 2.3.d that the level of investment in maintenance is decreased and the energy is diverted towards higher reproductive output and a shorter lifespan, as was experimentally verified in natural population of guppies (Reznick et al., 2001; Bryant and Reznick, 2004). The above changes are based on the principle that the disposable soma theory predicts that the organism invests enough in maintenance to extend its lifespan, without wasting resources by extending life longer than what is normally likely to be experienced in wild populations subject to external mortality.

After the analysis of the effects that the mortality parameters have on the optimal strategy of the organism, we will now consider what happens when the schedule of reproduction is altered. Increasing ages of maturity push the optimal level of investment to the left of the graph, as can be seen in figure 2.3.e, towards higher values of s while the maximum rate of increase is lowered. Longer periods of pre-reproductive life require more energy to be invested towards the soma to ensure that

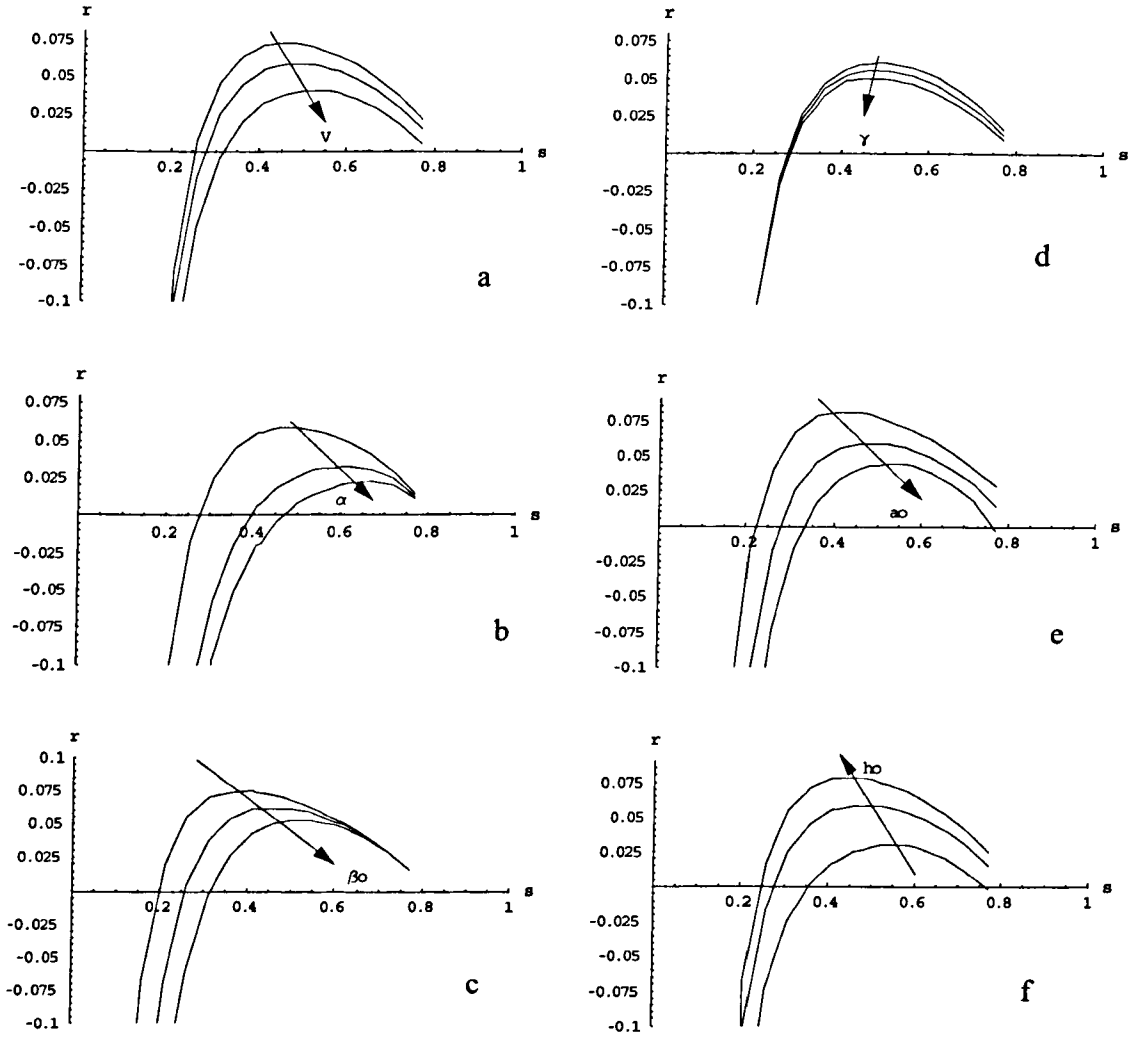


Figure 2.3. Graphs representing the changes of the optimum strategy when the parameters of the disposable soma are changed **a)** changing juvenile mortality V (values: 0.2, 0.4, 0.6) **b)** changing the basal vulnerability α of the population (values: 0.001, 0.0055, 0.01) **c)** changing the actuarial rate of ageing β (values: 0.15, 0.20, 0.25) **d)** changing the age independent mortality γ (values: 0.0025, 0.02125, 0.04) **e)** changing the minimum age of maturity a_0 (values: 6,8,10) **f)** changing the maximum rate of reproduction h_0 (values: 0.5, 1, 1.5).

the organism will remain in good condition at least till the onset of reproduction. On the other hand, when the maximum level of reproduction is increased the investment in maintenance required to keep the population at its maximum rate of increase is lower (figure 2.3.f). Since the ratio of offspring reaching adulthood remains the same while in the same time the ability to reproduce increases, the organism can achieve a higher level of fitness by investing towards reproduction. This has also been verified by studies of human historical population where a higher number of children was associated with a shorter lifespan (Westendorp and Kirkwood, 1998; Korpelainen, 2000). Moreover, as we saw in figure 2.1.b, lower investment in maintenance leads to an increase of early reproduction compared with that later in life. Again, experiments with *Drosophila* lines selected for early fecundity saw a pronounced decrease both in later reproduction and lifespan (Rose and Charlesworth, 1981; Luckinbill et al., 1984). The complete notebook for the computer program of the disposable soma theory model can be seen in Appendix 1.

2.5 Conclusions drawn from the disposable soma theory of ageing.

We have already discussed the general predictions of the disposable soma theory and the evidence supporting them. Our analysis has elucidated the detailed effects of changes in the investment in maintenance on survival and reproduction, as well as their combined effect on the optimal strategy that the organism has to follow to achieve maximum fitness under a specific set of conditions. Testing the disposable soma theory has also provided us with a number of interesting observations. While intrinsic rates of increased mortality increase the required level of maintenance for the optimal strategy, higher external mortality increases the investment towards reproduction. Selection for higher early or total reproduction will decrease lifespan,

while selection for late reproduction will increase longevity. Some weaknesses or shortcomings of the models were also made clear. For example, the model, in its present simple form, can not account for menopause in humans, something that we will try to improve in the immunogenetic trade-off model in the next chapter. Additionally, we saw that there is no distinction between the reproductive investment in offspring production and the investment in parental care. The later is an important parameter of the life history of the organism and especially humans, where the transfer of energy between generations constitutes a great proportion of the total energy available to an individual and can affect future reproduction and the fitness of the offspring.

Despite its simplicity and a small number of problems, the disposable soma theory model is a very useful descriptive and predictive tool for research into ageing. It can account for a number of different experimental observations in a concise and clear way and point towards basic systems involved in ageing. Furthermore, its simplicity is making the model a very good platform where additional parameters can be added to describe more detailed trade-offs with specific gene mechanisms, as we will show in the next chapter.

3 Immunogenetic Basis For Trade-Offs Between Human Fertility And Lifespan

3.1 Introduction

As we saw earlier, the genetic basis of human longevity is an intriguing puzzle (Cournil and Kirkwood, 2001). The emergence of the modern evolutionary theory of ageing (Kirkwood and Cremer, 1982; Rose, 1991) follows many centuries of theoretical speculation about the factors influencing length of life, which began around 350 BC with the writings of Aristotle (Barnes, 1984). The modern evolutionary theory recognises that ageing is a non-adaptive character that arises through the decline in the power of natural selection at later ages. A key aspect of the evolutionary theory is the prediction that trade-offs which favour fitness in early life will be favoured by selection, even if the later deleterious consequences of such trade-offs result in senescence and death (Partridge and Barton, 1993; Kirkwood and Austad, 2000).

We have already discussed that among the ideas that make up the current evolutionary understanding of ageing, the one that most explicitly addresses the nature of potential trade-offs in an organism's physiology is the disposable soma theory (Kirkwood and Rose, 1991; Kirkwood, 1992; 1993; 1997; 2002). The disposable soma theory recognises that organisms must allocate resources among a number of metabolic demands, including somatic maintenance, growth, storage, reproduction, immune function, and so on. Under pressure of natural selection, the investment in survival must be balanced against investments in traits such as fertility, and the optimal balance will be influenced by ecological factors such as the level of

extrinsic mortality (Kirkwood, 2002). Thus, the disposable soma theory generally predicts a negative relationship between fecundity and age-dependent survivorship (although in some instances there might be trade-offs between, for example, somatic maintenance and fat storage that involve fecundity only indirectly). We also show that in addition to these specific trade-offs involving investments in somatic maintenance, Williams (1957) suggested that there might exist a broad class of pleiotropic gene actions such that genes favouring an organism's fitness early in life might have deleterious consequences for its fitness at later ages, which are also generally predicted to produce a trade-off between fecundity and age-dependent survival.

As we have already discussed at length, numerous studies have examined the existence of inverse relationships between longevity and fecundity. The species most commonly used for this purpose have been members of the *Drosophila* fruitfly family. Trade-offs between fecundity and survivorship have been observed using a variety of methods ranging from artificial selection to genetic manipulation (Rose and Charlesworth, 1981; Luckinbill et al., 1984; Zwaan, 1993; Tatar et al., 1996; Sgro and Partridge, 1999; Zwaan, 1999; Stearns et al., 2000). Work on nematodes, mice and guppies has provided further evidence for the existence of trade-offs in other species.

In recent years, studies have begun to investigate trade-offs between longevity and fecundity in humans. Despite the difficulty of obtaining the necessary records for historical populations Westendorp and Kirkwood (1998; 1999b) found that among British aristocrats, there was a significant inverse association between fertility and lifespan. Further studies by Thomas et al (2000), Korpelainen (2000) and Doblhammer (2000) have confirmed this negative association between human longevity and reproduction, although Helle (2002) did not find it, while Lycett et al

(2000) have shown that the trade-off appears to be affected by socioeconomic conditions.

Our main focus in this chapter is a recently formulated immunogenetic mechanism which might explain a trade-off between human survival and fertility (Westendorp et al., 2001). This is based on evidence that innate immune factors influence both the ability to withstand fatal infections and the capacity to initiate and complete a successful pregnancy (Piccinni, 2002), but in opposite directions. A major difference is characterised in terms of the T-helper type-1 (Th1), associated with low interleukin (IL)-10 and high tumour necrosis factor (TNF-α) levels, and T-helper type-2 (Th2) phenotypes, associated with the inverse cytokine profile. In the Westendorp et al (2001) data set only 1out of 9 women with the Th1 phenotype exhibited normal fecundity, while 3 out of 9 with the Th2 phenotype had any reproductive impairment, defined as having at least 3 consecutive spontaneous abortions before the 16^{th} week of gestation. Although the importance of immunological factors in human reproductive failure remains controversial, there is clear evidence of the need for a special immunological relationship between mother and foetus which permits tolerance of the foetus as a foreign graft (Medawar, 1960; Choudhury and Knapp, 2001). Women with an innate Th1 phenotype have a greater probability of spontaneous abortion than women with an innate Th2 phenotype. Conversely, individuals with an innate Th1 phenotype show lower susceptibility to fatal infection, indicative of a greater likelihood of survival in an environment where infectious diseases exert a significant mortality pressure. The hypothesis that we examine in our model is that the selection push of the population towards Th1, based on increased survival, will have been countered by selection towards Th2, based on increased fertility (Westendorp et al., 2001).

3.2 Methods

Modelling life history trade-offs

Our model begins from the formalism described by Kirkwood & Rose (1991) to examine the predictions of the disposable soma theory. This required that fecundity and survival should both be dependant on the allocation of resources to somatic maintenance, so that by combining the schedules of fecundity and survival to compute fitness (as measured by intrinsic rate of natural increase), the optimal investment in maintenance could be determined. We have already discussed the underlying theory in section 1.2.6 while in the second chapter we described and analysed the model proposed by Kirkwood and Rose (1991) in some detail.

Survivorship

For the purposes of modelling the immunogenetic trade-off, we assume a variable η, equal to Th2/Th1 ratio expressing the phenotypic response of the immune system. The effects of varying η on survivorship are complex and must take account of effects on susceptibility to infectious disease across a range of ages within the population. Four parameters can, in principle, affect the survivorship schedule: child mortality V, and α, β, and γ of the Gompertz equation (eq. 1.1). We focus our attention on modifying α and γ, assuming β to be a characteristic of the species (see also Finch, 1990). We do not modify V, which corresponds to juvenile mortality, since our specific concern here is to model life-history trade-offs that affect the balance between fertility and immunity to life-threatening infectious disease during the reproductive years (our calculation of fitness – see below – uses the standard Euler-Lotka measure, which integrates the product of the fecundity and survivorship

terms; this product is zero during the juvenile phase of the life-history (eq. 2.1)). However, by changing γ we do account for significant effects on the early mortality in the young mature population.

We express γ as a function of η in the form:

$$\gamma(\eta) = \kappa\eta + \gamma'$$

where κ is a scaling constant and γ' is the basic extrinsic mortality of the population. This causes a large effect on mortality rates at younger adult ages when the age-dependent increase in the intrinsic death rate is only small, but it has only a small effect at older ages (Figure 3.1a). Such an effect might describe diseases such as malaria, cholera and measles.

The effect of increasing η on α will also be to increase mortality but this will have relatively little effect until later ages when the age-dependant increase in the intrinsic death rate will be greatly increased (Figure 3.1b). We represent this in the model as:

$$\alpha(\eta) = \lambda\eta + \alpha'$$

where λ is a scaling constant and α' is the original basal vulnerability of the population. This might describe diseases like pneumonia and tuberculosis.

We can insert these expressions for $\alpha(\eta)$ and $\gamma(\eta)$ into the Gompertz model and, following the methodology of Kirkwood and Rose (1991), obtain an expression for survivorship $l(x, \eta)$ as:

$$l(x,\eta) = (1-V)\exp[\frac{-\alpha(\eta)(e^{\beta x} - e^{\beta a})}{\beta - \gamma(\eta)(x-a)}]$$

The impact of varying the investment in maintenance, as we saw in section 2.3, is incorporated by making the variables a (age at maturity) and β (actuarial ageing rate)

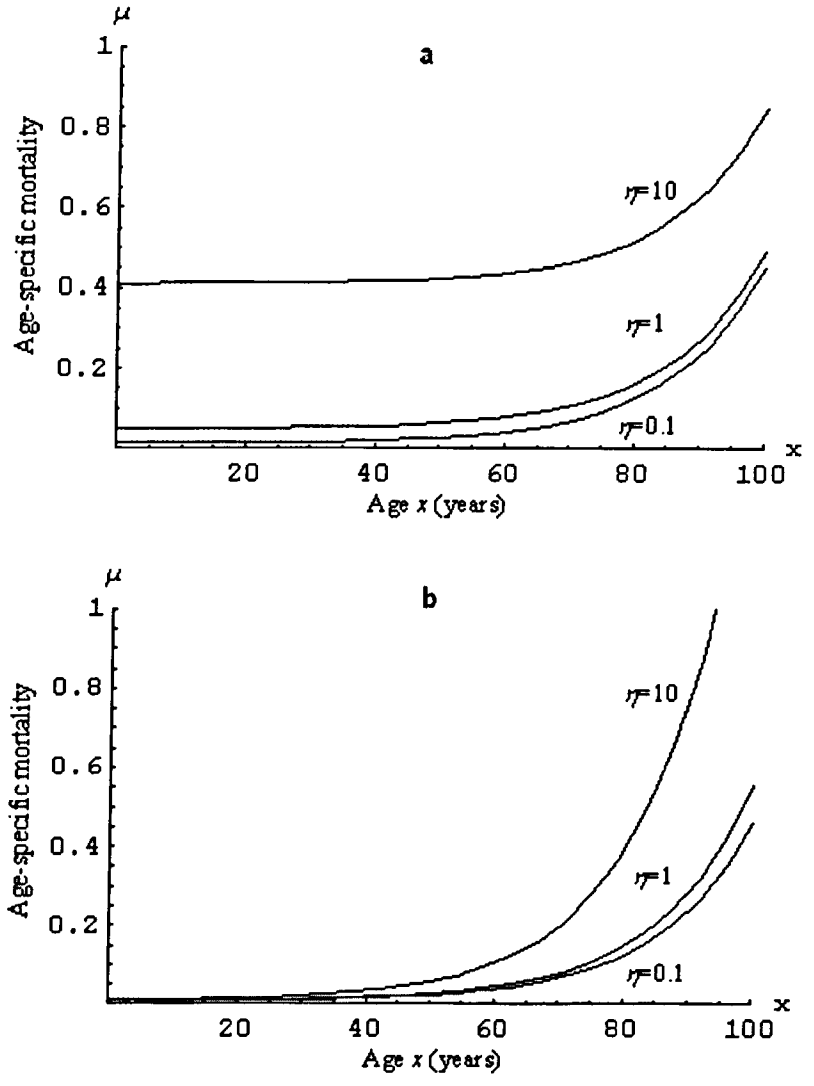


Figure 3.1. Effects on age-specific mortality of varying η assuming that: **a**). η affects the parameter γ (extrinsic mortality) of the Gompertz-Makeham equation; **b**). η affects the parameter α (basal vulnerability).

depend on a parameter s, which takes values in the range 0 to 1. The rationale for these assumptions is that the age of maturation should increase as more resources are directed towards maintenance, whereas the ageing rate should fall and can reach zero if sufficient resources are invested in maintenance.

Fecundity

We suppose that a reduced value of η, i.e. a shift towards Th2, is associated with improved reproduction so that

$$m(x,\eta) = m'(x)\frac{Th2}{Th1 + Th2}$$

or

$$m(x,\eta) = m'(x)\frac{\eta}{\eta + 1}$$

where $m'(x)$ is the basic fecundity predicted by the Gompertz law under the disposable soma theory. So fecundity, $m'(x)$, is assumed to be:

$$m'(x) = h\exp[\frac{-\alpha(e^{\beta x} - e^{\beta a})}{\beta}]$$

where h denotes the maximum reproduction rate, which is assumed to occur at the age of maturation a, and the multiplier represents a Gompertz-like decline in intrinsic fertility, due to senescence. As we have already discussed, the original model was postulated for the life history of the mouse where decline in fertility plays a much less important role than in the case of humans where the reproductive lifespan, especially in women, is much shorter than total lifespan. To account for menopause we assume that the reproductive system is ageing with a different rate than the rest of the soma. We can incorporate that in the model by considering a new actuarial ageing rate β_R for

the equation of fecundity. The relationship between β_R and β can be mathematically expressed as:

$$\beta_R = \theta \beta$$

where θ stands for a constant describing the factor of change between the decline of the reproductive system relatively to the soma of the organism. The use of θ links the age of menopause to the investment in maintenance s through its dependence to β, thus permitting the postponement of the age of menopause with increasing energy used for repair and maintenance increasing the realism of the model. Although menopause is almost an exclusively human characteristic, a number of other organisms also show a faster decline of fecundity with age compared to survival (Finch and Sapolsky, 1999). Even though the biological mechanisms for such a difference are not known we propose that it might be the result of intense selection among the precursors of gametes in order to ensure the soundness of the information passed to the next generation. In the case of humans, fitting the equation describing the reproductive decline to data from females in a contemporary western population (Office of National Statistics, UK) reveals that θ is equal to 4, thus the rate of reproductive ageing in a women is four times higher than that of the rest of her body.

Calculation of fitness

To assess the trade-offs associated with Th2/Th1 ratio we use as a measure of fitness the intrinsic rate of natural increase r given by the Euler-Lotka equation,

$$\int e^{-rx} l(x,\eta) m(x,\eta) dx = 1$$

which supposes that the population has a stable age distribution with the mortality and fecundity schedules remaining static in time. Although these assumptions do not

usually hold exactly for natural populations, minor deviations have a very small impact on the quantitative predictions of the model (Stearns, 1992).

Software and parameter values

Our models were programmed and run using *Mathematica* Version 4.0. Default parameter values were chosen so that around the fitness optimum the age-specific mortality schedule was similar to that observed for contemporary human populations (see Figure 3.2). Contour plots were obtained under various assumptions for fitness r as a function of η (assumed to range from 0.1 to 10) and s (assumed to range from 0 to 1). The entire Mathematica Notebook can be seen in Appendix 2.

3.3 Analysis of model

To analyse the model for the immunogenetic trade-off between fecundity and survival we will use two populations, from Denmark and The Gambia, as generic examples of exposure to high and low infectious load. Using the life tables statistics of Denmark (Lopez, 2001) we can calculate the parameters α and β of the Gompertz model, with the use of least-squares fitting to a non-linear parametric model, considering that no external mortality is affecting the population (Figure 3.2). In the case of the Gambian life tables, we used the Danish β parameter and fit a new α and γ for increased basal vulnerability and external mortality due to infections (Figure 3.2). Given the new values of α and γ though we can also calculate κ and λ of the model for the Gambian population. The next step was to assume a genetic structure such that investment in maintenance, s, and the immunogenetic phenotype, η, were

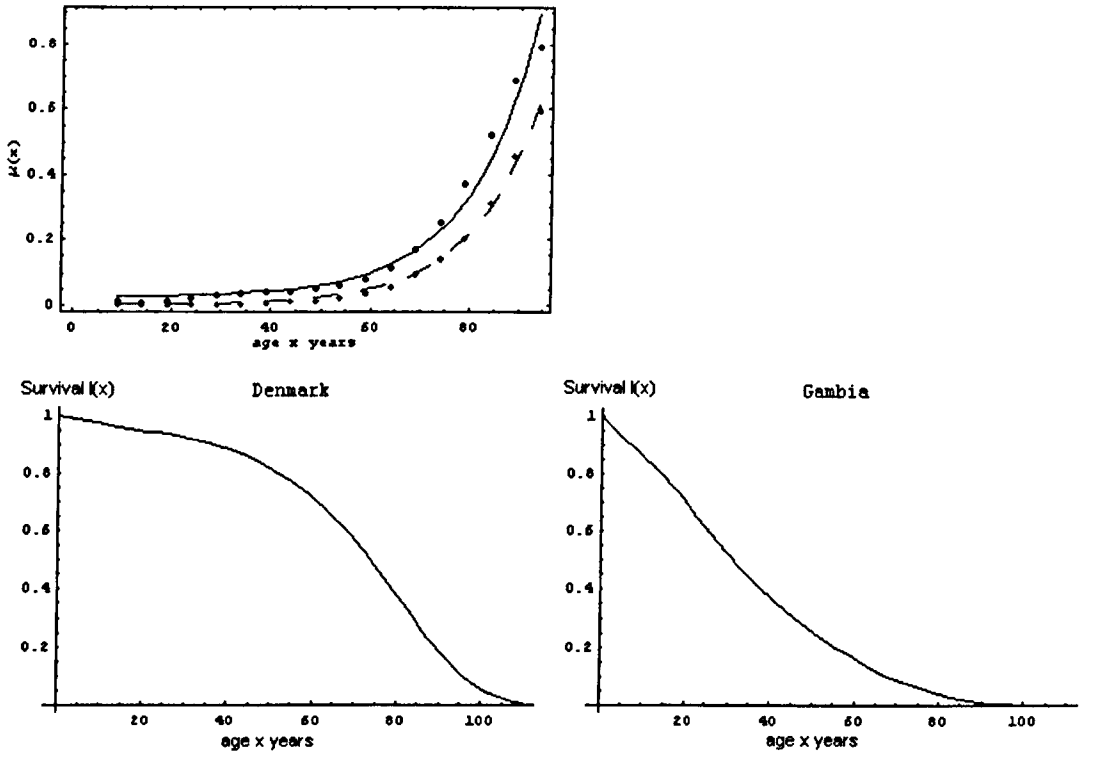
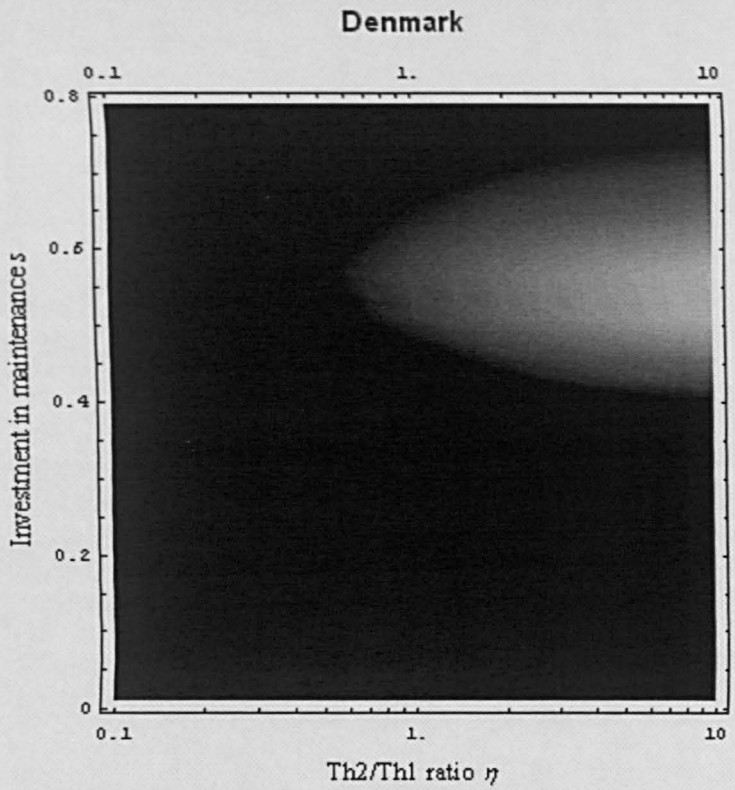


Figure 3.2. a) Comparison between WHO data values (points) and their fitted models (lines), for Denmark (dashed line and grey points) and the Gambia (solid line and black points) for the increase of the force of mortality with age **b)** Percent survival for Denmark and Gambia as a function of age, as predicted by our models in each case. All population values were either calculated by the fitted Gompertz curve or selected to mirror the expected realistic values in each case.

continuous characters ranging from 0 to 1 and from 0.1 to 10, respectively. Our assumption that η can vary from 0.1 to 10 is based on earlier estimations of an inter-individual ten-fold variation in production of both the Th2-type cytokine IL-10 and the Th1-type cytokine TNF-α (Westendorp, personal communication).

Based on the above we can now solve numerically the Euler-Lotka equation for both populations and get the fitness profiles expected from our equations for the evolution of life history (Figure 3.3). It is clear from Figures 3.3a and b that there is an optimum Th2/Th1 ratio, just as there is an optimum investment in somatic maintenance. The optimum is the point that best balances the advantage of better immune protection from fatal infectious diseases against the risk of increased spontaneous abortion, just as the optimum level of somatic maintenance balances the advantage of increased resistance to damage accumulation against the disadvantage of drawing too many resources from growth, reproduction, etc. Figure 3.3 shows how the fitness of the population is affected by assuming different exposure to mortality pressure such as might be exerted by differential exposure to fatal infectious diseases. As described earlier, to generate Figure 3.3a, the population was assigned life table data based on a contemporary West African population (The Gambia) where exposure to infectious diseases, like malaria, still causes significant mortality pressure at all ages while Figure 3.3b used life table data for a contemporary European population (Denmark) where the risk of fatal infectious disease is very much smaller. It can be seen by comparing Figures 3.3a and 3.3b that there is a marked rightward shift in the optimum value of η, signalling that the optimum strategy has moved in the direction of favouring a higher Th2/Th1 ratio in a contemporary European context.


Denmark
0.1
1.
10
0.8
0.6
0.4
0.2
0
Investment in maintenance s
0.1
1.
10
Th2/Th1 ratio η

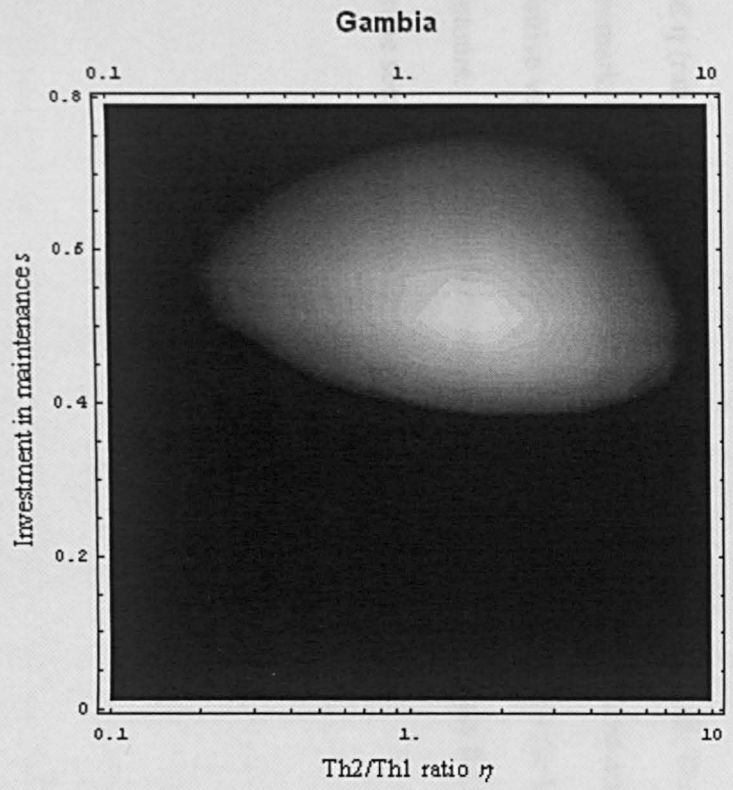

Gambia
0.1
1.
10
0.8
0.6
0.4
0.2
0
Investment in maintenance s
0.1
1.
10
Th2/Th1 ratio η

Figure 3.3. Contour plots of fitness r as a function of s (investment in maintenance) and η (ratio of Th1 to Th2), using mortality parameters for **a**). The Gambia, and **b**). Denmark, chosen to represent high and low infection load, respectively. Only positive values of r are shown with a maximum value of 0.03 for both cases and 30 contours. Darker areas indicate lower values of r. The values for the parameters used can be seen in Appendix 2

Given that Figure 3.3 suggests a strong selection pressure to alter the cytokine profile in response to the kind of demographic change that will result from reduced exposure to fatal infectious disease, we asked how fast the genetic determinants of innate immunity might alter in a population that underwent rapid change from Gambian to Danish mortality pressures. For simplicity we assumed here that η is determined by a single gene with two alleles, such that the heterozygote would have the value $\eta = 1$ and that the two homozygotes would have $\eta = 0.1$ and $\eta = 10$, respectively. Denoting the two alleles of the gene responsible for determining η as G and g, so that GG homozygotes have $\eta = 10$ while gg homozygotes have $\eta = 0.1$, we used the calculated fitness (r) values to predict the changes in genotype and allele frequencies in future generations of a population which started with the genotype distribution that was optimal under Gambian conditions but which now existed with the selection pressures imposed by Danish conditions (Table 3.1). It can be seen that the frequency of the g allele within the population declined to very low levels within about 6 generations, consistent with the fact that in the contour plot in Figure 3.3b the optimum zone includes the right-hand margin of the parameter space.

3.4 Discussion

Our immunogenetic model establishes the evolutionary plausibility of a trade-off between human survival and fertility mediated by selection on genes that determinate innate immunity. Furthermore, the model shows how within the context of the disposable theory, and given the ecological factors that affect mankind, this immunogenetic trade-off, which represents a clear example of the kind of pleiotropic

	Freq. of genotypes			Freq. of alleles	
Generation	GG	Gg	gg	G	g
1	0.315333	0.492424	0.192243	0.561545	0.438455
2	0.083951	0.411584	0.504465	0.289743	0.710257
3	0.023263	0.258516	0.718222	0.152520	0.847480
4	0.006579	0.149069	0.844351	0.081114	0.918886
5	0.001881	0.082980	0.915139	0.043371	0.956629
6	0.000541	0.045431	0.954028	0.023256	0.976744
7	0.000156	0.024667	0.975177	0.012490	0.987510
8	0.000045	0.013336	0.986619	0.006713	0.993287
9	0.000013	0.007193	0.992794	0.003610	0.996390

Table 3.1. Predicted changes in genotype and allele frequencies when a population that has evolved to optimise the immunogenetic trade-off under Gambian mortality conditions is transferred to Danish mortality conditions.

gene action envisaged by Williams (1957), seems critical for human longevity. The model shows how the advantage of increased resistance to fatal infectious disease can be offset by the increased risk of spontaneous abortion to produce an optimal zone of gene frequencies where these effects are balanced. Although in the model we have assumed that the genetic basis of innate immunity is simple, it is sufficient to describe the evolutionary implications of the Th2/Th1 balance and illustrate the effect of trade-offs on longevity and reproduction. A more realistic model might allow for a larger number of genes, with possible epistasis among them. However, in order to develop such a model, the methodological and mathematical difficulties would be enormously increased. In a model with a larger number of responsible genes it is likely that individual gene effects would be smaller than has been assumed here, which might predict a somewhat slower genetic change than is shown in Table 3.1. It is also worth noting, that when Figures 3.3a and 3.3b are compared with regard to the optimal values of s, the optimum in the case of Denmark is slightly higher (0.542) than that in Gambia (0.526). Such a difference is consistent with the predictions of the disposable soma theory since increased survival requires a higher investment in maintenance and repair.

The immunogenetic model described in this work confirms the plausibility of the hypothesis that the relatively high prevalence of genetic factors that impair fertility in present-day populations may be the legacy of a trade-off that was of great importance until quite recently. If this is true, it implies that the rapid reduction in exposure to the risk of fatal infectious disease that has occurred over the last century might be responsible for altering the selection forces on major determinants of the human life history and that a process of adaptation in developed countries is likely to be still ongoing.

4 Apolipoprotein E as an Example of the Use of Epidemiological Data in an Evolutionary Model

4.1 Introduction

According to the British Heart Foundation diseases of the heart and circulatory system (cardiovascular diseases (CVD)) accounted for more than 235,000 deaths in the UK during 2000, with half of them due to coronary heart disease (CHD) and a quarter from stroke. At the same time CVD is the second most common (after musculoskeletal conditions) cause of longstanding illness encountered in the population. CHD is the main form of heart disease and the single most common cause of death in the UK, killing one in four men and one in six women (Petersen, 2002).

From the above, it is clear that genes affecting CHD and CVD are extremely important to human longevity and successful ageing. The first such gene to be discovered was the apolipoprotein E (*apo E*) gene situated on human chromosome 19 at locus 19q13.2, a 3.7 kb long gene containing four exons and coding for a 317 amino-acid polypeptide that after cleavage gives rise to a 299 amino-acid long mature protein (Hagberg et al., 2000). Apolipoprotein E (APO E) is a member of a diverse family of proteins collectively known as apolipoproteins, carrier proteins specializing in lipoprotein particle formation, secretion, transport, binding and metabolism (van Bockxmeer, 1994). Another member of the apolipoprotein family closely linked with Apo E on chromosome 19 is APO C2, a necessary cofactor for the activation of lipoprotein lipase, and the cell surface receptor for LDL that plays a significant role in cholesterol homeostasis LDLR (Online Mendelian Inheritance in Man). APO E is

synthesized in many different areas of the body such as the liver, brain (primarily astrocytes), skin, macrophages and steroidogenic organs. The plasma pool is derived mainly from the liver while the central nervous system pool is produced locally (Smith, 2002). In plasma, APO E is a component of triglyceride-rich chylomicrons, very-low-density lipoprotein (VLDL) particles and their remnants, and high density lipoproteins (HDL), while it also serves as a ligand for the low density lipoprotein (LDL) receptor family and binds heparan sulphate proteoglycans during lipid metabolism (Hagberg et al., 2000; Shih et al., 2000).

Apolipoprotein E has three major and more than thirty minor isoforms usually linked with diseases. The three common alleles are epsilon (ε) 2, ε3 and ε4 producing three homozygous (ε2/ε2, ε3/ε3 and ε4/ε4) and three heterozygous (ε2/ε3, ε3/ε4 and ε2/ε4) genotypes (Rall et al., 1982). The difference between them lies at two amino acid residues 112 and 158; with the most common allele ε3 having cysteine and arginine respectively while ε2 has cysteine and ε4 arginine at both of them (Hagberg et al., 2000).

What makes *apo E* such an interesting topic for longevity studies is the fact that the carriers of different alleles show differences in the incidence of coronary artery disease (CAD), peripheral atherosclerosis, Alzheimer disease, possibly stroke and even ability to recover from trauma (Smith, 2002). The underlying mechanism for the action of the gene involves an interdomain interaction between the amino- and carboxyl- terminals of the protein producing isoform-specific lipoprotein preferences such that E2 and E3 protein isoforms bind preferably to HDL (the "good" cholesterol), while the E4 isoform shows a preference for VLDL (Dong et al., 1994). The APO E2 isoform is defective in binding the LDL receptor although it retains its ability to interact with LDL-receptor-related protein and other related receptors

(Smith, 2002). A small number of ε2 homozygotes suffer from a condition known as type III hyperlipoproteinemia, characterized by accumulation of cholesterol-rich remnant lipoproteins due to incomplete catabolism of chylomicrons and VLDL leading to premature atherosclerosis. The fact that although 90% of the patients are ε2 homozygotes, but only 5% of the ε2 homozygotes suffer from the condition, suggests that further factors are required for the phenotypic expression of the condition (Marz et al., 1998).

4.2 Apolipoprotein and global distribution

The global frequencies of the three common alleles vary significantly between populations with different ethnic backgrounds, with ε2 almost absent from the Inuit of Ammassalik, Mayans of Mexico, Yanomame of Brazil and Amerindians in general and Australian aborigines, while on the other hand ε2 is relatively common in Chinese, Malaysians and Papua New Guinea aborigines (Gerdes et al., 1996b). In general ε3 is globally the most common allele ranging from a frequency of 0.536 in African Pigmies to 0.911 in Mayans. The rarest of the three alleles is ε2 with a fluctuating frequency between 0.145 to just 0.02. The frequency of ε4 varies from 0.052 in Sardinians to 0.407 in Pigmies and shows a negative correlation to ε3 allele (Fullerton et al., 2000). In Europe, and between populations of Caucasian descent, there is a north-to-south gradient of decreasing ε4 frequency opposite to the spread of agriculture (Gerdes et al., 1996b; Corbo and Scacchi, 1999).

Generally in Caucasian populations it is believed that the carriers of the ε2 allele are more likely, and the ε4 carriers less likely, to reach old age than the ε3

homozygotes. The common study strategy has been to compare gene frequencies between a young and an old group from the same population and check for any significant differences of the candidate gene between the two. Although this kind of study has provided significant insight into the way that genes affect longevity, it must be treated with caution. Small groups, especially those for the old, and low representation of the uncommon genotype, in this case ε2, can lead to a poor statistical calculation of the true allelic effects. Also we should be careful when we suppose that the same selective pressures have acted on both groups, since the quality of life, technology and medical science have changed significantly during the last 100 years. The assumption that the penalizing effect of the candidate gene is not apparent in the young group can be misleading since the precise age of gene action and the presence of any pleiotropic effects are not fully studied yet. Finally the effects of linked genes and the composition of the groups should be carefully considered and checked, especially when healthy individuals are used, which are not always representative of the general population or the gene action, even more so when they include individuals from different ethnic backgrounds.

4.3 The role of apolipoprotein E in CVD

Using a group of 300 French centenarians and a group of 160 adults with between the ages of 20 and 70, Schachter et al (1994) found a significant decrease in the frequency of the ε4 allele among centenarians while the ε2 allele was significantly increased. Although the use of an adult control population of up to 70 years of age is expected to decrease the overall significance of the gene effect, since it would have already started producing *apo* E related mortality, the differences observed were large

(from 11.2% in control to 5.2% in centenarians for ε4, and from 6.8% to 12.8% for ε2). In an enlarged study (600 centenarians) of the same cohort six years later, Blanche et al managed to confirm the original findings (Blanche et al., 2001). Participants of the Framingham offspring study were used for a cross-sectional prevalence and cohort study to describe the association between *apo* E alleles and CHD (Wilson et al., 1994; Lahoz et al., 2001). It was observed that an elevated risk of CHD was associated with a gender specific effect of *apo E* ε4 allele.

But not all studies showed a relationship between the *apo E* genotypes and risk of death. Galinsky et al (1997) did not found a significant difference in the frequency of the *apo E* gene between the young and old groups concluding that it may be too simplistic to associate *apo E* isoforms with increasing risks for early death. Kuusisto et al (1995) using a non-diabetic Finish cohort of 1067 subjects between 64 and 74 years of age concluded that the *apo* ε4 allele cannot be regarded as an important risk factor in the elderly. Instead they proposed that the ε4 allele exerts its increased risk during middle age. The exclusion of subjects "too ill to participate", and the small number of subjects homozygous for the ε2 and ε4 alleles, could have played a significant role in dampening the effect of the gene. In contrast, a study in Sweden (Corder et al., 1996) reported that selective survival involving the *apo E* polymorphism was important in the oldest old (>85 years) with good cognition, but had no effect in survival of younger individuals.

In a meta-analysis of 10 studies Wilson et al (1996) reported an association between the presence of the ε4 allele and CHD for both sexes, but they could not see a significant association between the ε2 and ε3 alleles and CHD. A longitudinal study of elderly Finish men in two cohorts, one in Eastern and the other in South-western Finland, reported greater odds of CHD for different genotypes at each site but this risk

was generally linked with the presence of ε4. They concluded that *apo E* genotypes confer risk information dependent on the mortality history of the population, its environmental factors and its genetic background influencing the risk of CHD (Stengard et al., 1996). Gerdes et al (2000) using a relative mortality risk and the ε3ε3 group as a reference, found that there are very modest differences between the relative risks of the genotypes compared with what is seen when odds ratio is used. For the *apo E* alleles they estimated that ε2 carriers have a lower and ε4 carriers have a higher risk of mortality compared with the ε3ε3 group, concluding that *apo E* is a "frailty gene" and not a "longevity gene". In a study using a group of 179 Finnish centenarians and comparing them within the group Frisoni et al (2001) found that ε2 allele increases with increasing age above 100 years of age, concluding that the ε2 carriers might be predisposed to reach extremely old age. Finally in a study in Belfast using subjects of the MONICA study (MONItoring of CArdiovascular trends) the age-associated decrease of the ε4 allele was further supported and evidence was also found for the link of the ε2 allele with longevity (Rea et al., 2001).

Up to now the studies mentioned were done almost exclusively on Caucasian populations, and as we mentioned before there is a great amount of variation between the frequencies of the three common alleles in populations with different ethnic origins. In a study of Japanese ischemic heart disease (IHD) patients, it was found that ε4 allele was an independent risk factor for silent myocardial ischemia without showing any association to the lipid profile, while the ε2 allele was consistent with its role in lowering LDL (Nakata et al., 1996). Gerdes et al (1996b) found that in a Greenland Inuit population, which exhibits a high frequency of the ε4 allele, carriers of this allele showed higher mean plasma lipoprotein-related variables. In a cohort study in Chinese Han *apo E* also proved to be an important factor for longevity of

individuals >85 years of age, with the ε4 allele linked to higher mortality. To examine the effect of *apo E* polymorphism in different ethnic groups Lee et al (2001) studied the *apo E* associated mortality risk in Caucasians, Hispanics, and African-Americans living in northern Manhattan, New York. They observed that the ε2 allele seemed to reduce mortality risk for Caucasians and Hispanics but not for the African-Americans. The ε4 allele was not associated with higher risk in any of the groups. On the contrary, it seemed to lower the mortality risk for the African-American group. Finally Stengard et al (1998), using nine population-based samples of middle-aged males from the WHO Monica project, found that around 50% of interpopulation variation in average serum total cholesterol level and approximately 40% of the variation in CHD mortality rates could be explained by the relative frequency of the ε4 allele. The fact there is so much difference between the *apo E* allele frequencies between different populations, despite the same effects of this alleles on lipoprotein profiles, suggests that either there is a strong component of genetic background that can deal with the extra lipids in serum, or that there are differences in the selection pressures applied in the past of each ethnic group.

4.4 Role of apolipoprotein E in other disorders

Apolipoprotein E has also been implicated in other age dependent conditions such as Alzheimer's disease (AD), which is the most common form of senile dementia (Smith, 2002). Genetics, lifestyle and environment are all involved in the disease. AD changes are similar in some degree to the "normal" changes during ageing in human brain (Finch and Sapolsky, 1999). The main pathologic traits by which AD is defined are: senile plaques containing the amyloid β-peptide (Aβ) in the

form of highly aggregated and extracellular amyloid; neurons containing neurofibrillary tangles; and regional neuron loss in the hippocampus and entorhinal cortex (Finch and Sapolsky, 1999). The ε4 allele of *apo E* has been shown in a number of studies to be associated with both familial and sporadic forms of the disease, causing higher incidence and earlier age of onset, and affecting its pathology and rate of progression (reviewed in Roses, 1997). Although there are still no established mechanisms to account for the link between AD and *apo E* the majority of hypothesis can be summarized in three main categories. The first proposed mechanism is through the effect of APO E on Aβ production, aggregation and catabolism, with APO E being co-localized with Aβ in senile plaques with ε4 having an impact on the amyloid cellular uptake. Secondly APO E may have a direct effect on neurons or glial cells, including tau phosphorylation, neuronal survival, neurite extension, and on the allele-specific protective effects on neurotoxicity and oxidative cytotoxicity, with ε2 being the most protective. Finally, the third proposed mechanism is that the link between *apo E* and AD is due to changes in cerebral blood flow related to atherosclerosis, something that has evidence both for and against from various studies (reviewed in Smith, 2002).

Taking into account the effects that APO E has within brain tissue, it is easy to understand that it may affect a lot more of brain function and pathology than just AD. Normal age-associated cognitive decline, and mild cognitive impairment (MCI) in particular, have been repeatedly associated with the presence of the ε4 allele. MCI is defined as isolated memory impairment in an otherwise healthy individual. DeCarli et al (2001) found that the presence of the ε4 allele increased the risk of MCI and the risk of MCI conversion to AD. To test if *apo E* is a factor modifying general cognitive ability the genotype of 202 children was screened and compared with their

IQ scores; no association was found between the *apo E* isoforms or the level of its expression and cognitive ability (Turic et al., 2001), while in a second study with 134 young females (19-21 years old), ε4 was associated with only a modest increase in IQ (Yu et al., 2000) suggesting that the cognitive effect of the alleles is either irrelevant in a healthy population or that the alleles produce their effects during middle age.

Furthermore *apo E* has also been linked with response to head injury, where patients with the ε4 allele have a worse outcome after traumatic head injury. In a study of 93 patients with head injury it was found that 57% of the ε4 carriers resulted in death, vegetative stage or severe disability, compared with 27% of non carriers (Teasdale et al., 1997). To assess the association between chronic traumatic brain injury and *apo E* genotype Jordan et al (1997) studied 30 professional boxers and concluded that possession of the ε4 allele may be linked with increased severity of chronic neurological deficits.

In addition to the association of *apo E* with AD, cognitive decline and head injury, various other studies have found associations with a number of different brain pathologies. Both ε4 and ε2 were linked to cerebral amyloid angiopathy-related hemorrhage, irrespective of AD pathology, with individuals carrying the ε2ε4 genotype being at risk at a particularly young age. As with brain injury, ε4 was associated with adverse outcome from intracerebral hemorrhage. Less certain is the relationship of *apo E* with stroke where it is generally believed that ε4 may have a small effect of increasing the risk for an ischemic stroke but has no effect on the outcome after the event (reviewed in Horsburgh et al., 2000).

It is easy to see the importance of the *apo E* gene in longevity and health of the elderly. Its well documented association with CHD and AD and its emerging

importance in other age related pathologies makes it a very good candidate for the title of first human "ageing" gene, if something like that exists at all.

4.5 *apo E* evolution

Up to now we have seen the function of *apo E* and its importance in longevity and health, but we have avoided the important questions of how and why *apo E* polymorphism has evolved. One of the major ideas of the evolution of senescence is the mutation accumulation theory, which states that alleles with deleterious effects restricted to older ages can achieve higher mutation selection balance than those with such effects early in life (Medawar, 1952). It is widely believed that the *apo E* gene belongs to the category of genes acting at a sufficiently late age that it is immune to direct natural selection.

To consider the evolution of *apo E* we should start from its origin. Although the ε3 allele is the most common allele worldwide and both ε4 and ε2 alleles can be derived from it by a single base change, ε3 is not widely considered to be the ancestral human *apo E* allele (Mahley and Rall, 1999). In contrast to the human three allele system, non-human primates carry a single variant that resembles the human ε4 allele (Hanlon and Rubinsztein, 1995). This ε4-like isoform of the non-human primates (and other mammals in general) does not have the same lipoprotein preference as the human counterpart. Another residue of APO E that needs to be considered is residue 61. By changing the arginine at residue 61 to threonine (as in the rhesus and squirrel monkey) the lipoprotein specificity of the allele will change from VLDL to HDL, giving ε4 an ε3-like preference (Mahley and Rall, 1999). Finch and Sapolsky (1999) suggested that the evolution of *apo E* in primates went through

two stages, first the 61 residue was changed from a threonine to an arginine in early hominids, and then the substitution of cysteine 112 to arginine gave rise to the human ε3 allele. A final change of the 158 arginine to a cysteine is responsible for the production of the ε2 allele, which probably arose multiple times since the base pair transition responsible is common, due to the high frequency of methylation of cytosines within CpG dinucleotides such as those present in the codons involved (Hanlon and Rubinsztein, 1995; Finch and Sapolsky, 1999; Mahley and Rall, 1999).

The worldwide frequency of ε3 relative to the ancestral allele ε4 suggests that either there is a strong negative selection against the latter or a positive selection for the former, or even both at the same time. Since the diseases associated with *apo E* are evident usually after the end of the reproductive lifespan it was thought unlikely that they could play a role in human evolution (Hanlon and Rubinsztein, 1995). To account for the evolution of the gene a number of different suggestions were made based on the dietary changes during human evolution, APO E's role in neurological development, head injury and immunoregulation, or simply that the allele frequencies observed today are just due to chance and genetic drift (Hanlon and Rubinsztein, 1995). In a study of 585 Danish men Gerdes et al (1996a) found a significant difference in the number of offspring between men with different *apo E* genotypes, with ε3ε3 genotype carriers showing the highest number of children while the ε4 carriers showed the lowest. The lack of any known mechanism to link APO E and lipids to fertility and the fact that many other factors, that can or cannot affect the number of children through the action of *apo E*, were not considered, suggest that the results of the study might be due to reasons other than the association of *apo E* with male fertility.

Sapolsky and Finch (2000) put forward an interesting idea for the evolution of ε2 and selection against the ε4 based on these alleles' role in AD, proposing that the *apo E* ε2 isoform is the most recently evolved as a highly adaptive subtype increasing the probability of its female carriers to achieve higher fitness due to grandmothering. The "grandmother hypothesis", initially put forward to explain the long post-reproductive lifespan in women based on the grandmother-grandchild food sharing in hunter-gatherers (Hawkes et al., 1998), gives a possible insight into *apo E* evolution by the association of APO E, AD and oestrogen. Oestrogen treatment is considered to delay or even prevent AD in post-menopausal women (Tang et al., 1996) and it may be a signal linking female fertility and the onset of the disease (Breitner and Miech, 1999; Sapolsky and Finch, 2000). Thus post-reproductive women with the ε2 and ε3 alleles will have been more likely to contribute to the gathering and sharing of food, thus increasing the chances of their genes spreading in the population. Finch and Sapolsky (1999) also mention that in autopsy studies on baboons and monkeys some degree of atherosclerotic heart change was observed, but they still use the "grandmothering" hypothesis as the most likely factor affecting the selection of the *apo E* isoforms, although they recognise the possibility that the three alleles might have been established before the evolution of this social function.

A more interesting point in the story of the evolution of *apo E* is the presence of the ε4 allele. Since it is clearly disadvantageous, how can it be retained in the population? Martin (1999) proposed that the ε4 allele could be advantageous in cases of infections with pathogens requiring host lipids for survival. Parasites like *Trypanosome brucei* that cannot carry out de novo liposynthesis may depend upon host LDL for acquisition of lipids. So the high affinity of ε4 for some lipids might actually hinder their uptake by the pathogen.

Although CHD is a disease that is evident mainly after the end of reproductive age in humans the British Heart Foundation statistics show that there are on average 40 deaths per 100,000 individuals under the age of 65 every year (Petersen, 2002). But death is not the only result of CHD and atherosclerosis, morbidity is also important in an evolutionary context since it will lower the overall fitness of the individual and its offspring. Furthermore vascular disease and atherosclerosis may hinder reproduction directly by changing the sexual ability of the individual. A strong association is believed to exist between male impotence and vascular diseases (Bortolotti et al., 1997) with arteriosclerosis being an important factor (Jensen et al., 1999). Erectile dysfunction (ED) is a problem common mostly in middle age men and can be considered as analogous to CHD, arising from progressive blockage of small vessels and reduced arterial compliance. Importantly ED and CHD share the same risk factors leading to atherosclerosis (Feldman et al., 2000), and although there is no documented association between ED and *apo E* we suggest that it is very likely to exist.

Our hypothesis is that the increased energetic requirements of humans during pregnancy, lactation and early development, imposed by our large brain size, drove us towards the consumption of a higher energy diet and especially meat-eating. Since the effects of the three alleles are more pronounced while on a fatty diet, we propose that the acquisition of energy from meat provided the selective advantage towards the ε2 and ε3 genotypes that carry the lower penalties for high lipid consumption. We test our hypothesis with the help of a simulated human population characterised by a number of risk factors for CVD.

4.6 Methodology

Based on our previous discussion on *apo E* we will simulate its effects on a human population and use our model to investigate the selection and evolution of the three *apo E* alleles. The model will simulate the population directly instead of relying on use of the Euler-Lotka equation as in Chapters 2 and 3. Direct simulation avoids the assumptions of stable population structure needed for the Euler-Lotka equation to be valid. Moreover, having a model that made use of equations that statistically summarize the population would lead us to lose the contribution of the uncommon events, unless we make assumptions about how uncommon they are and how they affect the entire population. With the use of a direct simulation we can avoid these problems as long as the forces exerted on the population are known and the number of repetitions is large enough to give a statistical validity to our population. Additionally, direct simulation can provide valuable insight into epidemiological studies, which by definition use a relatively small number of individuals, by providing information on the frequency of the observations. Thus although we are unable to test 1,000,000 individuals we can test 1,000 and simulate the rest. We can then test how often our observations can arise in multiple sets of studies each with 1,000 individuals.

Before we describe the program used for the simulation we will explain the framework of the model as well as the difficulties encountered during its construction. The first step is to decide which are the relevant factors affecting the action of *apo E* on longevity. There are two kinds of risk factors associated with cardiovascular diseases. The first kind is the non-modifiable risk factors, of which the more important are heredity, in this case *apo E* genotype, and gender. The second kind is the modifiable risk factors. According to the British Heart Foundation the main

modifiable risk factors for CHD are: smoking, unhealthy diet, lack of physical activity, high alcohol consumption, poor psychosocial well–being, raised blood pressure, raised blood cholesterol, obesity and diabetes (Petersen, 2002). In section 4.6.1 we will review the evidence for the association of each risk factor with CVD and the action of *apo E*.

After identifying the relevant factors affecting the action of apo E we needed a measure of risk to describe their effect on longevity. The most straightforward and convenient measure of risk when the basic Gompertz model is used to account for the mortality of the population, is the relative risk as will be defined in section 4.6.2. However, as we shall see later, epidemiological studies usually express the increase or decrease of risk in terms of the odds ratio. Although both measures of risk describe an elevated or lowered probability of an event taking place within a group relative to another group, the change of an odds ratio to a relative risk is not a simple problem and we will address this in section 4.6.2.

Another complication during the construction of the model arose due to the lack of quantitative data for the gene × environment interaction between apo E and the risk factors used. This has led us towards the use of a novel approach of estimating these interactions based on independent measurements of risk for each factor and the qualitative form of the relationship between the gene and the environment. A number of assumptions are needed for such a view of the gene × environment interaction to work which will be presented in detail in section 4.6.3.

A basic aspect of our simulation program is the use of a Gumbel distribution to represent the mortality schedule of the individual. In section 4.6.3 we will explain what a Gumbel distribution is and its biological significance. For now it will suffice to say that we used the specific statistical distribution due to its close fit to a modified

version of the Gompertz model describing the change of mortality with age and , in this case, also the probability of death per year of age. The modified Gompertz curve and the Gumbel distribution were used throughout our model to account for the effect of risk factors on the mean lifespan of the simulated population and to randomly choose an age of death for the individual as we shall see later.

Another important aspect of our model is that it has a large stochastic component. That means that every time the program was run a different result was obtained. Each single run is only one of the almost infinite possibilities of the model, but a group of runs statistically summarized provides an indication of how the population is expected to evolve. To achieve a level of stochasticity mirroring nature we made extensive use of random numbers. In section 4.6.5 the basic theory behind random number generators and information on the algorithm used are given in detail. Finally, in section 4.6.6 we present the procedure followed to construct the simulated population and we describe in detail the method of random walk used to make the distinction between stochastic change and change due to selection.

Before we start to present in detail the different parts of our model we should note that in this part only the construction of a general form of the model, also called the default version, is presented. To fully explore the significance of the different factors affecting the action of *apo E* on longevity we use a number of different versions of the program with the same basic characteristics as the basic model but with quite different parameters values. These versions will be described and tested in the part 4.7.

4.6.1 Risk factors

As we saw earlier the first step for the development of our simulation is to identify the relevant CVD risk factors affecting the action of *apo E* on longevity and disease. Here we will review the evidence supporting our choice of two non-modifiable and four modifiable risk factors. For the first category of non-modifiable factors, genotype is the obvious choice since we are interested in apolipoprotein E, while gender is a significant modifier of mortality as we will see, especially in the case of cardiovascular disease. In the case of modifiable risk factors the choice was based on current knowledge of lifestyle parameters affecting CVD while keeping in mind the need for independence between the risk factors, simplicity and availability of data. Consequently, obesity and diabetes, although well known factors affecting the incidence of the disease, were not included as non-independent of the rest CVD risks factors, while psychological well being lacks the simplicity of association with the disease and quantitative data are extremely difficult to obtain. On the other hand, diet and physical activity are parameters that have played very important roles in our recent evolutionary history and have contributed both to our great success as species and to the contemporary prevalence of CHD. Smoking and alcohol consumption have emerged as significant factors in the last two centuries and are likely to affect CVD risk in the future, especially in developing countries. Moreover, as we shall presently see, a large number of studies have investigated the association of diet, alcohol, smoking and exercise with CVD and *apo E*.

4.6.1.1 Non-modifiable risks factors.

Genotype

The first risk parameter that we will consider is the *apo E* genotype. The importance of the *apo E* genotype and the possible mechanisms that can affect CHD and CVD were reviewed earlier. We should note here that we assume that the gene generates a risk factor only in conjunction with an environmental parameter. The gene function is to transport lipids through the bloodstream, when these are kept in their minimum amount we consider that all genotypes transport the essential quantities and no risk is generated. Thus, there are no risk factors for the different alleles but for the interaction between a specific allele and each environmental parameter. That is a concept that we will further explain in section 4.6.3 where the method used to find the genotype × environment interaction will be described in detail.

Gender

The second risk factor that should be taken into account is gender. It is well documented that women suffer less from CHD-related mortality than men (Lloyd-Jones et al., 2002; Panagiotakos et al., 2002; Petersen, 2002). Notably, women show a lag of 5 to 15 years in coronary mortality and morbidity relative to men (Panagiotakos et al., 2002). Although the precise reasons for that difference are not yet known it has been thought to be due to the protective role of female hormones, with observational trials of women undergoing hormone replacement therapy supporting this notion (Humphries and Gill, 2003). Recently, randomized trials showed that there is no protective effect of hormone therapy for CVD, casting serious

doubts on theories of their role in generating the difference of cardiovascular risk between men and women (Humphries and Gill, 2003; Lokkegaard et al., 2003). We will discuss the relative risk between men and women later on in section 4.6.2 together with the methodology to change odds ratios to relative risk.

4.6.1.2 Modifiable risk factors

Diet

The first modifiable risk factor that we will address is diet. Diet is a well known modifier of cardiovascular risk and extensive information has been accumulated on the different ways that CHD is affected by our nutrient and calorie intake. The currently accepted view is that long-chain saturated fatty acids, such as palmitic acid found in meat and dairy fat, increase LDL-cholesterol. When these are replaced by carbohydrates both HDL and LDL are decreased keeping their ratio constant. But the common form of carbohydrates in the Western diet contains refined polysaccharides with high glycemic index, increasing serum glucose, insulin and fasting triglyceride levels, another risk factor for CHD. Trans-unsaturated fats produced by hydrogenated vegetable oil, formed from their change to solid or semi-solid form for frying or baking, increase LDL while decreasing HDL. The most favourable diet change though, is the replacement of saturated fat with monosaturated or polyunsaturated fats, lowering LDL significantly, while HDL is decreased only slightly, and without any parallel increase in triglyceride levels, especially when the very long chain n-3 polyunsaturated fat found in fish is used (Sacks and Katan, 2002).

As we mentioned earlier, the model assumes that *apo E* is a risk factor when combined with a lifestyle parameter, in this case diet, and that the effects are no different between men and women. Although there is evidence for higher sensitivity

of women to triglyceride levels and lower sensitivity to LDL, the assumption is a reasonable approximation for the purposes of the model (Sacks and Katan, 2002). In the case of diet, which is believed to be the fundamental risk in the evolutionary history of CVD, the *apo E* genotype is considered in relation to nutrient intake to produce the diet relative risk. Evidence for the importance of diet and the change it can bring can be seen in a recent study of free-ranging baboons. Animals living close to tourist facilities, coming into contact with an easily accessible and energetically dense food source, showed two- to three-fold increases in serum insulin and very high HDL, VLDL and LDL cholesterol compared with wild foraging baboons (Kemnitz et al., 2002). We consider that in the absence of any lipid intake or excess energy, there is no distinction between the six genotypes, but as the fats and energy intake increases there is a genotype-specific risk associated with the dietary change. According to this, we expect that the same amount of lipid change will have a bigger effect on LDL for the ε4 carriers, both for the increase and decrease of lipid intake. There is a number of studies reporting that ε4 carriers have a greater plasma lipid response during dietary intervention while others failed to find any significant change (for a complete summary see Ordovas, 1999) with Weggemans et al (2001) concluding that the effect was too small for a successful therapeutic approach. Ordovas et al (1995) summarised the available literature for the gene-environment interaction for lipid response to diet concluding that the *apo E* ε4 allele carriers are more responsive to diet than the other two allele carriers and that the effect is likely be more intense when the total amount of fat is changed, a conclusion also shared by Hagberg et al (2000).

In order to quantify the effect of diet in terms of risk factors we first need to recognise the different possible categories of dietary patterns. Using cluster analysis, Millen et al (2001) described 5 distinct dietary patterns with unique food behaviours

and nutrient intake profiles. The dietary data collected among 1,828 women participating in the Framingham study revealed the presence of five clusters: Heart Healthy, Light Eating, Wine and Moderate Eating, High Fat, and Empty Calorie. Although the cluster names are self-explanatory we will describe them in a bit more detail. The Heart Healthy group was characterized by high consumption of foods typically recommended for healthy diets, as fruits and vegetables. The Light Eating group had an overall low food intake. The Wine and Moderate Eating group was similar with moderate intake of fats and sweets and high wine consumption. The last two groups of High Fat and Empty Calorie were characterized by fewer servings of healthy nutrient dense foods, micronutrients and fibre with the first group showing high fat consumption, while the later ate more sweets (Millen et al., 2001; 2002). In a later paper using the same dietary clusters, Millen et al (2002) calculated the risk factors of carotid atherosclerosis disease corresponding to each cluster. As was the case for the gender risk factor described earlier, the diet-associated risk was expressed as odds ratio instead of relative risk, thus we will deal with their calculation in section 4.6.2.

Alcohol

There is little doubt that alcohol consumption has an effect on the prevalence of CHD. A number of case control and cohort studies have demonstrated the presence of a U- or J- shaped curve relating alcohol intake and CHD risk, and alcohol and all-cause mortality (Corella et al., 2001). It is widely believed that moderate alcohol intake has a protective effect which diminishes with increasing dose after the point of heavy consumption. The levels of moderate and heavy alcohol intake are difficult to define, as well as the levels of desirable alcohol consumption in a

population. Usually the distinction is made at the level where alcohol stops having a protective effect. The UK Department of Health currently suggests that 3 to 4 units[7] (26 - 34g) per day for males and 2 to 3 units (17 - 26g) per day for women of all ages is "sensible", while higher doses carry progressive health risks (Department of Health and Inter-Departmental Working Group, 1995). Other studies found that consumption of up to 6 units per day is beneficial (Bovet and Paccaud, 2001) while most agree that more than 9 units of alcohol per day is heavy intake and should be avoided (Marmot, 2001a). Alcohol is associated with a wide range of medical and social problems, and in some groups, especially those with low CHD risk, all-cause mortality increases with intake even below 9 units per day, mainly due to accidents and violence which increase with ethanol consumption (Bovet and Paccaud, 2001). The pattern of alcohol intake has proven to be of almost equal importance to quantity. Episodic drinking sessions that reach the weekly intake of a moderate drinker can have the same detrimental effects as heavy intake (Marmot, 2001b). Binge drinking may be the reason for the absence of protective effects of alcohol in Northern European countries (Finland, Scotland) and may play a significant role in the current rise in mortality seen in Eastern Europe (Bovet and Paccaud, 2001). Another important aspect of alcohol intake is whether the lower risk is a result of the beverage or of the alcohol it contains. Studies have shown that although there are beverages containing bioactive compounds able to lower CHD risk, such as flavonoids and resveratrol in red wine (Kris-Etherton et al., 2002), the protective role of moderate alcohol consumption is independent of their action (Marmot, 2001b).

The biochemical mechanisms through which moderate alcohol consumption reduces the risk of CHD are not yet fully elucidated, but the main factors are believed

[7] One unit of alcohol in the UK is equal to 10ml or 8.7 g of absolute alcohol (density 0.87g/mL).

to be: the concentration of LDL and HDL in blood, the ethanol involvement in endothelial function, the platelet function in blood clot formation and their enzymatic dissolution (Zakhari, 1997). In terms of the effect of alcohol on blood lipoproteins, it is well established that there is an increase of HDL that seems to account for up to 50% of the lowered CHD risk (Rimm, 2001). This may be attributed to the increased hepatic production of apolipoproteins and lipoproteins, the increased lipoprotein lipase concentration and the decreased removal of HDL cholesterol as a result of alcohol (Rimm et al., 1999). Platelets are a key factor in thrombosis which can contribute to myocardial infarction or stroke. Alcohol's antithrombotic effect is related to granule secretion and fusion and fibrinogen activity (Zakhari, 1997). Furthermore moderate drinking affects endothelial function by increasing nitric oxide release, alcohol dehydrogenase activity in the blood vessel wall antagonizing lipoprotein oxidation, and may also be involved in prevention of inflammation due to fatty streaks in the arteries (Zakhari, 1997; Puddey et al., 2001).

While moderate alcohol consumption confers protection from atherosclerosis, heart attacks and stroke, heavy consumption not only increases other forms of mortality, but it also increases the risk of cardiomyopathy, cardiac arrhythmia, hypertension and hemorrhagic stroke (Zakhari, 1997). Alcoholic cardiomyopathy is a heart muscle disease seen most often in alcoholics in their late 40s. It is characterized by increased myocardial mass, dilation of the ventricles and wall thinning. It is the result of myocyte dysfunction due to high ethanol intake which causes changes in mitochondrial, endoplasmic reticulum and contractile protein function and disturbs calcium homeostasis (Piano, 2002). Alcohol-induced arrhythmias are considered as the basic factor of sudden alcohol-related coronary death especially in subjects with previous CAD history. These arrhythmias were first observed at weekends and

holidays ("holiday heart syndrome") and are related with high alcohol intake that causes atrial fibrillation (loss of synchrony between the atria and the ventricles) even in people with otherwise healthy hearts (Puddey et al., 1999a). Finally, the effect of alcohol in increasing blood pressure is well documented in both men and women. Consumption of more than 15 units/week in women increased the prevalence of hypertension (Nanchahal et al., 2000) while in studies where alcohol intake was lowered to moderate levels, blood pressure decreased significantly after just 4 days (Zakhari, 1997). Alcohol-related changes in the sympathetic nervous system, vasopressin release, the rennin-aldosterone system and atrial natriuretic peptide may play an important role in increasing the blood pressure of heavy drinkers (Kauma et al., 1998).

The genotype-environment interaction between *apo E* and alcohol is still a point of debate. Kauma et al (1998) reported a significant increase of blood pressure associated with *apo E* in men with moderate and heavy alcohol consumption, while no effects of the genotype were observed in men "abstainers" and women. Although the ε4 allele is linked with high LDL cholesterol, Kauma et al (1998) surprisingly found that subjects with the ε2 allele were more susceptible to the blood pressure increasing effect of alcohol compared to the ε4 carriers. In a reply to this article Puddey et al (1999b) questioned the validity of the observations and reviewed their own work, which showed that blood pressure changes due to alcohol were not affected by the *apo E* genotype. To examine whether *apo E* variation modulates the association between LDL cholesterol and alcohol, Corella et al (2001) used a healthy population sample from the Framingham Offspring Study. For the male participants, they were unable to find any difference of LDL with genotype for non-drinking men while a significant difference was observed for drinkers. LDL cholesterol in ε2

carriers was lower in drinkers than non-drinkers, being the other way around for the ε4 allele carriers. For women the effects of APO E on LDL cholesterol were obvious for both drinkers and non-drinkers alike (Corella et al., 2001).

The last aspect of alcohol consumption that we will discuss is the idea that recommending moderate alcohol intake at a national level will lower the prevalence of CHD. There is no doubt that in the last two decades people have become more health conscious than in the past and are willing to follow public health advice. Discussion of the health benefits of red wine is believed to be have been a driving force in the increasing sales of red wine in the US (Klatsky, 2001). Before advising moderate alcohol intake we should make clear that there is also going to be an increase of other forms of mortality associated with ethanol intake. Secondly, the population theory of alcohol consumption argues that increase of the mean alcohol consumption in the population will also increase the number of heavy drinkers, thus diminishing the benefits of moderate intake for the population in total (Marmot, 2001b).

Smoking

Smoking is mainly associated with mortality through cancer, but it is also a significant risk factor for cardiovascular disease. It is believed that the products of tobacco combustion damage the vascular epithelium, increasing the secretion of adhesion molecules promoting thrombosis and atherosclerosis (Humphries et al., 2001). Cigarette smoking has been linked to the insulin-resistance syndrome, causing elevated triglyceride levels, decreased HDL cholesterol, dysfibrinolysis and an increase in small dense LDL particles, all risks of CVD (Eliasson et al., 1997). The association of smoking and *apo E* genotype has been studied by Humphries et al (2001) using a sample of 3,052 middle-aged man with no history of CHD, concluding

that smoking is a risk factor for CHD especially in men having the ε4 allele while ε2 carriers exhibited no increase at all. Further studies have supported the effect of *apo E* genotype on the cardiovascular disease risk of smoking in hypertensive male (van der Meer and Witteman, 2002), while others failed to find any evidence for this genotype-environment interaction (Keavney et al., 2003; Liu et al., 2003).

Exercise

Physical activity is considered to be a significant factor in the prevention of both primary and secondary cardiovascular disease (Fletcher, 1997). It is estimated that some 37% of CHD deaths can be attributed to physical inactivity (Press et al., 2003). Regular exercise can improve myocardial contraction and its electrical stability, increase heart-stroke volume and cardiac output, and decrease heart rate. At the same time, endothelial function is also improved, the dilatory capacity of the coronary arteries is increased, while inflammatory factors are decreased. Physical activity also reduces platelet aggregation and increases fibrinolytic activity affecting the tendency of blood to clot. Finally lipid oxidation is stimulated; lipase activity is increased, accelerating triglyceride clearance and the conversion of VLDL to HDL, while a higher HDL to LDL ratio is observed together with a decrease in fasting insulin levels and its response to glucose (see review by Press et al., 2003). Interestingly, a number of studies found that although occupational physical activity causes much higher total energy expenditure than leisure-related physical activity, it is not associated with changes in the CHD risk factor (Koenig et al., 1997; Kaprio et al., 2000). Rothenbacher et al (2003) confirmed that moderate levels of leisure exercise were enough to decrease the risk of CHD, but they also found a positive association between work-related physical activity and risk of CHD. They attributed that discrepancy to the different duration and intensity associated with each mode of

physical activity, although they recognised that residual confounding of another associated risk is also possible. At the same time, a link between physical fitness and all-cause mortality and CVD has been shown to exist (Blair et al., 1995) which leads us to ask how closely linked is occupational physical activity with physical fitness or physical well being, what are the factors affecting that association and how they compare with those of leisure exercise? To my knowledge, these questions have not been answered yet despite their importance for revealing which are the main factors of exercise affecting cardiovascular and general health. Recently, an investigation has been carried out to assess the amount, type and intensity of exercise needed to lower CHD in men (Tanasescu et al., 2002). The study involved 44,452 men, whose information was updated every two years from 1986 to 1998, and revealed that there is a significant inverse dose-response relationship between total physical activity and cardiovascular risk, while intensity was also related to reduced risk, with walking pace showing an inverse relationship to CHD incidence.

Since, as we mentioned earlier, exercise has an effect on the lipid profile and APO E is a key factor involved in that process, we would have expected that the association between APO E phenotype and lipids is affected by physical activity patterns. A number of different studies using quite different populations and techniques have come up with conflicting results. Using young adults from the CARDIA study Schmitz et al (2001)concluded that fitness did not influence the association between APO E and HDL or LDL, a result in accordance with another study on Finnish males between the ages of 9 and 24 (Schmitz et al., 2001; Bernstein et al., 2002). While a population study by Bernstein et al (2002) found that ε4 carriers had a greater protective effect from intense exercise compared to ε2 carriers or ε3 homozygotes. To explain the contradictory results, Bernstein et al (2002) addressed

two important points, exercise intensity and the detailed association between physical activity and HDL change. We have already seen that the intensity of physical activity is modulated through a dose-response mechanism, as for the change in serum lipids there is evidence suggesting that exercise increases HDL in those who need it more (Thompson and Rader, 2001). Zmuda et al (1998) showed that the ability to increase the HDL cholesterol levels in blood through endurance exercise training is limited in subjects that initially had low HDL, probably due to the failure of exercise to change their triglyceride metabolism. Coillard et al (2001) using the HERITAGE family study concluded that endurance exercise is especially helpful in men with low HDL and high triglyceride levels, in contrast to those with isolated low HDL who seem to be much less responsive. Thus the discrepancy between the different studies can either be attributed to different intensities of the schedules which were followed or to the underlying heterogeneity of the population.

4.6.2 Relative risk and odds ratio

We have already presented the evidence for the risk factors chosen for our model and now we will describe the method used to quantify their effect on CVD. To achieve this we used a proportional hazard model describing the risk ratio between two different sets of conditions. According to this, if $\mu_1(x)$ is the hazard of death of genotype 1 at age x, and $\mu_2(x)$ is the hazard of death of a genotype 2 at the same age x, then $\mu_1(x) = R \times \mu_2(x)$, where R expresses the relative risk of genotype 1 compared to genotype 2 (Gerdes et al., 2000).

The first case that we will consider is the calculation of the relative risks for the six different genotypes produced by the combination of the three common alleles. For the calculation of the risk factors associated with each genotype we relied on the

work of Gerdes et al (2000). Using individuals with the ε3ε3 and ε2ε4 genotypes as the reference group (a convention that we will keep throughout our model) Gerdes et al (2000) estimated the average relative mortality risk in ε2 and ε4 carriers in 3 different studies. Out of the three we will use the values of 0.9 and 1.13 as the most representative for the relative risk of ε2 and ε4 carriers respectively. Due to the low frequency of ε2 and ε4 homozygotes in most populations, the great majority of studies combine the homozygotes and heterozygotes with the ε3 allele in a larger category of ε2 and ε4 carriers. Thus the relative risk measurements are done between the ε3ε3 homozygote and the combined ε3ε2 and ε2ε2 genotypes for the ε2 allele, and similarly for the ε4 carriers. In order to construct our model though, we need to know the risk factor for each one of the genotypes instead of just the risk for the carriers of an allele. To calculate the risk ratio for all the six different genotypes we first have to find their allelic effect, the change of the mean lifespan observed if we could substitute one allele of a population of homozygotes with the allele we are interested in. Using the relatives risks described by Gerdes et al (2000) and a form of the Gompertz curve representing the change of mortality with age we found that carriers of the ε2 allele were expected to live 1.12 years more than the homozygotes for ε3, while the carriers of ε4 showed a decrease of 1.3 years in average lifespan compared to the ε3ε3 genotype (Figure 4.1). Assuming that the alleles have an additive effect, as supported from experimental evidence (Bohnet et al., 1996; Zerba et al., 1996), we can use the frequencies of the two different genotypes within the carrier groups for each allele to calculate the allelic effects in each case. Thus we were able to calculate that the presence of one ε2 allele extends mean life span by 1 year, while a single ε4 allele decreases life expectancy by 1.1 years. Again using the assumption of an

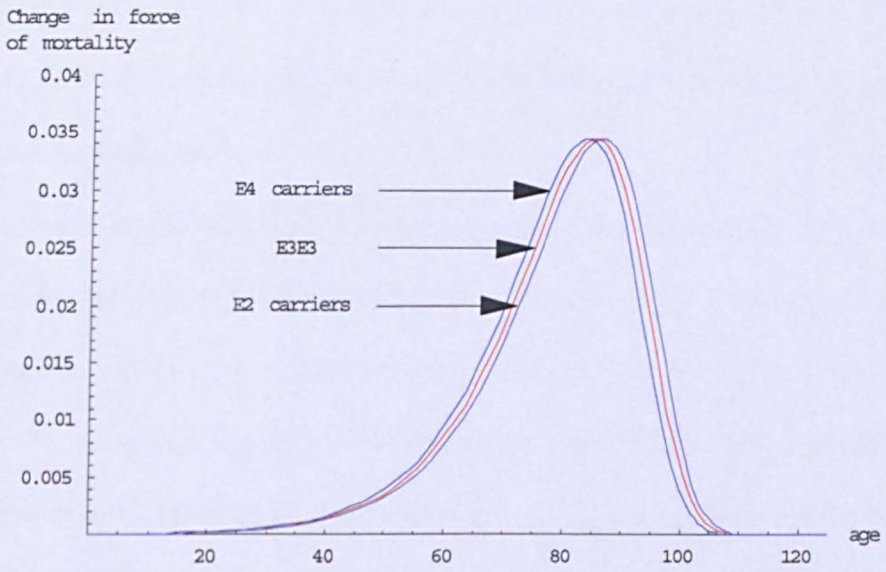


Figure 4.1. Change in the force of mortality with age for ε3ε3 homozygotes and ε2 and ε4 carriers.

additive allelic effect we can find the mean lifespan of each genotype and to finally turn those in the form of risk factors relative to the ε3ε3 genotype to be used in our model (see Appendix 3).

Many health studies though, use odds ratios to describe the data obtained, instead of the more accurate relative risk. The odds ratio is calculated by dividing the odds of an event in a treated or exposed group by the odds of the event occurring in the control population. Its popularity is mainly attributed to the widespread use of the logistic regression to analyse binary categorical data. The problem is that the logistic regression and the odds ratio are accurate only for rare events (<10%) but are often used to study common outcomes (McNutt et al., 2003). In these cases the odds ratio usually overestimates the true effect of the risk factor of interest (Robbins et al., 2002). To approximate the relative risk from the adjusted odds ratio we will use the very popular methodology of Zhang and Yu (1998) as the simplest one, and the one that requires less information for the dataset used. According to this

$$RR = \frac{OR}{(1-P_0)+(P_0 \times OR)} \qquad (1)$$

where RR is the relative risk, P_0 is the incidence of the outcome in the non exposed population and OR is the odds ratio. Despite the method's slight inaccuracy in calculating confidence intervals and inability to account for confounding factors, its balance between simplicity and precision makes it an especially useful tool (Robbins et al., 2002; McNutt et al., 2003).

We will use equation (1) to calculate the relative risk associated with gender using the odds ratio provided by Stevens et al (2001) and the incidence described by Panagiotakos et al (2002). Substituting the incidence for women with 20% and the OR with 0.525 we can calculate the relative risk as 0.58. Using the Gompertz model

for the probability of death per year of age we can see that the difference produced between males and females is almost 6 years which is within the limits of the expected lag of cardiovascular disease between men and women (Figure 4.2). Using the same method we estimate the relative risk associated with each dietary pattern group of the Framingham study based on the odds ratio provided by Millen et al (2002). If the Heart Healthy group is used as the reference with risk ratio 1.0, using equation (1) the Light Eating risk is corrected to 1.17, the Wine and Moderate is changed to 1.32 which is very close to the original odds ratio mainly due to the small representation of the group in the population. The relative risk for a High Fat diet is 1.36 and finally the Empty Calorie diet risk factor was estimated to 2.07. The Mathematica notebook with the detailed calculation can be seen in Appendix 4.

For the remaining risk factors considered, relative risks were provided by Stampfer et al. (2000). So in the case of alcohol consumption and for the purposes of the model we considered seven groups of alcohol intake: the non-drinkers group, three groups for various levels of moderate intake, and three groups for heavy intake including those with episodic binge drinking despite their mean alcohol consumption. For simplicity we assumed that there is no difference between men and women for the risk factors in each group. As a basis we used the risk factors as calculated by Stampfer et al (2000) using a sample of 84,129 women in the Nurse's Health Study.

Although, as we saw earlier, the association of smoking with apo E is still debatable, in the model we assumed that there is an interaction between them and again used the risk factors presented by Stampfer et al (2000) using the four groups provided. Here we should mention that in the case of former smokers there might be an increased representation of those already suffering from CHD which will disturb the calculated risk factor, and that the group with 1-14 cig/day can include very

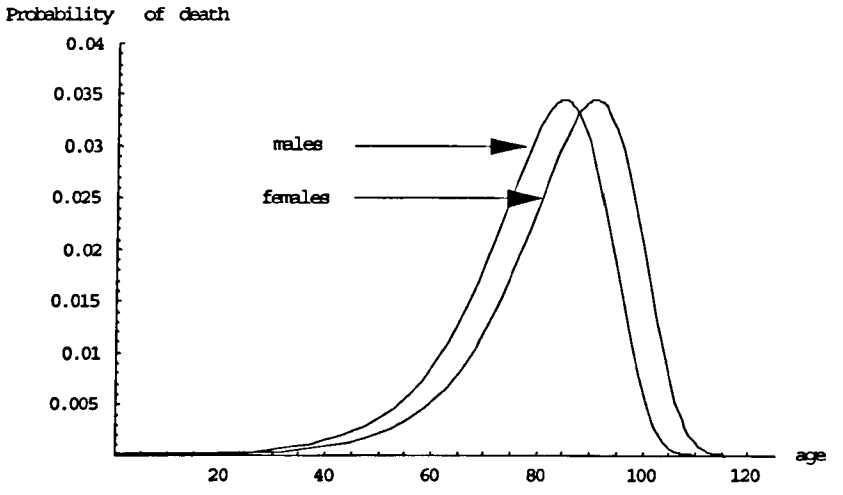


Figure 4.2. Difference of average age at peak mortality between men and women.

different patterns of smoking ranging from irregular smokers to moderate but everyday ones.

Finally, to incorporate the changes of CHD incidence in relation to physical activity we used five different groups of activity patterns depending on their duration of exercise[8] in hours per week as described in Stampfer et al (2000).

4.6.3 Genotype environment interaction

As we have seen, there are studies that link *apo E* genotype with all the modifiable risk factors described, but their precise interaction in quantitative terms remains almost completely unknown. Although the description of the interaction between genotype and the modifying parameters of the CVD risk will require a series of epidemiological studies in a big population, we will try to compensate for the lack of data using the given risk factors and qualitative results of the existing studies to predict the expected associations between environment and *apo E* genotype.
We will go through the method used using the risk factors for CVD associated with smoking. The first step is to assume that all the risk factors considered up to now are independent of each other, so that for example people that smoke more have the same chance of drinking more as those that never smoke. Although there is some evidence to support the opposite (Twisk et al., 2001) the simplicity that this assumption confers significantly outweighs the small changes that the clustering of risk factors would cause to our calculations. This assumption permits us to consider that if all other factors are equally distributed among smoking groups, then the differences observed between the six genotypes can be considered as the result of smoking × genotype interaction. Moreover, since the ε3ε3 genotype is by the far the most common

[8] Defined as "strenuous enough to build up sweat"

genotype in western populations, we can assume that the risk factors described by Stampfer et al (2000) are those corresponding to the ε3ε3 genotype. The next step is to construct a graph of relative risk with category of smoking and fit a line through the data in accordance with our knowledge of the quantitative relationship between the two. This provides us with a continuous curve of changing risk with level of smoking (Figure 4.3.a). Using this curve we can find a series of curves relative to the one we already have that their risk ratio across all categories will correspond to the relative risk observed between the genotypes (Figure 4.3.b). Thus we have an approximation of the apo E genotype × smoking interaction for the risk of CVD. Similarly we can calculate the relative risk between *apo E* genotype and the rest of the modifiable risk factors. Thus in the end, the total risk of an individual can be simply calculated as the product of the gender relative risk and all four modifiable risk factors corresponding to the individual's genotype. Details of the method used can be seen in Appendix 5. A simulation of 500,000 individuals in the program revealed that our method overestimates the average risk of the population by 3.5. Calibrating the model according to this we achieved a simulated population with a similar mean and variance to a real population (real mean ~75.9, while simulated mean ~75.7 years of age). Further tests of our program proved that the calculated relative risks accurately reflect the situation of real life populations.

4.6.4 Gompertz model and the Gumbel distribution

As we mentioned earlier we can use a curve derived from the Gompertz model to link relative risk to changes of mean life span and to represent the probability of death at each age. To derive the equation for that curve from the equation of the

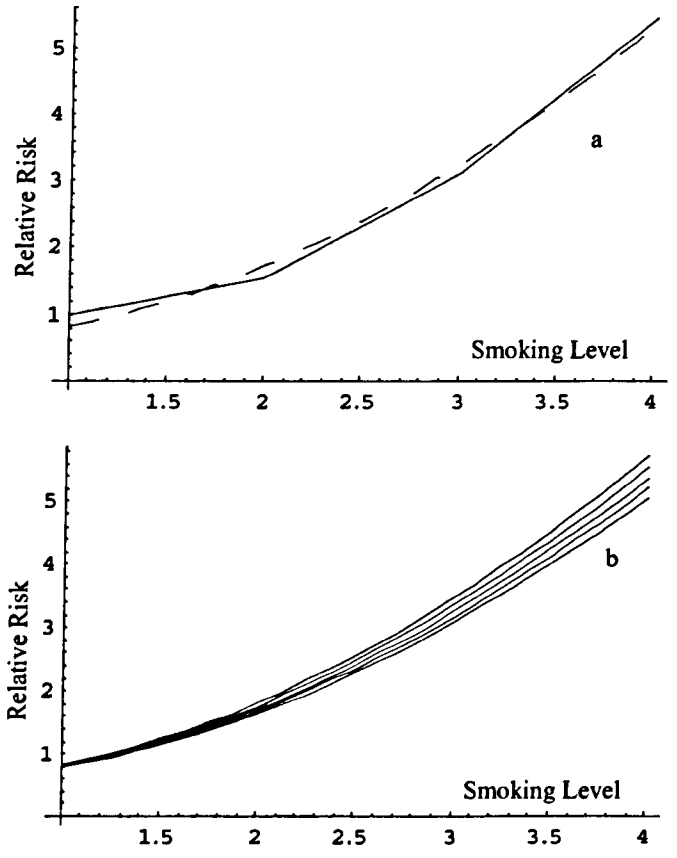


Figure 4.3. **a)** Graph representing the original (solid) and fitted (broken) lines for the change of relative risk with category of smoking. **b)** Calculated lines for the interaction between smoking and *apo E* genotype. From top to bottom the lines are for ε2ε2, ε2ε3, ε3ε3 and ε2ε4, ε3ε4, ε4ε4 genotypes.

Gompertz model, seen in Chapter 1, we consider that the survival $l(x)$ at an age x can be found as:

$$l(x) = e^{-\int \mu(x)dx}$$

where $\mu(x)$ is the force of mortality at age x. Thus the change in survival or equally the change of mortality from age $x-1$ to x can be written as:

$$M_{Diff} = e^{-\int_1^x \mu(y-1)dy} - e^{-\int_0^x \mu(y)dy}$$

or

$$M_{Diff} = e^{-\frac{(-1+e^{(-1+x)\beta})\alpha}{\beta}} - e^{\frac{(1-e^{x\beta})\alpha}{\beta}} \qquad (2)$$

This transformation of the Gompertz curve is equivalent to a probability density function (PDF), having a total probability of 1 and each of the points of the curve denotes the chance of death at the specific age for an individual with basal vulnerability α and rate of ageing β. Comparing the modified Gompertz curve relative to the PDFs of other well known statistical distributions commonly used in mortality studies we found that the Extreme Value Distribution for minima is almost identical to it (Figure 4.4 and Appendix 6). To understand the biological significance of that, we first have to explain what the extreme value distribution is. Probabilistic extreme value theory is mainly concerned with the stochastic behaviour of the maximum or minimum of distributions. Although it has a long history in statistics, dating back to the beginning of the 18th century, its use has been limited in engineering and hydrology to predict the reliability of complex machines and the frequencies of extreme physical phenomena. There are two classes of extreme value distributions: the first class known as the generalized extreme value distribution, and

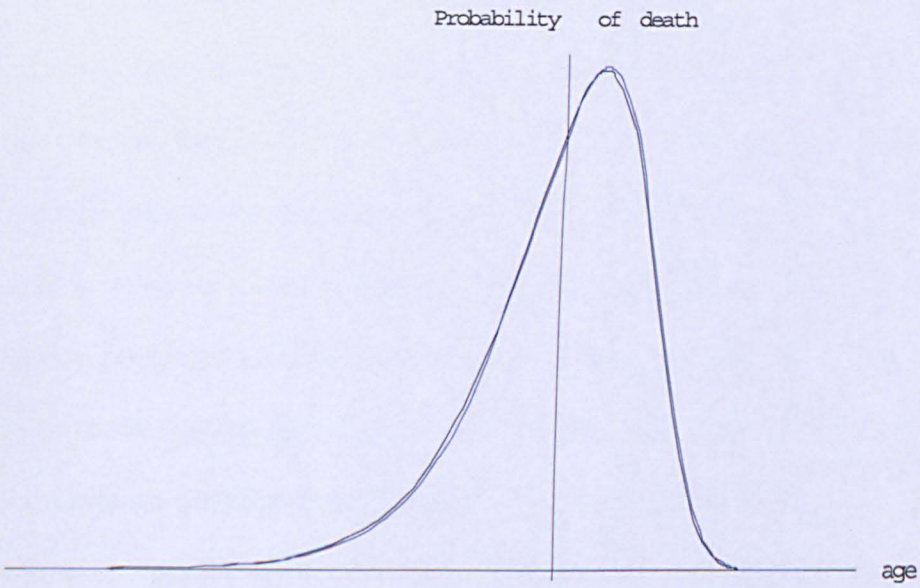


Figure 4.4. Comparison of the modified Gompertz curve (black line) and the probability density function curve for the Gumbel distribution (blue line). The y-axis represents the mean age of death for the population.

the second one as the generalised Pareto distribution (Bali, 2003). We will consider the first class which includes three different families of extreme value distributions: the Gumbel, the Frechet and the Weibull, each alternatively called type 1, 2 and 3 extreme value distributions respectively (Kotz, 2000). All three of them are mainly used in reliability theory in engineering; especially the Weibull model that can be used to predict the lifetime of complex mechanical objects and which has also found some use in demographic statistics. Here we will use a form of the Gumbel model for the minimum which gives the distribution of the smallest extreme from a set of normal or logistic curves. In terms of biological significance we can consider humans, and animals in general, as complex biological machines that will fail when a core part of the system fails. Each part will be expected to fail based on a normal or logistic distribution in accordance with the ideas of Fisher, stating that a quantitative genetic characteristic that is influenced by the environment will show a normal distribution around a mean. The use of an extreme value distribution, in the writer's opinion, is an important alternative to consider age related mortality and changes. In this way we can account both for the heterogeneity in ageing rates of different organs within a single organism and for the deceleration of the force of mortality at later ages.

To accurately construct the age-of-death probability density curves in our program, we made use of recent UK population data to calculate the α and β parameters of the Gompertz model. To do that we used the interim life tables from The Government Actuary's Department for males in England and Wales for the years 1999-2001. In Appendix 7 the details of the methods used to calculate the Gompertz parameters can be seen while the graph representing the number of deaths in each year of age can be seen in figure 4.5.

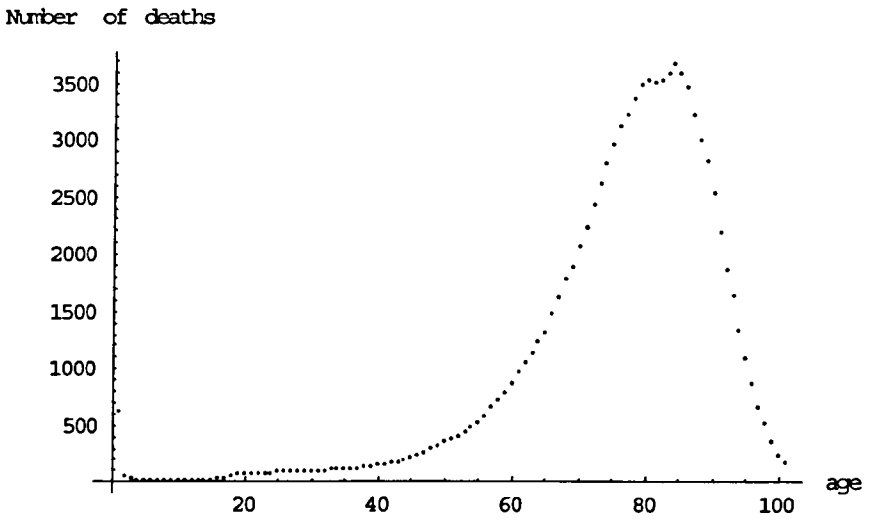


Figure 4.5. Number of deaths with age for England and Wales for the years 1999-2001. The close fit of real data to what is expected form the Gumbel distribution and the modified Gompertz is evident.

4.6.5 Random number generators

To simulate a real-life population as realistically as possible we will make extensive use of random number generators. Although very sophisticated true random numbers generators exist today that make use of radioactivity, radiowave noise or even lava lamps, they are mainly used for cryptographic applications and have high demands of processing powers, making them prohibitive for population models. Instead we shall use a pseudo-random number generator from the GNU Scientific Library (GSL version 1.4) for C and C++. These generators, commonly used for simulations, employ an algorithm to generate a deterministic set of numbers which behave and look similar to genuine random numbers (for more information on random numbers and random number generators see L'Ecuyer, 1998). The algorithm that we used was MT19937 by Makoto Matsumoto and Takuji Nishimura which is a variation of the "Mersenne Twister" generator and has a period of 2^{19937} - 1 (about 10^6000) and is equi-distributed in 623 dimensions (GSL Reference Manual).

The model uses random numbers to allocate the genetic and non-genetic parameters of the model for each simulated individual based on the frequencies found in real populations. Then, an overall risk ratio was calculated and a probability curve constructed, describing the probability of death at each age for the given combination of factors. Again using random numbers, an age of death is chosen from the probability distribution of the individual to represent the stochastic component of longevity.

4.6.6 Simulating evolutionary change

Since we have already described the basic aspects of the methods used to obtain the various elements of model, we are now ready to present the procedure followed by the program to simulate a human population and the process of evolutionary change for the frequencies of the three *apo E* alleles. To start, our population consists of an equal number of males and females. These are grouped into couples which reproduce either till they reach their 50^{th} year of age or until one of them dies. The reproductive schedule is the same as found in contemporary western populations, so that the probability of giving birth decreases with increasing age. Each offspring is given a random genotype based on the genotypes of its parents and in the end all the produced genotypes are pooled into a matrix which constitutes the initial conditions for the next generation. The process is repeated as many times as the number of the generations required. If the *apo E* alleles have a selection differential between them, we expect to see that there will be a directional change of their frequencies in the population with each successive generation. In the model this can be observed in the case that the proportion of early deaths in the population is higher for one of the alleles compared to the other two, and this effect is within the normal reproductive lifespan so as to limit the number or survival of the offspring.

The results of the model, as expected, have a very big stochastic element. This is expected in nature and seldom will selection be so intense so that the stochasticity of the system is negligible in relatively small groups. To overcome the problem and obtain reliable results we had to run the model many times so we could further analyse the data statistically. A simple version of the program can be seen in Appendix 8.

4.6.7 Statistical manipulation of the results

Our model can be used to test if, in a western population, there is selection for a particular allele or genotype. To minimize the random variation of our results we used a set of at least 60 runs, each comprising 200,000 individuals followed for 50 generations. These provided a number of different possibilities for the system to evolve. We summarized the data calculating the mean and standard deviation of all 60 runs for each generation and each allele, using the results to plot the mean and twice the standard deviation around it to indicate the confidence intervals of the line (Figure 4.6). Since a regression analysis assumes that the variance of the sample remains the same, it is not appropriate to use it in this case where we can clearly see that the variance increases with each generation. To overcome the problem we used the method of random walk in one dimension to obtain the average change with each generation. First, we have to assume that the change of the allele frequencies is linear and can be described by a simple equation. As long as the alleles do not have a pleiotropic effect, which will balance the selection for or against them, we will expect their frequencies to increase until they are fixed or extinct in the population. A simple equation to describe the change can be written as:

$$\alpha_t = \alpha_{t-1} + d + \Sigma_t$$

where α_t is the frequency at time or generation t, α_{t-1} is the value of the frequency in the previous generation, d is the change between two consecutive points in the series and Σ_t is random noise with an underlying normal distribution. In order to calculate d we find a y_{jt} such that:

$$y_{jt} = \alpha_{jt+1} - \alpha_{jt}$$

with j = 1,.....,60

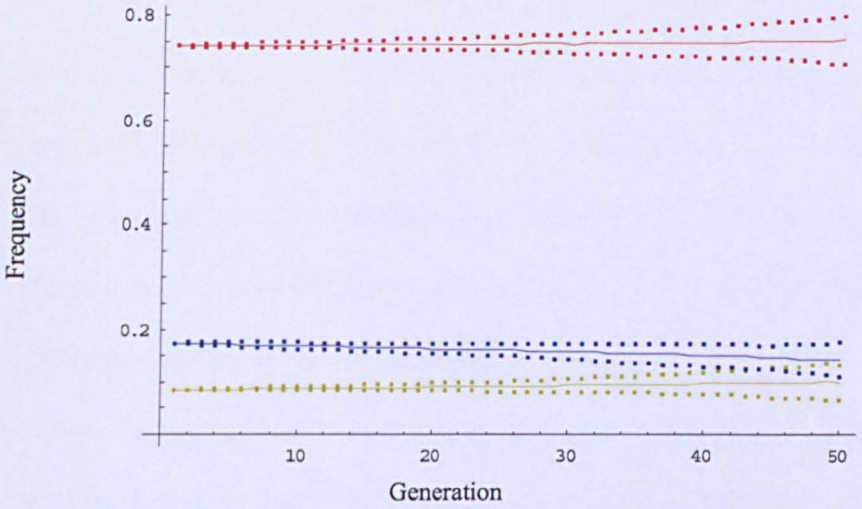


Figure 4.6. Change of the three apo E alleles with each generation. The red line represents the ε3 allele, the blue the ε4 and the yellow the ε2. The broken lines represent the 95% intervals of the tree lines.

and $i = 1,\ldots,49$

giving us a sample of 49×60 elements. From this sample we can easily find the mean and its confidence interval (CI). If the CIs are positive then there is evidence for a significant increase of the allele frequency with each generation, while if negative the allele goes towards extinction. In the case that 0 is included within the confidence intervals, we cannot consider the change of frequency to be significant and selection is probably extremely weak, with the change attributed to the randomness of the system (Appendix 9). The 49×60 matrices obtained were also transferred into Minitab statistical software to test the significance between and within different variants of the program. Furthermore, the mean life span of the population can be easily calculated by running the simulation without reproduction for 500,000 individuals giving us the expected changes in lifespan associated with each parameter of the model.

4.7 Use and results of the model

Having explained the methods used for the construction of our model and simulation program we are now ready to present the results obtained by our default version of the program and start testing the selection pressure on the *apo E* alleles under a number of different conditions. In section 4.7.1 we will explore the parameters used in our model by running the simulation and changing a single parameter every time. For part of 4.7.2 we will use our model to elucidate the origin and early evolution of the three alleles by the use of data from contemporary hunter-gatherers to parallel the conditions of human lifestyle during the spread of the two novel ε2 and ε3 alleles in a prehistoric population.

4.7.1 Analysing the parameters of the model

A number of different variants of the program were considered. The default version of the model, with parameter values similar to those found in a typical western country, clearly shows that there is selection for the ε2 and ε3 alleles against the ε4 one, with the ε2 frequency increasing slightly more rapidly than that for ε3, although there is no significant difference between them.

The second variant of the model is concerned with the selection forces applied to the alleles when only a fraction of the population is affected by the action of the gene. If we consider that only about 30% of the total mortality in the population is *apo E* dependent and that *apo E* dependent and independent mortality have the same underlying shape (Health Trends - A Working Document, http://www.health.gov.ab.ca/public/document/Health_Trends/index.html) then we can run the entire simulation again to see if there is any selection for the gene in a heterogeneous population with various causes of mortality. The program shows that even when *apo E* independent mortality is considered, selection for the alleles is still evident and approximately one third of the original. Thus if no antagonistic pleiotropy for the *apo E* alleles exists, then in a contemporary western population a clear selection against the ε4 allele and a positive one for the ε2 and ε3 alleles can be observed. Again, the rates of increase for the ε2 and ε3 alleles do not differ significantly.

Changes in diet

The model was also used to test how the selection forces on the *Apo E* alleles change when the lifestyle parameters in the model vary. The first parameter considered was diet. For the first change the Heart Healthy and Light Eating groups

were decreased by 20% while the High Fat and Empty Calorie groups increased to compensate for the change. Then we altered the model in the opposite way, with High Fat and Empty Calorie groups decreased by 20%. As can be understood the second change will be much smaller since the two groups have much lower frequencies, as can be seen in Millen et al (2002). In the first case the average lifespan of a population with only *apo E* related mortality decreased from 75.69 to 74.93 years, while in the second case the average lifespan increased to 75.98 years. The selection forces for the alleles in both cases remained significantly different from zero, but when their frequencies for the two diets were compared with each other or with the original pattern, no significant differences were found. Additionally, there was no significant difference between ε2 and ε3 change with any of the diet variants considered; from now on this will be the case unless otherwise stated.

Changes in alcohol consumption

As we saw earlier, alcohol is a very important parameter for the cardiovascular health of the population, but it is also linked with risk of accidental mortality and many social problems. What makes alcohol a special case is that together with smoking it is closely monitored by governments and both are commonly subjects of high taxation, making them central in policy decisions. As with diet, we used two different variants of the simulation. In the first, light drinkers were decreased by 25%, while in the second the decrease was in the groups of the heavy drinkers. The mean lifespan increased by 0.28 years in the first case, while in the second case a very slight decrease to 75.54 years of mean lifespan was observed. This does not mean that an increase of alcohol consumption will be beneficial for the population. Our model measures the longevity of a group dying from cardiovascular *apo E* related

aetiology; it is well known that heavy drinking will have a direct adverse effect on the heart muscle that will overcompensate for the slight increase we observed. Our results just point to the complex association of alcohol and longevity and our inability to follow simple strategies to successfully use the beneficial effects of ethanol while avoiding its abuse. In terms of the selection at the level of the entire population, there seems to be no significant difference between the two variants of the model, so that the 25% change in either way will not affect the selection pressure on the alleles, although in the case of ε3 and ε4 the effect is more intense than for the ε2 allele.

Changes in exercise level

The next parameter to be addressed was the level of physical activity of the population. First we tested the changes in longevity and selection when the minimum level of the exercise in the population was 2.5 hours per week, or the equivalent of walking fast for half an hour every weekday. This resulted in an increase of mean lifespan to 76.19 years, half a year more than the present mean. The second variant of the program tested the results of a decrease by 25% of the two most active groups. In this case the population lifespan dropped to 75.41 years, showing that the small frequency of very active individuals is too small to play a significant role in the longevity determination in the population. Selection for the gene remained significant but there was no difference for it between the default and the first exercise pattern and between the first and the second change in the level of exercise.

Changes in smoking

The final parameter investigated was the prevalence of smokers in the population. Although a significant decrease has been seen in smoking in western

countries over the last decade, the problem still remains and in many developing nations is increasing rapidly. Once more, two variants of the model were considered. In the first, the smokers increased by 25% while in the second the population consisted solely of non-smokers. The model predicts that a 25% decrease in non-smokers would result in a mean life span of 75.57 years which is fairly close to the initial lifespan. On the other hand when we take out all the smokers from the population the mean longevity jumps to 82.95 years. For both variants there was no significant difference between the ε2 and ε3 alleles, but now there was more dissimilarity than usual. In the case of decreasing the non-smokers, the selection pressure was still significantly different from zero and non significant in relation to the default form of the model. In the second smoking variant of the simulation we see that ε2 and ε4 retain their CIs for the mean all positive and all negative respectively, while in the case of the ε3 allele the value of the average random walk for the first time loses its significance, thus it can be attributed to the stochastic nature of the model. To demonstrate that this result was not due to some extreme case we ran the model 120 times instead of the standard 60 but the non-significance for the ε3 allele remained stable. At the same time, the change in smoking caused a significant difference relative to the default situation, with the change of the ε3 allele being significant at the 0.949% level while the other two alleles showed even higher levels of significance. Since the difference in the parameter tested was much larger than any other difference previously tested, it was expected that the random walk would now be different from what we saw earlier. As for the loss of significance in the case of the ε3 allele, it does not mean that the selection for it is solely based on the effects of smoking. Running another variant of the model with no smokers in the population for a group affected solely by *apo E* related mortality we see that the selection force on

the alleles is strong, but when only 30% of the population is affected, then the selective pressure is diluted enough to be indistinguishable from the randomness of the system. Furthermore even though the change for the ε3 allele does not reach the 95% significance level, most of the area of the CI for the mean remains above zero indicating that it is expected more often to have a small positive change rather than a negative or no change at all. All results are summarized in tables 4.1 and 4.2.

4.7.2 Origin of the apo E alleles

Further variants of the model addressing the problem of the past evolution of the *apo E* gene were also tested. For this purpose we used the program which included *apo E* un-related mortality and two different patterns of diet. In the first one the main source of energy was provided by a meat diet with a high fat content. There is evidence that around 2.5 million years ago our early ancestors started to consume food requiring more tearing than grinding, signifying the change towards meat eating (Boyd Eaton et al., 2002; Richards, 2002). This meat will have initially come from small mammals and scavenged skulls, left over by bigger predators, with the use of tools to get access to the brain, a part of the body that carnivorous animals do not always consume (Cordain, 2002). Of course this would have been very opportunistic in the beginning but as the size and sophistication of hominids increased so did their ability to hunt larger prey, as can be seen in the case of Neanderthals (Hockett and Haws, 2003). Although humans retained a diverse diet, modern hunter-gatherers show that they probably relied on meat for more than 50% of the daily calorie consumption (Kaplan et al., 2000). In the second variant of the model, the diet was changed from high fat to high carbohydrates. In this case the relative risk of an

Run	Parameters	Mean Lifespan
Run 1 default	Default	75.69
Run 3 diet 1	Heart Healthy and Light Eating groups decreased by 20%	74.93
Run 4 diet2	High Fat and Empty Calorie groups decreased by 20%	75.98
Run 5 alcohol 1	Light and non-drinkers decreased by 25%	75.97
Run 6 alcohol 2	Heavy drinkers decreased by 25%	75.54
Run 7 exercise 1	Entire population moderately active (2.5h/week)	76.19
Run 8 exercise 2	Most active groups decreased by 25%	75.41
Run 9 smoking 1	Non-smokers decreased by 25%	75.57
Run 10 smoking 2	No smokers in the population	82.95

Table 4.1. The simulated average lifespan in years of each variant considered in the absence of non Apo E related mortality. All the changes, except the last one, are moderate changes in lifestyle parameters that can take place in the population within relatively short term (~ 25 years).

Run	Apo E allele	Mean change of allele frequency per generation	95% confidence intervals for the mean	
Run 1 default	ε2	0.0003352	0.0002661	0.0004043
	ε3	0.0002785	0.0001809	0.0003761
	ε4	-0.0006137	-0.0006946	-0.0005327
Run 2 deafult 2	ε2	0.0000846	0.0000497	0.0001195
	ε3	0.0000846	0.0000316	0.0001377
	ε4	-0.0001692	-0.0002140	-0.0001244
Run 3 diet 1	ε2	0.0000900	0.0000547	0.0001252
	ε3	0.0000899	0.0000360	0.0001438
	ε4	-0.0001798	-0.0002257	-0.0001340
Run 4 diet2	ε2	0.0000809	0.0000480	0.0001139
	ε3	0.0000643	0.0000130	0.0001157
	ε4	-0.0001453	-0.0001905	-0.0001000
Run 5 alcohol 1	ε2	0.0000836	0.0000524	0.0001148
	ε3	0.0000818	0.0000326	0.0001309
	ε4	-0.0001654	-0.0002081	-0.0001226
Run 6 alcohol 2	ε2	0.0000847	0.0000500	0.0001195
	ε3	0.0000588	0.0000074	0.0001103
	ε4	-0.0001436	-0.0001880	-0.0000991
Run 7 exercise 1	ε2	0.0000895	0.0000555	0.0001235
	ε3	0.0000629	0.0000108	0.0001149
	ε4	-0.0001524	-0.0001959	-0.0001088
Run 8 exercise 2	ε2	0.0001088	0.0000749	0.0001427
	ε3	0.0000637	0.0000115	0.0001158
	ε4	-0.0001725	-0.0002171	-0.0001279
Run 9 smoking 1	ε2	0.0001283	0.0000897	0.0001668
	ε3	0.0000974	0.0000383	0.0001564
	ε4	-0.0002256	-0.0002766	-0.0001746
Run 10 smoking 2	ε2	0.0000406	0.0000205	0.0000607
	ε3	0.0000259	-0.0000057	0.0000574
	ε4	-0.0000665	-0.0000936	-0.0000393

Table 4.2. The mean change of the each allele frequency for the various variants of the model with the 95% CI of the standard error of the mean in each case. For all runs we had 60 repetitions each, except run 10 which had 120.

empty calorie diet was used to signify the overabundance of calories in the diet of a human adapted to long periods of near famine. Agriculture started around 10,000 BC somewhere in the Near East and later independently in other parts of the world. This led to an increasingly sedentary population with access to food high in carbohydrates and calories. At the same time a general decline in health and stature was observed, together with the appearance of some nutritional disorders (Richards, 2002). Farmers probably expanded due to their ability to sustain their bigger numbers, thereby spreading their new technology. This movement can still be seen in Europe today and displays a south-east north-west gradient similar to the frequencies of the *apo E* alleles (Cavalli-Sforza et al., 1993). For both cases the other parameters considered will be changed to mirror what we expect to find in a population in the Palaeolithic and Neolithic ages. Thus, alcohol consumption and smoking will be set to zero, while we will assume that the entire population is moderately active. The final variant of the program to be tested takes into account that only 10% of the entire population will be affected by the action of the gene. This is based on the fact that only 20% of a Hadza population, a modern equivalent of hunter gatherers, reach the age of 65 and that even from those, not everyone will die from *apo E* related diseases (Kaplan et al., 2000). The result for the last three variants can be seen in Table 4.3. Again for ε2 it seems that there is no significant difference between the two diets although the high carbohydrate diet seems to have a more pronounced effect. In the case of the ε3 allele the significance seen here is probably due to the randomness of the model used rather than a real effect. For the last run of the program it can be seen that the selection for ε2 remains significant even in the case where a very small proportion of the

Run	Apo E allele	Mean change of allele frequency per generation	95% confidence intervals for the mean	
Run High Fat	ε2	0.0003212	0.0002829	0.0003596
	ε3	0.0001759	0.0001240	0.0002278
	ε4	-0.0004971	-0.0005586	-0.0004357
Run High Carbo	ε2	0.0003603	0.0003102	0.0004105
	ε3	0.0003267	0.0002621	0.0003913
	ε4	-0.0006871	-0.0007652	-0.0006089
Run High Fat with 10% Apo-E mortality	ε2	0.0000252	0.0000044	0.0000461
	ε3	0.0000215	-0.0000074	0.0000504
	ε4	-0.0000467	-0.0000814	-0.0000120

Table 4.3. The mean change of the each allele frequency for the three variants concerned with changes in diet mirroring those found during human evolution. The first and third variants have the same parameters but for the last one the effect of Apo E is limited to only 10% of the entire population.

population is affected by the action of the gene, while for ε3 although the difference can not easily be distinguished from the random events it seems that it is enough to play a role if the time interval is sufficiently long.

4.8 Discussion

Our aim was to find if the *apo E* gene is acted upon by natural selection in a contemporary population despite the fact that its main effect is beyond the end of reproduction. We have shown that given the environmental parameters often found in Western countries the level of selection can be calculated as the change of frequency of the alleles between generations. We found that the ε2 and ε3 alleles are selected at the expense of the ε4 allele, with ε2 showing a larger but not significantly different change per generation compared with the ε3 allele. Furthermore the selection remained significant even when the parameters changed between 20-25%, showing that the model is not very sensitive to the frequencies considered, thus our results will stand even if there is a degree of error in the values used. In the case of smoking we saw that when the smokers are taken out of the population the change of ε3 allele per generation loses its significance while that for ε2 remains positive. We also checked for the expected changes in longevity when we vary the environmental parameters of the model which has provided us with a useful way to test the effect of these changes and get a quantitative view of the benefits or losses to the mean lifespan.

Up to now the fact that APO E has an effect beyond the reproductive lifespan of humans has posed an interesting puzzle for the evolution of the gene. As we saw earlier, one of the prevailing views is the one described by Finch and Sapolsky (1999; 2000) based on the grandmother hypothesis (Hawkes et al., 1998). The grand-mother

hypothesis is based on the idea that the mother-child food sharing seen among hunter-gatherers may allow post-reproductive grandmothers to enhance their daughters' fertility thus elevating their own fitness and increasing the selection for long postmenopausal lifespan. Finch and Sapolsky argued that this evolutionary advantage will have caused selection for an *apo E* allele that will have delaed neuropathology and mortality. Although this is a compelling hypothesis, despite some problems with the underlying theory (see Kennedy, 2003), which may work at a secondary level to increase selection for the *apo E* alleles, our model showed that it is not strictly necessary. According to our simulations the direct effect of *apo E* on the mortality or morbidity of the population is sufficient, at least in contemporary Western populations, to produce a differential of selection between the alleles in a wide number of different situations.

Our results clearly indicate that although the main effect of *apo E* is beyond the end of the reproductive lifespan of humans, the relatively rare events of early mortality or morbidity are enough to produce selection against the ε4 allele. Considering these findings in the light of mutation accumulation theory, we can see a new important aspect of this concept. Since most longevity associated genes, such as *apo E*, will cause a distribution around a mean for the corresponding mortality, evolution will very rarely, if ever, manage to push a gene action entirely out of the reach of selection. Most often the effect of the gene will be gradually moved to older and older ages. Furthermore this action will take place simultaneously in many genes since if a single mortality distribution is exposed significantly more often to the action of selection, it will be moved towards older ages faster, until it reaches a balance with the rest of the mortality genes. This process will continue to drive the evolution of the longevity of the organism until the age-related mortality becomes non-significant in

relation to the age-independent mortality. The above view of the evolution of longevity can account for the fact that in the wild very few animals die of natural causes as well as the difference between longevity in the wild and in captivity. In addition it can explain how the same genes in different but closely related species can produce big differences in longevity. In this respect the evolution of longevity, although based on selection of individual genes, is more of an evolutionarily coordinated transfiguration of the whole organism to achieve the maximum lifespan in the given conditions (for similar ideas see Hamilton, 1966). Based on the above we can see how common age related diseases today, with late age of action, could have played a role in our evolutionary history even if their action is currently much beyond the expected lifespan of our early ancestors. Cardiovascular disease can be considered as an example of just such an evolutionary process. All primates seem to follow the same basic rules for the evolution of late acting diseases, with differences in their sensitivity towards various environmental and physiological parameters (Rudel et al., 1998). If they live in a protected environment and are exposed to similar diets to those of contemporary humans, they will develop similar age-related diseases (Kemnitz et al., 2002). So, can we apply our model to make predictions about the past evolution of the *apo E* gene? If we assume that cardiovascular disease played a role even in early humans, which as we saw earlier is possible, and the relative risks are comparable with those seen today then we can cautiously use the model to get an indication about the presence or not of the selection in our early evolutionary history and its approximate intensity.

Assuming that the ancestral allele was ε4, since this is found in almost all other primates, and considering the prevalent view that ε3 originated from a C-T transition at codon 112 while ε2 arose from another C-T transition at codon 158 from

the ε3, we can use the changes in frequency per generation predicted by the program to obtain an indication of the evolutionary history of each allele (Finch and Sapolsky, 1999; Mahley and Rall, 1999). As we saw earlier in a population showing high consumption of fat, according to our model, the expected change of the ε2 allele per generation will be 0.0000252 while for ε3 it will be 0.0000215. Given the current prevalence of the gene in Europe of 0.0845 for ε2 and 0.7415 for ε3 and assuming that the conditions remained comparable to the parameters used in the final run of the model, then we will expect to have an evolutionary history of around 60,000 years for ε2 allele, with a minimum of 34,000 and maximum 350,000 years, and 600,000 for ε3 with a minimum history of 265,000 years. Of course conditions have changed significantly during our evolutionary history so that the linear process assumed in our calculation is a simplification of the problem. What is instead expected is a nearly exponential increase of the selection for the gene, increasing very slightly in our early ancestors. As cardiovascular disease becomes more important, the increase becomes bigger and bigger. Furthermore we do not know the exact fraction of the population affected by the action of the gene. We assumed a 10% ratio while we used a diet of meat. Thus we probably overestimated the very early levels of selection. According to these observations we can hypothesise a history for the human *apo E* gene in two stages. The first stage starts with the change of the ε4 from its ape form to the one found in humans today, either due to an unknown advantage conferred by this isoform or more probably due to a bottleneck in our early history. Later as early hominoids started to add protein and fat to their diets, probably by scavenging subcutaneous fat, bone marrow and brains from discarded kills, the previous mutations of ε4 to ε3 began to be selected. The intensity of selection would have increased with increasing brain size and reliance on meat and of course longevity (Finch and Stanford, 2004).

During this time, as the ε3 allele increased in frequency, a mutation of it arose that conferred even better protection from the adverse effects of meat eating, the ε2 allele. This time can probably be characterised by a linear change in allele frequencies. The second stage probably started with the beginning of agriculture where humans were now able to live a much more secure life, increasing their lifespan and the selection on the alleles of *apo E*, most likely starting an exponential rate of change for the allele frequencies. Furthermore agriculture also brought a decrease of health, mainly due to overcrowding and high level of carbohydrate in the diet, with the first significant appearance of diseases common today, such as diabetes and, importantly, diseases of the vascular system (Richards, 2002). We are probably still in the second stage of the gene selection which will end either by the extinction of the ε4 allele or with a balance between the three alleles due to some as yet unknown pleiotropic effect of them. A two-stage evolution of the gene may explain the global differences of the alleles. So with the first stage we can account for the high frequency of the ε4 allele in Africa, where we expect the selection to be present but much attenuated by external mortality and the relatively recent use of agriculture, while the second stage can explain the north-south gradient of the ε2 allele in Europe coinciding with the spread of agriculture.

Of course all our work is based on our current knowledge and understanding about the operation of the *apo E* gene. Further work is needed to elucidate the precise quantitative relationship of the *apo E* gene with the parameters of the model so that our predictions can become more accurate. In that respect, the simulation described can work as a framework of a gene-environment interaction where new data can be added easily to see their effect on the longevity of the population and selection applied on different alleles. Nevertheless, from our discussion already we can see that

although the *apo E* gene is known to be a part of the lipid transport system in the body it may worthwhile to ask what is its role in cases of caloric restriction, a high carbohydrate diet and dietary related diseases.

To summarize, our results show that *apo E* is selected in a normal western population even when different lifestyle parameters change. One of the most significant changes made in the model was smoking and how this affects the action of *apo E*. We were also able to elucidate some new aspects of the mutation accumulation theory and how it might operate, moving distributions of gene effects during the selection for longevity. Additionally we gave an example of how an extreme value distribution can describe an altered form of the Gompertz model. Finally we tried to answer how the *apo E* evolved and we proposed a two-stage process in relation to meat eating and the spread of agriculture.

5. The Effects of Social Class in Longevity

There is little doubt that ageing is the product of a set of complex interactions between our genes and the environment with time. In this Chapter we will explore how social class in human societies might affect the ageing process of an individual and ultimately its longevity. Examples will be drawn from a wide variety of sources. Furthermore, our previously developed models will be considered in the context of living conditions within an unequal society while an analysis of global data will be used to illustrate our main conclusions. The genetic and non-genetic factors involved in the heterogeneity of ageing between castes or social groups will be addressed. Finally we will explore the effects of inequality of longevity on the evolution of human populations and ask whether we will expect to find evolutionary selection differences between social groups.

5.1 Introduction

Most commonly the definition of society refers to the patterns of relationships among individuals within a definitive territory (Scupin, 1992) and is free from any necessity for a social stratification. The emergence of social classes, as sub-groups of the population given a specific role in society, is the result of a social structure where the relationships among individuals are driven by transfers of resources between specialist groups. These groups share common attributes and dexterities and interact in particular ways with other groups in the society. The nature and balance of the interactions between the different classes are central in political theory and have been the subject of rigorous investigation through the disciplines of sociology, philosophy

and economics. In our current work we will concentrate only on the effects of social status and its measures in relation to public health and ageing.

Inequality in health between different social groups is a fact long recognized. Frederick Engels (1845) described disturbing pictures of urban mid 19th century Britain, where inappropriate dwellings, poor hygiene conditions and malnutrition, were widespread among the working class. These resulted in 57% of children in the poorest classes dying before the age of five, compared to 20% in the most affluent families. Fatal cases of small-pox, measles, scarlet fever, and whooping cough, in children living in industrial cities were four times more frequent than the national mean. Children were reported working as young as eight or nine years old. Engels also cites a Dr Hey describing the deformities of the spine and limbs due to overwork and poor nutrition, while Dr Hawkins (the medical commissioner for Lancashire at the time) described the factory workers of Manchester as surprisingly short, lean and pale with a high percentage of alcoholism. Others reported that most men were unfit to work after 40 years of age and they were by then considered as too old. Women were forced to work up to delivery and to resume 3 to 4 days after giving birth for fear of being laid off, while young girls working in factories very commonly exhibited menstruation problems (Engels, 1845).

5.2 Markers of inequality and their link to health and mortality

Despite the fact that in developed countries today this kind of situation sounds alien and unacceptable, it is sadly a condition that still exists, to some extent, in some developing countries. Even in today's advanced economies there is significant inequality in health and longevity between social groups. Many recent studies have

found that there is a correlation between markers of social status and lifespan. Usually education, income and occupation are used as indicators of socioeconomic level and their association with morbidity and longevity is tested in a sample of the population. Each marker of social class though, captures only a part of what constitutes our status in society. Education refers to the early part of life, it is sometimes linked with the status of the parents and it is also predictive of later success. Occupation is a marker of the conditions of adult life and a link between education and income. On the other hand income relates directly to material conditions at the specific time and can be independent of occupation or education. Income can affect health directly in the short- or long-term since it is the key factor in determining access to better housing, food, health care, recreation etc (Avlund et al., 2003). Regardless of their differences, the markers discussed can give consistent and reliable associations between social class and inequalities in health and lifespan as long as their specific properties are kept in mind. Bopp and Minder (2003), using data from the Swiss National Cohort, found that mortality from age 25 to 85 is negatively correlated with the level of education for both men and women. Moreover for males the differences seem to decrease with ageing so that the mortality curves converge at around the 89th year, in accordance with the majority of literature; while in the case of women the difference is much more difficult to observe.

Although health status is a much more subjective measure than lifespan, in most studies higher social status means not only a longer life but also a longer health expectancy, with the differences between social classes being greater than that observed for longevity (Bronnum-Hansen, 2000). People in lower classes, in addition to living a shorter life, are expected to also spent more years in ill health, both in absolute terms and as a percentage. For example, for a 25 year old Belgian with the

lowest level of relative education, out of the 43 years of further expected longevity it is predicted that 25 will be in good health and the remaining 18 in perceived poor health, in contrast to those at the highest level where out of the 46.5 years only 4.8 are expected to be in perceived poor health (Bossuyt et al., 2004). Of course the question remains whether individuals with a higher education consider their health good even when someone with a lower education will consider it poor. Generally, when the calculations are based on perceived health, the health expectancy is higher in all socioeconomic groups than it would have been if longstanding illness was used (Bronnum-Hansen, 2000), but no evidence exists that this overestimation is much higher in specific parts of the population. Further evidence for the differences between social classes are provided from the work of Huppert et al (2000) where event-based prospective memory, the ability to remember to carry out indented actions, at ages 65+ was negatively associated with less education and lower social status irrespective of the objective perception of health of the participants.

Not all studies though are able to observe this inequality in health between social classes. Notably a study in an elderly population (mean age 80 years) in Hong Kong showed no association between income after retirement and lifespan, which led the authors to attribute the lack of correlation either to the survivor effect in the cohort used, or to the well developed and easily accessible health and social services in this Chinese province (Woo et al., 2000). What the writers overlooked in this case though, is the fact that the mean age of the population used in their study was very close to the crossover point of mortality between a high and low mortality group. As was shown by Bopp and Minder (2003) the differences in longevity in this area are much reduced compared to that seen earlier in life. Nevertheless, the characteristics of a society or of an economic program can alter, up to a point, the level of inequality

between privileged and disadvantaged members of a population, but sadly not to eliminate it. This can be seen in the work of the EU Working Group on Socioeconomic Inequalities in Health, where 11 western European countries were compared based on their inequalities with respect to morbidity and mortality (Mackenbach et al., 1997). Despite the methodological problems arising from historical differences in the timing of the epidemiological transition between European countries, and differences in the causes of mortality between countries, it was found that France has the highest inequality in mortality between social classes while the Nordic countries exhibit the highest degree of inequality relative to morbidity. Generally within Europe a north to south gradient of inequality has been reported, with Mediterranean countries showing much smaller differences between social classes (Mackenbach et al., 2000). The authors proposed that "open" societies, where social position exhibits a low "heritability" and is more affected by personal characteristics, will lead to a larger inequality in health (Mackenbach et al., 1997).

A higher social status is not always beneficial for health. In developing countries the trend between class and health is commonly reversed. Jamaica, a middle income developing country, shows a complex association between income, education and hypertension, with both the low- and high social classes having elevated blood pressure relative to the intermediate groups (Mendez et al., 2003). In China, higher social status was linked with an unfavourable serum lipids profile (Yu et al., 2002). In these cases, a high socioeconomic status is linked with an early adoption of westernized lifestyles characterized by smoking, sedentary lifestyle, and a high fat diet. Generally, distinct groups of the population progress through epidemiological transitions at different times, with a different starting point and a different rate, with the more affluent classes starting earlier and progressing much

more rapidly (Gulliford, 2003). In some more recently industrialized countries the after-effects of a socioeconomic change can still be seen in the prevalence of diseases linked with lifestyle. A study in Barcelona, Spain, showed that although total deaths were negatively associated with employment status, cardiovascular mortality in men was more common in the higher-level professional group (Borrell et al., 2003). These effects can be addressed with the "diffusion theory", where the higher socioeconomic groups in affluent countries were the first to afford the change to an unhealthy lifestyle which then spread to other populations, lower social groups or poorer countries, until the more well off started a second transition to a more health-conscious lifestyle (Mackenbach et al., 2000). A very good example of the above is the variation in tobacco consumption and physical activity among social classes in Spain, where it was found that smoking habits passed through a series of four diffusion states from their adoption in the highest classes to their spread in the disadvantaged ones (Borrell et al., 2000).

5.3 Non-communicable disease and social status

As we have already mentioned, cardiovascular disease and cancer incidence contribute to the differences observed in mortality and morbidity between social classes. Generally, non-communicable diseases, largely heart disease, stroke, cancer, diabetes and obesity, are the major factors in the inequalities in health observed in both developed and developing countries. It has been shown that greater increases in systolic blood pressure and decreases in diastolic blood pressure are linked with lower socioeconomic categories after the age of 50, while low social class also increased the risk of hypertension (Roux et al., 2002). Prevalence of diabetes was also found to

increase with decreasing income and education in Denmark (Tang et al., 2003). Nilsson et al (2003a), investigating five biomarkers of biological ageing, concluded that men and women from a higher social class appeared to be biologically younger than individuals of the same age but lower class. The recent increase of people living into old age and the low mortality in infants and young adults have contributed to elevating the importance of non-communicable diseases which are mainly age-related. The overall risk for the group is dependent on the past and cumulative risk experienced by the specific group (Beaglehole and Yach, 2003). Since disadvantaged social classes seem to be less interested in their future health (Roos et al., 2001) socioeconomic status is one of the main determinants for the risk factors of non-communicable diseases. So, particularly in the north and west Europe, those with higher education consume more fruit and vegetables than people with lower education (Roos et al., 2001). At the same time, intake of saturated fats is increasing with occupation category further elevating the risk factors for cardiovascular disease (Lopez-Azpiazu et al., 2003). Lower social status individuals gain weight more rapidly than those with a higher status, leading to overweight and obesity, one of the most important emerging dangers of public health (Martikainen and Marmot, 1999), while increased incidence of smoking and drinking leads to a pronounced further accumulation of unhealthy behaviour in lower educational groups (Prattala et al., 1994).

5.4 Social status and psychological stress

Here we will also briefly address the idea that inequality in health can be, at least partly, attributed to psychological factors. Under this hypothesis one of the key

factors of social position is social participation and control of one's life. The latter can be clearly seen in the example of occupation, where jobs with high psychological demands and low control put people at high risk for cardiovascular disease (Marmot, 2003). Psychological distress is a well known factor is coronary heart disease, as was exhibited among others, by Stansfeld et al (2002) in a cohort study of London-based civil servants. Power of control thus, can be an important factor in translating social status to health (Marmot, 2003). Similar phenomena have been observed in non-human primates, those with a low status in the group exhibiting signs of stress with enlarged adrenal glands, exaggerated cortisol response, changes in dopamine metabolites concentration and most importantly increased atherosclerosis (Kaplan et al., 2002; Marmot, 2003). Insufficient participation in society can also lead to elevated psychological stress, and can be perceived as the comparison of an individual's ability to own articles that have a social status currency. Socially oriented items and items of luxury have been both related to health, independently of economic hardship (Marmot and Bobak, 2000).

5.5 Environmental conditions and longevity

Differences in health between social groups may extend much further than just differences in behavioural characteristics, lifestyle preferences and psychology. Exposure to even mild environmental stress may contribute to even further widening of the inequalities in mortality between classes, especially for members of vulnerable groups such as the old. A characteristic example is the link between temperature and mortality. Increased temperatures increase demands on the cardiovascular system, elevating cholesterol levels and blood viscosity, whereas low temperatures are linked

with influenza epidemics, increased blood pressure and fibrinogen levels (Curriero et al., 2002). It was estimated that a five day heat wave in England during 1995 (max temperature 35.2°C) was responsible for an 8.9% increase in mortality with an even larger effect in the Greater London area (Rooney et al., 1998). Also in 1995, a heat wave hit Chicago from the 12th to the 16th of July causing at least 700 excess deaths, mainly associated with cardiovascular diseases (Semenza et al., 1996).

In most countries, though, the peak of the mortality rate is observed during the winter months. Cardiovascular disease and ischaemic heart disease account for half of the excess winter mortality mainly attributed to thrombosis due to haemoconcentration, increased plasma fibrinogen and endotoxin inhibition of fibrinolysis. Nearly half of the remaining surplus mortality is due to respiratory diseases, mainly because cold suppresses, through stress hormones, the resistance of the immune system to respiratory infection and assists the survival of bacteria in droplets (Keatinge et al., 1997). Aylin et al (2001), over a period of 9 years, found an average of 30,000 excess cold-related deaths per annum in Great Britain, making cold a significant public health problem. One might think that living in countries with milder winters might avoid the effects of excess winter mortality. Paradoxically this is not true and Mediterranean countries show larger seasonal changes than Scandinavia (reviewed in Mercer, 2003). For a given fall in temperature, mortality increased to a greater extent in countries with milder winters, probably due to the fact that in warmer regions people live in colder houses, wear less protective clothes and are less active outdoors (Keatinge et al., 1997). Also the fact that these populations are acclimatized to a different set of conditions should not be overlooked. Populations in warmer regions seem to be more vulnerable to cold, while those living

in colder regions are much more sensitive to elevated temperatures (Curriero et al., 2002).

But is exposure to temperature linked with socioeconomic status? It is easy to imagine that disadvantaged groups will be more exposed to temperature stress than the affluent classes. Studies in Brazil (Gouveia et al., 2003) and Britain (Lawlor et al., 2000) found no association between deprivation and temperature-related mortality. This can be attributed to some methodological problems of the first paper (see Keatinge, 2003) and the complex effects of council housing in Britain and the confounding factors of deprivation in the second case. Aylin et al (2001) were also unable to find the association between quality of dwelling and excess winter mortality, but at the same time they found that it was linked to the lack of central heating calling for further investigation of the relationships. Certainly the association between low temperature and excess winter mortality is a complex one and although lower indoor temperature increases the effect (Keatinge et al., 1997), behavioural means of minimizing heat loss also play an important role both in- and outdoors (Mercer, 2003). Furthermore the quality of insulation or status of the house may not necessarily be good indicators for the home-heating habits of the occupants. For extreme heat, the association of increased mortality with social class is much easier to see. The excess number of deaths in the Greater London area in the 1995 heat wave can be attributed to the "heat island" effect in urban densely populated areas, high air pollution and poor housing, all characteristics of a disadvantaged environment. Furthermore in the heat wave of the same year in Chicago, living conditions and especially the presence or absence of air-conditioning were very important for the differences in mortality with increasing temperature (Semenza et al., 1996).

Generally, low education level and poverty have been associated with increased mortality during high temperature (Curriero et al., 2002).

5.6 Early development and phenotypic plasticity

The effect of socioeconomic status on health and mortality is an example of the interaction between the genome and the environment. We have treated this interaction as a direct association, where the interplay between the two factors is based in the real-time conditions experienced by a fully matured organism with a given set of responses. But the balance of these responses between the genes and the environment can also be significantly affected by both the environmental challenges during development and the expectations of what these challenges will be. Developmental plasticity and maternal effects are two major factors influencing the phenotype of the individual and its ability to express the right responses in accord with the environmental cues it receives.

Phenotypic plasticity, the variation of phenotype of a given genotype, due to early developmental or maternal effects has been extensively studied in insects. Social hymenoptera are an extreme example, where from a single genotype a queen or a worker can develop. These alternative morphologies are the product of different rearing conditions between individuals with otherwise identical potentials, causing large differences in lifespan and reproductive ability (Page and Peng, 2001; Chapuisat and Keller, 2002). This kind of polyphenism is the result of differential gene expression due to hormonal regulation (Evans and Wheeler, 1999). Similar mechanisms can also be found in other insect species where division of labour exists, as in weaver and leaf-cutting ants (Keller and Genoud, 1997; Hughes et al., 2003).

Less extreme examples of phenotypic plasticity due to pre- and post-fetal events have been observed in model organisms, such as the *D. melanogaster* and *C. elegans* and in mammals and birds (Lindstrom, 1999; Ackermann et al., 2001; Dillin et al., 2002).

When it comes to humans and specifically the effect of socioeconomic status on longevity, we will focus our attention on the effects of undernutrition during early pre- and post-fetal development on the health of the individual when adult. Different ageing markers have been associated with weight at 1 year old, leading Sayer et al (1998) to question if rates of ageing are determined in utero, concluding that the association between early growth and ageing are causal, with some systems being programmed in early life. A number of age-related diseases are also linked with early development and maternal effects. Cardiovascular disease is associated with small size at birth, which in turn is related to maternal bodyweight (Stein et al., 1996). Coronary heart disease also seems to be associated with low weight gain in infancy, and moreover with rapid childhood weight gain in boys that were thin at birth (Eriksson et al., 2001; see also for commentary Osmond et al., 2001). Studies in Britain have revealed that low growth rates increase the prevalence of cardiovascular risk factors including blood pressure, and plasma concentrations of glucose, insulin, fibrinogen, factor VII and apolipoprotein B (reviewed in Barker et al., 1993). Diabetes and glucose intolerance as well as insulin resistance in adult life have also been linked with abnormal fetal growth, both for underweight and overweight fetuses, in North American and European studies (Phillips, 1998). Further effects of undernutrition during early development include lowering of immunocompetence in adulthood through the early programming of thymic function (McDade et al., 2001) and increase of obstructive airways disease in adulthood due to permanent effects of the structure and physiology of airways during periods of rapid growth (Lopuhaa et

al., 2000). The above have given rise to the "fetal origins" hypothesis which proposes that the nutrient and hormonal environment of the fetus can alter gene expression, resulting in developmental adaptations leading to permanent changes in physiology and metabolism (Godfrey, 1998). Evidence for the theory is provided by the seasonality of births observed in humans, as in other animals, and its effects on adult health. Seasonal patterns of nutrition and disease in rural African societies increase environmental stress which can lead to intra-uterine growth retardation and reduced birth weight (Moore et al., 1997). Mortality in young adults due to fatal infection in these groups shows a strong correlation with the season of birth, so that babies born up to two to three months after a hungry period have a carry-over effect even if maternal nutrition improves later in pregnancy (Moore et al., 1997; 1999). Similar effects can also be seen in Europe where those born in autumn have a higher life expectancy at age 50 than those born in spring (Doblhammer and Vaupel, 2001). Since these studies are based mainly on cohorts born in the early part or the first half of the 20th century we can see that the effect is decreasing significantly over time, in accordance with the lack of seasonal patterns in access to nutrients in contemporary Europe (Lummaa et al., 1998; Doblhammer and Vaupel, 2001).

Early developmental events not only affect the health and mortality of an organism, they can also have a profound effect on its ability to reproduce and may even affect subsequent generations. This would have been expected if adverse conditions early in life could affect organs producing or regulating reproductive hormones (Lummaa, 2003). Women with reduced early growth and low weight give birth to small babies with an increased risk of pregnancy complications and early mortality, while women born underweight have a higher chance of stillbirth or of losing their baby within the first days (Lummaa and Clutton-Brock, 2002). A study

from a British cohort showed that although birth weight was not associated with age at follicle depletion, women with lower weight at 2 years of age had earlier menopause, highlighting the influence of early post fetal development with reproductive ageing (Hardy and Kuh, 2002). Furthermore in 19th century Canada, Lummaa and Tremblay (2003) found that women born in the best months of the year (November-March, June and September) had a longer reproductive lifespan, larger number of live births and more children raised to adulthood. Lumey and Stein (1997), on the other hand, observed that women born in the five month famine in Denmark in 1945, show no evidence for a fertility-reducing effect of early exposure to undernutrition, although these women seem to have an excess of perinatal deaths.

Although the exact mechanisms are not known, it is believed that early nutritional environments can be "memorized" by the organism over a narrow window during ontogeny. We will follow the reasoning of Waterland and Garza (2002b) in describing the process as "metabolic imprinting" rather than "programming". This way our held view that lifespan cannot be programmed or determined is satisfied and at the same time we can describe a mechanism where early experience may have a significant effect later in life (see also Lucas, 2000; Waterland and Garza, 2002a). The two necessary characteristics for any such mechanism to be credible are that it must be adaptive to early environmental conditions and that it must occur during a limited period of susceptibility (Waterland and Garza, 2002b). The precise stage of pregnancy during which undernutrition occurs will lead to different effects on the organism. So, early pregnancy undernutrition will probably result in symmetrically small low birth-weight babies, while undernutrition in mid-pregnancy can change the interaction between the fetus and the placenta. Finally, low nutrients in late pregnancy could lead to fetal wasting as energy is diverted from the fetus to the

placenta to maintain its function (Barker et al., 1993). Detailed physiological mechanisms through which early undernutrition may affect adult lifespan could involve: 1) morphological variations in organ structure through different inductive interactions among different germ layers during organogenesis; 2) differences in cell numbers, where growth is achieved instead of the normal cellular proliferation, which is sensitive to nutrient availability, by increase of cell size (hypertrophy); 3) disproportional population growth of the most proliferating cells during clonal selection in conditions of some limiting nutrient factors; 4) metabolic differentiation, changing the capacity for adaptive and basal gene expression through DNA methylation, modulation of chromatin structure and changes in the autoregulation of DNA binding proteins; and 5) changes in signals determining time and extend of polyploidization of vital organs (for full description see the excellent work of Waterland and Garza, 2002b).

At least in industrialized countries, we do not expect to find such extreme differences between social classes, where pregnant women and babies are exposed to famine and undernutrition, while our previous discussion seems more appropriate for countries of the third world or the previous experience of Europe. But recent experimental studies have shown that even subtle differences of early nutrition can have a significant effect later in life. What is more important is that these differences are usually linked with the social status of the parents. Smoking, for example, during pregnancy has been associated with restricted fetal growth, a habit that is more common in people of a low social class (Moore and Davies, 2002). In India, mothers with low weight during pregnancy had a higher chance of having a child who developed cardiovascular problems when adult (Stein et al., 1996). But the opposite is also true, maternal obesity, another current marker of low status, also has

detrimental effects for the health of the child later in life (Godfrey, 1998). The father's social class is strongly correlated with the social class of the child in adulthood, and independently it was also linked to non-fatal myocardial infarction and ischaemic heart disease, probably through the persisting effect of class early in life (Wannamethee et al., 1996). Probably the most important effect of social status and later health is the quality of food received by the mother. A Scottish study has shown that a high-animal protein, low-carbohydrate diet is linked with raised blood pressure in the adult offspring, while other studies found that a low ratio of protein to carbohydrate and fat is also linked to lifelong elevated blood pressure (Godfrey, 1998; Shiell et al., 2001). In general, imbalance of maternal body composition during pregnancy is associated with unfavourable outcomes in the child's later life (Godfrey, 1998). Keeping in mind the specific dietary habits and lack of proper health advice, either through lack of education or access to the relevant information, commonly found in lower social classes, it is easy to understand how an adverse early environment can affect the health of the population. This is not to say that the differences described up to now are solely due to maternal effects. On the contrary, we chose the term of imprinting to describe a process that very early in life will shape the way the organism responds to its environment, so that the early experience will make the individual more or less susceptible to later challenges. Metabolic imprinting and maternal effects in general will provide only the background where our future decisions and lifestyle are going to interact with our genetic make-up to increase or decrease the chances of events.

5.7 Social status and access to health services

We have considered the differences between health and mortality of social classes, but until now we have avoided discussion of equity in access to the health services and ultimately treatment. In today's industrialized countries social development is closely linked to provision of health care to all members of the society with no class distinction. The two prevailing model strategies for social sector development are aimed directly at overcoming health service inequalities in an advanced market economy (Bloom, 2001). But the truth is that inequality seems to exist even in the most developed health care models. In Kentucky, USA, patients with either colorectal, lung, breast or prostate cancer on private insurance had the best survival, in contrast with those whose insurance was unknown who fared worse (McDavid et al., 2003). Considering that private insurance is a marker of income and occupation, the results are in accordance with the majority of studies in USA, Europe and Canada where patients from deprived areas had lower survival compared with those living in more affluent areas (McDavid et al., 2003). In the case of Spain, where the National Health Service (NHS) provides universal and free coverage, inequalities can be found in services that are only partially covered by the public health care system. Additionally, those that paid for a private service had to wait on average 18.8 minutes less for treatment than NHS patients, and even among the latter group, those of a low social class waited longer than the more privileged individuals (Borrell et al., 2001). In Finland during the increase in the supply of services from general hospitals in the 1990s, the lowest income groups gained access to fewer newer services than those belonging to higher classes (Keskimaki, 2003). A study involving NHS Direct, a health advice and information service in Britain, in an area of southeast London, including areas of extreme deprivation and relative affluence, found that call

rates increased with deprivation up to a point, then extreme deprivation reversed the pattern (Burt et al., 2003). It seems that, in developed countries, the worse-off groups receive more hospital care and see general practitioners more often, whereas specialist visits and surgical care, in almost most cases, are more accessible to higher socioeconomic groups (van Doorslaer et al., 2000; Keskimaki, 2003). Although a clear picture of inequality in access and utilization is emerging from the above, it may actually be quite misleading in terms of spread of the variation and significance of the results. The truth is that we know very little about the causes of these differences and their underlying patterns. As we saw earlier the needs for health care between social groups are very different, both in terms of incidence of poor health and frequency of diseases. Access and utilization may sound quite solid as measures of use, but in fact they do not give much information on the outcome or the quality of treatment which ultimately can increase the inequity (for a more detailed discussion of the problems see Goddard and Smith, 2001).

5.8 Socioeconomic inequalities and evolution

The rest of our discussion will focus on the question of whether and how social structure might affect the evolution of human populations. Evolution is commonly defined as the change of allele frequencies within a given population, although other definitions are also available. In order to have evolved differences within a single population, reproductive isolation is essential to split the original group into smaller compartments. In time, genetic drift and difference in the environment experienced by each smaller group will lead to a divergence between their genotypes. If the isolation is long enough it can lead to speciation.

H. G. Wells, in his book The Time Machine, inspired by the social inequality in the late 19th century Britain described a terrifying future, in which the delicate Eloi and the sub-terranean Morlocks are the products of thousands of years of evolution of the aristocrats and the working classes respectively. But is that science fiction or scientific probability? Can social class differences lead to gene changes through evolution or even speciation between classes in humans? In our species the degree of variability both between and within groups is so high that it is extremely unlikely, or just impossible, for something like that to occur. The entire *Homo sapiens sapiens* species is a single breeding population, especially now when modern ways of travelling contribute to the mixing of the previously localized populations (Chakraborty, 2003). On the other hand, caste systems in traditional societies, and segregation between social classes in more recent cases, that put barriers against the free flow of genes within the population, can lead to a low degree of genetic differentiation. This kind of isolation is usually linked with groups of a different religious or ethnic background relative to the wider population, the best examples being the Ashkenazi Jews and the Amish groups (Mitchell et al., 2001; Behar et al., 2004). Nevertheless, isolation due to social criteria has also been found. The traditional caste system in India and the European royal families are two good examples (Papiha, 1996; Stevens, 1999). Even in these cases the group is not completely isolated, and as is the case in other human populations (Bereczkei and Dunbar, 1997) women are allowed to marry up the scale, thus maintaining the gene flow with the background population. Differences in the cases of isolation mentioned, though, are the product of random genetic drift and inbreeding, and we are still far from finding any adaptive changes of allele frequencies between social groups. In truth very little work has been done towards finding any such case, but the frequent -

in evolutionary terms - changes in the economic situation and the structure of our societies, are probably a contributing factor in the homogeneity of gene frequencies between different socioeconomic groups.

Moreover we discussed the effects that early development and maternal effects can have in the health and lifespan of the individual, and we related those with socioeconomic status. The association between the genome and the environment is very complex, with a single genotype able to produce more than one phenotype. Since natural selection is applied at the phenotypic level but its effect are inherited at the genotypic level, it is difficult to predict how selection will operate in relation to social class (Mousseau and Fox, 1998). Two opposite possibilities can be envisaged in this case: 1) maternal effects wrongly predict the future environment, for example presenting a poor nutrient environment that changes in energy abundance later in life, causing insulin resistance due to early undernutrition, and development of diabetes in adulthood (Phillips, 1998), and 2) maternal effects accurately predict future conditions of nutrient availability. In the first case, the effect of a disadvantageous allele will be amplified by the outcome of development carrying a severe selective penalty, while in the second case the organism is already phenotypically adapted to handle the excess energy, with its genes being less severely selected against, even if they are not optimal.

5.9 Models of human longevity and social status

We have earlier presented two different models for the evolution of longevity and its association with the genes and the environment experienced by the population. The first model was based on the disposable soma theory of ageing (Kirkwood and

Rose, 1991) and involved an immunogenetic trade-off between the mother's ability to complete a pregnancy and her ability to withstand infection. In terms of socioeconomic status in industrialized countries, the trade-off seems to be not applicable since deaths from infections are rare. Despite that, since evolution progresses much slower than human history, it will be interesting to test whether social classes that originated from the working classes of the previous century have different frequencies of the genes involved in the trade-off compared with the old aristocracy. The case should be very different in developing countries, where deprivation significantly increases the incidence of life-threatening infection and the presence of the trade-off should be clear and associated with socioeconomic status. The same should be the case for the disposable soma theory in general. Better survival and health, together with a much more favourable early development should be able to dampen the effects of the trade-off since the selection for the genes responsible decreases with decreasing selection for survival, while environmental parameters during early life and adulthood become more important. If this is the case, then the result obtained by Lycett et al (2000), where the longevity-reproduction trade-off was increasingly strong with increasing poverty, can be explained, as well as the fact that pre-1700 women were mainly responsible for the relationship found in the paper from Westendorp and Kirkwood (1998).

With our second model we examined the effect of the *apo E* gene on cardiovascular disease, the major non-communicable disease in the western world and a rapidly increasing problem in developing countries. In this case the gene-environment interactions were the central factor linking the gene with morbidity or mortality. All the risk factors used for cardiovascular disease are also linked with social status, income and education. It is easy to see how social class will affect the

progression of the population and if the conditions remain stable, the segregation is long enough and the gene flow small, it will ultimately produce two groups with quite different allele frequencies. Unfortunately our model cannot account for the presence of maternal effects and early development that may disturb our predictions.

5.10 Longevity and global data

One of the most straightforward tests to see how longevity can be affected by social markers is to analyse the available global data. In this case we used information extracted from the *"The World Factbook 2004"*, a publicly available document profiling all countries in the world based on a number of different parameters ranging from political situation to the age-structure of the population (http://www.cia.gov/cia/publications/factbook/index.html). In our current analysis we used 10 different measures of reflecting social stratification in the population, including population, birth and death rate, infant mortality rate, total fertility rate, country GDP, GDP per capita, unemployment, internet users per 10,000 population, and %GDP spent in military expenditure. The Minitab statistical program gives us the ability to select the "best subset" of variables with the higher coefficient of regression and the minimum number of variables using a number of regression analyses. During the procedure the program examines all the one-predictor models and returns the two with the highest value of the coefficient of determination R^2, then all the two predictor-models are examined and the two with the largest R^2 are displayed, the process is repeated until the model contains all predictors. Using life expectancy as the response and running a regression test for the best subset of variables able to describe the changes of longevity we identified that two of the five-

variable models can explain 94.5% of the total variation of life expectancy in the population (Figure 5.1). Increasing the number of the variables used, had no effect on the adjusted R-square value of 94.5%. Thus one of the two five-variable model including birth, death, infant mortality, and total fertility rates plus one of GDP per capita or Internet users is adequate, while the rest of the parameters do not provide additional information. To distinguish between the two alternatives we ran a regression analysis for each model comparing their F-values, and a regression analysis for the model fitting all six of them. From those we can conclude that the subset containing GDP-per capita gives a better description of life expectancy both due to a higher F-value in comparison with the alternative subset regression and retaining its significance in the six-variable model, in contrast to Internet users per 10,000 of population. A correlation analysis between life expectancy at birth and GDP-per capita revealed that there is significant correlation between them with a Pearson coefficient of 0.606, thus an increasing GDP per capita is associated with a longer lifespan. The population, birth and death, infant mortality, and total fertility rates included in the best subset model describe the demographic state of the population and given that the effects of social status on longevity are dependent on demographic transitions, as discussed earlier for the changes of disease risk and income between developing and developed countries, the close association was expected. The variables Internet users and GDP per capita, on the other hand, describe access to information and goods in general respectively. In the case of the per capita GDP, which gave the best fit, what is measured is the purchasing power of the individual in its country of living, thus its ability to consume and get access to goods and services. The correlation between Internet users and GDP per capita (the latter explaining more than 65% of former), can partially explain why both of them can give models of

Best Subsets Regression: Life expecta versus Population, Birth rate, ...

Response is Life exp

136 cases used 89 cases contain missing values.

Vars	R-Sq	R-Sq(adj)	C-p	s	Populati	Birth ra	Death ra	Infant m	Total fe	GDP	GDP-pe	Unemploy	Internet	Military
1	76.5	76.3	440.6	5.7566				X						
1	63.4	63.1	759.3	7.1822			X							
2	90.7	90.5	97.0	3.6386		X	X							
2	89.9	89.7	116.9	3.7940			X	X						
3	93.4	93.2	33.2	3.0774		X	X				X			
3	93.3	93.1	35.3	3.0978		X	X		X					
4	94.2	94.0	14.8	2.8871		X	X		X				X	
4	94.2	94.0	15.3	2.8922		X	X		X		X			
5	94.7	94.5	5.5	2.7790		X	X	X	X				X	
5	94.7	94.5	5.7	2.7814		X	X	X	X		X			
6	94.8	94.5	5.2	2.7653		X	X	X	X		X		X	
6	94.7	94.5	6.0	2.7741		X	X	X	X		X			X
7	94.8	94.5	6.0	2.7627		X	X	X	X		X		X	X
7	94.8	94.5	6.9	2.7732	X	X	X	X	X		X		X	
8	94.8	94.5	7.5	2.7685		X	X	X	X		X	X	X	X
8	94.8	94.5	7.7	2.7702	X	X	X	X	X		X		X	X
9	94.9	94.5	9.3	2.7766	X	X	X	X	X		X	X	X	X
9	94.8	94.5	9.4	2.7781	X	X	X	X	X	X	X		X	X
10	94.9	94.5	11.0	2.7848	X	X	X	X	X	X	X	X	X	X

Figure 5.1. Best subset analysis describing changes of variables with population parameters. The first column contains the number of the predictors used, while the X mark in the last column identifies which ones they were. R-Sq is the R^2 coefficient of determination while R-Sq (adj) is the R^2 adjusted for degrees of freedom. C-p is a measure of error for the fit of residuals while s is the standard deviation about the regression line.

good fit for lifespan, although the possibility that information itself can be a contributing factor can not be discounted.

An interesting aspect of our analysis was the reverse association of lifespan expectancy at birth and fertility measurements, both as birth rate and as total fertility rate. The birth rate describes the average annual number of births per 1,000 population and is dependant on the level of fertility and the age structure of the population, while the total fertility rate is a more direct measure of fertility and it is defined as "the average number of children that would be born per woman if all women lived to the end of their childbearing years and bore children according to a given fertility rate at each age" (http://www.cia.gov/cia/publications/factbook/docs/notesanddefs.html). As expected both measures are very closely linked with one another (adjusted R-sp 96.7%) but birth rate seems to be more closely correlated with lifespan than total fertility (Figure 5.2), with birth and total fertility rates explaining 62.8% and 58.9% of life expectancy variance respectively. Due to the complicated nature of the trade-off between reproduction and longevity, especially in humans, it is not clear if the reverse association observed is due to physiological factors or behavioural patterns, or a combination of the two. A similar procedure used for expected lifespan (removing life expectancy and total fertility from the variables) in order to identify the best subset of parameters describing birth rate, showed that death rate, infant mortality and unemployment accounted for 74.5% percent of the variance of the birth rates with infant mortality being the most important predictor.

The analysis of the global data has provided us with a number of interesting observations mainly concerning the effect of demographic parameters on longevity and the importance of the individual's purchase power. Moreover, we were able to

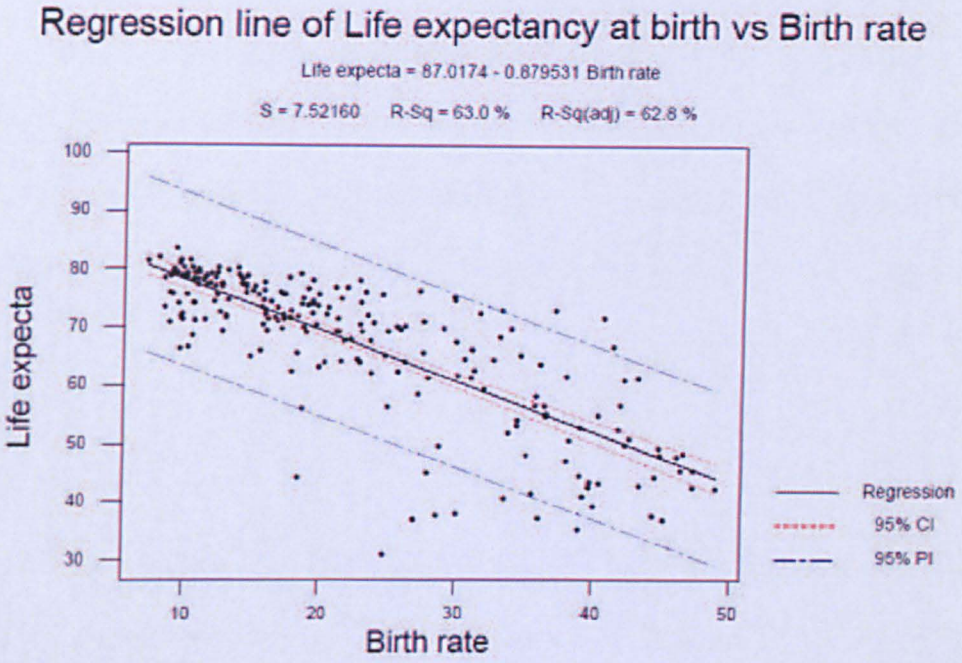


Figure 5.2. Graph representing the regression line between Life expectancy at birth and birth rate based on global data. Also seen are the 95% confidence intervals (CI) as red and the 95% prediction band (PI) as blue lines.

identify core components of the disposable soma theory concerning the link between life expectancy and measures of fertility. Unfortunately, differences in culture, policies and resources between each country limit the applicability of our results. That does not mean though, that strong associations are not identifiable rather that their quantitative values can be miscalculated and that weaker association can be lost.

5.11 Epilogue

From our everyday experience we see that there are marked differences between people belonging to different social groups no matter which criteria we use to distinguish those groups. We saw that education, occupation and income are predictors of lifespan and health in old age. A number of different factors responsible for these associations were presented. People with lower education or income are less worried about the future health implications of their diet and lifestyle, having a higher incidence of obesity, cardiovascular problems and diabetes. Psychological stress can also be considered as an important determinant of longevity and mortality. Inability to control one's life and incomplete social participation are two factors that elevate psychological stress in members of many social classes. Even mild environmental stresses, and especially temperature related stress, can have a profound effect on mortality. When the ability to avoid exposure is linked with income, social inequalities in excess deaths are observed. Early development and maternal effects can result in predispositions for a number of diseases when adult. A balanced diet and a healthy lifestyle of the mother have significant effects on her offspring. Access to health services also has a social gradient. Disadvantaged groups lose out on many services provided, despite being more often in contact with non-specialist health care

professionals. We concluded that frequent changes in social structure, the high variability in human populations, the ability of females to move through social classes and the plasticity of the phenotype, make evolutionary changes on a social basis unlikely to occur. Finally, we discussed the impact that social structure will have in the models we developed and their abilities to predict changes between socioeconomic groups and showed evidence for the association of life expectancy and GDP per capita on a global scale.

Utopias always describe the perfect society where all live happilly, in harmony and equity. Unfortunately reality is much different, inequality in life and health are easy to find in developed and developing countries alike. We may have improved the conditions for all classes in the last 100 years, today's lower socioeconomic groups living much better and longer than the richest in the beginning of the 20th century. But the improvement is not the same for everyone. Since 1976 the highest social class show a 5.7 years extension in lifespan, compared with just 1.7 for the more disadvantaged (Marmot, 2003). We started by making clear that we will restrict the discussion to biology, but while biology can observe and find associations between parameters, the solutions should be looked for in sociology, philosophy and most importantly policy planning. Maybe there is no perfect health system that will alleviate inequality, since its roots go much further than mere treatment. Maybe the only way to abolish inequality is to abolish disease (Jessop, 2000). Maybe we should try to change society instead of changing the individual (Marmot, 2003). No utopias exist or will ever do. Harmony and equity are goals that we all feel obliged to achieve but we constantly fail. No two people are equal in terms of abilities and needs and there is no equity where inequality is the building block. That is not to say that we should feel complacent about the inequalities we see around us. On the contrary, we

should strive to eliminate them and to provide everyone with the chance to fulfil their needs. We should all have the same chances in our lives and the ability to pursue our dreams.

6 Conclusions

This project aimed to elucidate the gene-gene and gene-environment interactions relevant to human ageing, by using epidemiological studies to construct a framework model that could both account for existing data and also be used to help identify the critical genes involved. It also aimed to provide a theoretical account for the intrinsic variability of the ageing process, especially between different socioeconomic groups within a single population. To address these aims we used the published literature and the established evolutionary theories of ageing to develop mathematical and computerized statistical models. The use of theoretical tools provided us with the ability to infer more general principles for the genetics of ageing than is possible when dealing experimentally with the effect of single (or a small number of) genes. Furthermore, recent advances in computing power made possible the construction of models containing complex relationships between a large number of environmental and genetic parameters, thus permitting programs with detailed and realistic interactions between the relevant factors.

We started by discussing the prevailing views for the evolution of ageing and the experimental evidence supporting them. We have seen that evolutionary theory considers ageing as a non-adaptive characteristic of life history which has evolved either through a mutation accumulation process based on a mutation-selection balance or through the action of trade offs and pleiotropic genes, or most likely through a combination of both. Furthermore we reviewed what is currently known for the genetic mechanisms underlying the ageing process in a variety of model organisms and in humans. The importance of maintenance and repair of DNA, and generally the ability to withstand physiological stress constantly emerge as central mechanisms during ageing both in model organism and human progeria mutations. A promising

genetic pathway extending lifespan is the IGF pathway which seems to be evolutionary conserved from yeast to mammals. Its action signifies the ability of lifespan to respond rapidly to changing environmental conditions through genes which are central during early development and gene-environment interactions during adulthood.

The modelling aspect of the work started with simple computer programs, e.g. with the aim of representing the disposable soma theory of ageing in the advanced Mathematica programming language. Analysis of the model parameters elucidated the detailed predictions of the model as well as some weaknesses of its simple mathematical representation in the case of humans. The next step was to use the disposable soma theory to construct a new model of trade-offs. The disposable soma theory, as we saw, considers the organism as a 'good economist' that seeks both to maintain its body (soma) and produce offspring with its limited resources in a way that will maximise its evolutionary fitness (Kirkwood and Rose, 1991). The expected survival of the organism in the wild will be the factor that determines the optimal energy allocation between maintenance and reproduction. Too high an investment in survival would be wasted when the organism dies, while too low an investment would result in premature death. Although the theory was first proposed in the context of physiological ecology, it can also account for the existence of genetic trade-offs between longevity and reproduction. In our model we changed the original equations used for the disposable soma theory model so that we could include further trade-offs, specifically the immunogenetic trade-off suggested by Westendorp et al (2001). In their paper Westendorp et al (2001) provided evidence that women with an innate T-helper type-1 (Th1) phenotype have a greater probability of spontaneous abortion than women with an innate T-helper type-2 (Th2) phenotype, while individuals with the

Th1 phenotype show lower susceptibility to fatal infection, indicative of a greater likelihood of survival in an environment where infectious diseases exert a significant mortality pressure. In the model we included parameters that could account for the reaction of the immune system towards the foetus and towards infections. Using population data from the World Health Organisation we were able to calculate the optimal level of the trade-off in two populations, Denmark and The Gambia, with very different infectious loads. Moreover, the model could be used to predict the rate of evolutionary change in the frequencies of genes specifying the immunogenetic trade-off in a way that could address whether or not the relatively recent reductions in exposure to infectious disease in developed countries could be responsible for significant, and perhaps ongoing, genetic change.

In our model we established for the first time the evolutionary plausibility of such a trade-off in humans. Furthermore, the model shows how within the context of the disposable soma theory, and given the ecological factors affecting mankind, this immunogenetic trade-off may have been critical for determining human longevity until relatively recently in the now-developed populations, and may still be crucial in developing countries. Finally, we illustrated in clear, quantitative terms how the advantage of increased resistance to fatal infectious disease could be offset by the increased risk of spontaneous abortion to produce an optimal range of gene frequencies where these effects are balanced according to the prevailing environmental parameters.

Our predictions for the immunogenetic trade-off are in accordance with recently acquired results by Westendorp (unpublished data) collected from an African (Ghanaian) population with high infection load. Moreover our assumptions about the genetic structure responsible for the trade-off and our predicted speed of frequency

change for the relevant genes seem to correspond very well with what is seen in epidemiological studies (Westendorp, personal communication).

Considering the estimates that only 30% of all human pregnancies reach successful completion while 50% of those lost are missed before the recognition of late menses (Choudhury and Knapp, 2001) we can easily understand the importance of work associated with a possible immunogenetic trade-off between reproduction and survival. Moreover, since the great majority of reproductive failure is due to unknown factors in a system where the immune status is of paramount importance, it is necessary to focus our attention to the action of the immune system of the mother towards the fetus. Our model provides theoretical support for an under-recognised potential source of reproductive problems and may explain their prevalence in a Western population, based on our recent history and evolution.

The second model followed a different path and led to a different theoretical framework. Our objective was to examine the evolutionary pressures on the *apo E* gene and to identify whether or not it is still the subject of selection. Apolipoprotein E (APO E) is a member of a diverse family of proteins collectively known as apolipoproteins, carrier proteins specializing in lipoprotein particle formation, secretion, transport, binding and metabolism (van Bockxmeer, 1994). Carriers of different *apo E* alleles show a difference in the incidence of cardiovascular disease, peripheral atherosclerosis and Alzheimer disease, possibly stroke and even in the ability to recover from trauma (Smith, 2002). One of the fundamental ideas in the evolutionary theories of ageing is that the force of natural selection decreases with increasing age. Under the mutation accumulation theory of ageing it is expected that deleterious mutations expressed after the end of reproduction will be almost selectively neutral, presenting an interesting puzzle for the evolution of genes with

late action, such as *apo E* in humans. Recently, the "grandmother hypothesis" has sought to provide support to the notion that late-acting genes are not selectively neutral by proposing an idea based on the food sharing between close kin in social species (Hawkes et al., 1998). A possible insight into *apo E* evolution has also been provided by its association with Alzheimer disease and oestrogen (Finch and Sapolsky, 1999; Sapolsky and Finch, 2000). We investigated these ideas using computer simulated populations with a number of lifestyle parameters corresponding to a contemporary European population. We treated the problem as a gene-environment interaction, where the gene by itself does not produce any differences in mortality but in association with the environment it can produce different hazards for the individual. The lifestyle parameters used were well known risk factors for cardiovascular disease, such as diet, alcohol consumption, smoking and physical activity. Using a number of randomly generated probability functions we were able to follow the changes of the *apo E* alleles' frequencies under different situations as well as to compute the expected mean life span under each parameter set.

Our results showed that *apo E* is still subject to some selection in a typical Western European population even when only a fraction of the population experiences *apo E* related mortality. We were able to examine quantitatively how this selection varied with different lifestyles. We were also able to elucidate some new aspects of the mutation accumulation theory and how it might act upon the distributions of gene effects within populations during the process of selection for longevity. At a more technical level, we provided an example of how an extreme-value distribution can describe an altered form of the Gompertz model to account for the mortality patterns seen within the population. Finally, we sought to answer how

the *apo E* gene might have evolved and we proposed a two-stage process related to meat-eating and the spread of agriculture.

Cardiovascular diseases are the main causes of mortality, at least in the Western world. Our work provides a framework through which epidemiological studies can address this complex problem in a more methodological way. More precisely our simulation is an example of the use of epidemiological data to account for the gene-environment interaction and its change with each generation. Furthermore we can use the model to predict how lifestyle changes are likely to affect longevity, and how selection might work to extend the lifespan of organisms. In our work we also show the important questions that we should ask when dealing with complex diseases and the factors that need to be quantified in order to obtain reliable results.

Finally, we addressed the problem of social stratification in relation to heterogeneity of morbidity and mortality during ageing. Education, occupational status and income were considered in relation to lifespan and health. The effects of low and high temperature on seasonal mortality were also discussed. Psychological stress due to status and social participation was linked with longevity and social position. The importance of early development and maternal effects for the incidence of common diseases in adulthood was highlighted considering a number of different studies. We discussed the idea of unequal health service access and utilization between different social groups and their effects in different parts of the world. Moreover we considered how our two models can account for the heterogeneity and variability of the ageing process in individuals from different social groups within the same population based on the differences in environmental challenges faced, and whether these differences are enough to unevenly shape evolution and gene

frequencies between social groups found within a single population. We concluded that the plasticity of early life and the adaptive nature of maternal effects, plus the common changes in human history and the ability of women to cross social boundaries in marriage, will make such differences between classes improbable. Our discussion made evident the importance of a balanced health service and the necessity for everyone to expect a healthy and long life irrespective of social class.

With this work we have shown how theoretical modeling of genetic and non-genetic factors in human ageing can contribute a strong predictive framework for further studies on this important topic. We will be able to use insights from the work in a major population-based study of the factors influencing maintenance of health among the oldest age groups (the Newcastle 85+ study). We will also contribute insights into a new European Union Framework 6 integrated project on 'Genetics of healthy ageing' that will commence later this year.

ὅστις τοῦ πλέονος μέρους χρῄζει τοῦ μετρίου παρεὶς ζώειν,

σκαιοσύναν φυλάσσων ἐν ἐμοὶ κατάδηλος ἔσται.

(Sophocles, The Oedipus at Colonus, 1211)

References

Ackermann, M., Bijlsma, R., James, A. C., Partridge, L., Zwaan, B. J. and Stearns, S. C. (2001). "Effects of assay conditions in life history experiments with Drosophila melanogaster." Journal of Evolutionary Biology 14(2): 199-209.

Akaike, H. (1973). Information theory as an extension of the maximum likehood principle.In: *Second International Symposium of Information Theory*. B. Petrov and F. Csaki. Akademiai Kiado. Budapest: 267-281.

Amaranayake, N., Church, J., Hill, C., Jackson, J., Jackson, V., Myers, C., Saeed, Z., Shipsey, C. and Symmonds, T. (2000). Social Trends 30, National Statistics.

Avlund, K., Holstein, B. E., Osler, M., Damsgaard, M. T., Holm-Pedersen, P. and Rasmussen, N. K. (2003). "Social position and health in old age: the relevance of different indicators of social position." Scandinavian Journal of Public Health 31(2): 126-136.

Aylin, P., Morris, S., Wakefield, J., Grossinho, A., Jarup, L. and Elliott, P. (2001). "Temperature, housing, deprivation and their relationship to excess winter mortality in Great Britain, 1986-1996." International Journal of Epidemiology 30(5): 1100-1108.

Bali, T. G. (2003). "The generalized extreme value distribution." Economics Letters 79(3): 423-427.

Barker, D. J. P., Gluckman, P. D., Godfrey, K. M., Harding, J. E., Owens, J. A. and Robinson, J. S. (1993). "Fetal Nutrition and Cardiovascular-Disease in Adult Life." Lancet 341(8850): 938-941.

Barnes, A. I. and Partridge, L. (2003). "Costing reproduction." Animal Behaviour 66: 199-204.

Barnes, J. (1984). The Complete Works of Aristotle. The Revised Oxford Translation. Princeton University Press, Princeton.

Beaglehole, R. and Yach, D. (2003). "Globalisation and the prevention and control of non- communicable disease: the neglected chronic diseases of adults." Lancet 362(9387): 903-908.

Behar, D. M., Garrigan, D., Kaplan, M. E., Mobasher, Z., Rosengarten, D., Karafet, T. M., Quintana-Murci, L., Ostrer, H., Skorecki, K. and Hammer, M. F. (2004). "Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations." Human Genetics 114(4): 354-365.

Bereczkei, T. and Dunbar, R. I. M. (1997). "Female-biased reproductive strategies in a Hungarian Gypsy population." Proceedings of the Royal Society of London Series B-Biological Sciences 264(1378): 17-22.

Bernstein, M. S., Costanza, M. C., James, R. W., Morris, M. A., Cambien, F., Raoux, S. and Morabia, A. (2002). "Physical activity may modulate effects of ApoE genotype on lipid profile." Arteriosclerosis Thrombosis and Vascular Biology 22(1): 133-140.

Bitterman, K. J., Medvedik, O. and Sinclair, D. A. (2003). "Longevity regulation in Saccharomyces cerevisiae: Linking metabolism, genome stability, and heterochromatin." Microbiology and Molecular Biology Reviews 67(3): 376-399.

Blair, S. N., Kohl, H. W., Barlow, C. E., Paffenbarger, R. S., Gibbons, L. W. and Macera, C. A. (1995). "Changes in Physical-Fitness and All-Cause Mortality - a Prospective-Study of Healthy and Unhealthy Men." Jama-Journal of the American Medical Association 273(14): 1093-1098.

Blanche, H., Cabanne, L., Sahbatou, M. and Thomas, G. (2001). "A study of French centenarians: are ACE and APOE associated with longevity?" Comptes

Rendus De L Academie Des Sciences Serie III-Sciences De La Vie-Life Sciences 324(2): 129-135.

Bloom, G. (2001). "Equity in health in unequal societies: meeting health needs in contexts of social change." Health Policy 57(3): 205-224.

Bluher, M., Kahn, B. B. and Kahn, C. R. (2003). "Extended Longevity in Mice Lacking the Insulin Receptor in Adipose Tissue." Science 299(5606): 572-574.

Bohnet, K., RegisBailly, A., VincentViry, M., Schlenck, A., Gueguen, R., Siest, G. and Visvikis, S. (1996). "Apolipoprotein E genotype epsilon 4/epsilon 2 in the STANISLAS cohort study - Dominance of the epsilon 2 allele?" Annals of Human Genetics 60: 509-516.

Bohr, V. A., Brosh, R. M., von Kobbe, C., Opresko, P. and Karmakar, P. (2002). "Pathways defective in the human premature aging disease Werner syndrome." Biogerontology 3(1-2): 89-94.

Bopp, M. and Minder, C. E. (2003). "Mortality by education in German speaking Switzerland, 1990- 1997: results from the Swiss National Cohort." International Journal of Epidemiology 32(3): 346-354.

Borrell, C., Rue, M., Pasarin, M. I., Rohlfs, I., Ferrando, J. and Fernandez, E. (2000). "Trends in Social Class Inequalities in Health Status, Health-Related Behaviors, and Health Services Utilization in a Southern European Urban Area (1983-1994)." Preventive Medicine 31(6): 691-701.

Borrell, C., Fernandez, E., Schiaffino, A., Benach, J., Rajmil, L., Villalbi, J. R. and Segura, A. (2001). "Social class inequalities in the use of and access to health services in Catalonia, Spain: what is the influence of supplemental private health insurance?" International Journal for Quality in Health Care 13(2): 117-125.

Borrell, C., Cortes, I., Artazcoz, L., Molinero, E. and Moncada, S. (2003). "Social inequalities in mortality in a retrospective cohort of civil servants in Barcelona." International Journal of Epidemiology 32(3): 386-389.

Bortolotti, A., Parazzini, F., Colli, E. and Landoni, M. (1997). "The epidemiology of erectile dysfunction and its risk factors." International Journal of Andrology 20(6): 323-334.

Bossuyt, N., Gadeyne, S., Deboosere, P. and Van Oyen, H. (2004). "Socio-economic inequalities in health expectancy in Belgium." Public Health 118(1): 3-10.

Bovet, P. and Paccaud, F. (2001). "Commentary: Alcohol, coronary heart disease and public health: which evidence-based policy?" International Journal of Epidemiology 30(4): 734-737.

Boyd Eaton, S., Eaton, S. B. and Cordain, L. (2002). Evolution, Diet, and Health.In: *Human Diet, Its Origin and Evolution.* P. S. Ungar and M. F. Teaford. Bergin & Garvey.

Breitner, J. C. S. and Miech, R. A. (1999). "Sex and sociobiology: a role for APOE?" Neurobiology of Aging 20(4): 445-445.

Bronnum-Hansen, H. (2000). "Socioeconomic differences in health expectancy in Denmark." Scandinavian Journal of Public Health 28(3): 194-199.

Brown-Borg, H. M., Borg, K. E., Meliska, C. J. and Bartke, A. (1996). "Dwarf mice and the ageing process." Nature 384(6604): 33.

Bryant, M. J. and Reznick, D. (2004). "Comparative studies of senescence in natural populations of guppies." American Naturalist 163(1): 55-68.

Burt, J., Hooper, R. and Jessopp, L. (2003). "The relationship between use of NHS Direct and deprivation in southeast London: an ecological analysis." Journal of Public Health Medicine 25(2): 174-176.

Cavalli-Sforza, L. L., Menozzi, P. and Piazza, A. (1993). "Demic Expansions and Human-Evolution." Science 259(5095): 639-646.

Chakraborty, R. (2003). "Genetics of castes and tribes of India." American Journal of Physical Anthropology 120(2): 206-208.

Chapuisat, M. and Keller, L. (2002). "Division of labour influences the rate of ageing in weaver ant workers." Proceedings of the Royal Society of London Series B-Biological Sciences 269(1494): 909-913.

Charlesworth, B. (1994). Evolution in Age-structured Populations. Cambridge University Press, Cambridge.

Charlesworth, B. and Hughes, K. A. (1996). "Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence." Proceedings of the National Academy of Sciences of the United States of America 93(12): 6140-6145.

Charlesworth, B. (2000). "Fisher, Medawar, Hamilton and the evolution of aging." Genetics 156(3): 927-931.

Chen, L., Lee, L., Kudlow, B. A., Dos Santos, H. G., Sletvold, O., Shafeghati, Y., Botha, E. G., Garg, A., Hanson, N. B. and Martin, G. M. (2003). "LMNA mutations in atypical Werner's syndrome." The Lancet 362(9382): 440-445.

Choudhury, S. R. and Knapp, L. A. (2001). "Human reproductive failure I: Immunological factors." Human Reproduction Update 7(2): 113-134.

Clancy, D. J., Gems, D., Harshman, L. G., Oldham, S., Stocker, H., Hafen, E., Leevers, S. J. and Partridge, L. (2001). "Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein." Science 292(5514): 104-106.

Corbo, R. M. and Scacchi, R. (1999). "Apolipoprotein E (APOE) allele distribution in the world. Is APOE 4 a 'thrifty' allele?" Annals of Human Genetics 63: 301-310.

Cordain, L. (2002). "Fatty acid composition and energy density of foods available to hominids: Implications for encephalization." American Journal of Physical Anthropology 117(S34): 57-57.

Corder, E. H., Lannfelt, L., Viitanen, M., Corder, L. S., Manton, K. G., Winblad, B. and Basun, H. (1996). "Apolipoprotein E genotype determines survival in the

oldest old (85 years or older) who have good cognition." Archives of Neurology 53(5): 418-422.

Corella, D., Tucker, K., Lahoz, C., Coltell, O., Cupples, L. A., Wilson, P. E. F., Schaefer, E. J. and Ordovas, J. M. (2001). "Alcohol drinking determines the effect of the APOE locus on LDL-cholesterol concentrations in men: the Framingham Offspring Study." American Journal of Clinical Nutrition 73(4): 736-745.

Coschigano, K. T., Clemmons, D., Bellush, L. L. and Kopchick, J. J. (2000). "Assessment of Growth Parameters and Life Span of GHR/BP Gene-Disrupted Mice." Endocrinology 141(7): 2608-2613.

Coschigano, K. T., Holland, A. N., Riders, M. E., List, E. O., Flyvbjerg, A. and Kopchick, J. J. (2003). "Deletion, But Not Antagonism, of the Mouse Growth Hormone Receptor Results in Severely Decreased Body Weights, Insulin, and Insulin-Like Growth Factor I Levels and Increased Life Span." Endocrinology 144(9): 3799-3810.

Couillard, C., Despres, J. P., Lamarche, B., Bergeron, J., Gagnon, J., Leon, A. S., Rao, D. C., Skinner, J. S., Wilmore, J. H. and Bouchard, C. (2001). "Effects of

endurance exercise training on plasma HDL cholesterol levels depend on levels of triglycerides - Evidence from men of the health, risk factors, exercise training and genetics (HERITAGE) family study." Arteriosclerosis Thrombosis and Vascular Biology 21(7): 1226-1232.

Cournil, A. and Kirkwood, T. B. L. (2001). "If you would live long, choose your parents well." Trends in Genetics 17(5): 233-235.

Curriero, F. C., Heiner, K. S., Samet, J. M., Zeger, S. L., Strug, L. and Patz, J. A. (2002). "Temperature and mortality in 11 cities of the eastern United States." American Journal of Epidemiology 155(1): 80-87.

De Benedictis, G., Tan, Q. H., Jeune, B., Christensen, K., Ukraintseva, S. V., Bonafe, M., Franceschi, C., Vaupel, J. W. and Yashin, A. I. (2001). "Recent advances in human gene-longevity association studies." Mechanisms of Ageing and Development 122(9): 909-920.

de Boer, J., Andressoo, J. O., de Wit, J., Huijmans, J., Beems, R. B., van Steeg, H., Weeda, G., van der Horst, G. T. J., van Leeuwen, W., Themmen, A. P. N., Meradji, M. and Hoeijmakers, J. H. J. (2002). "Premature aging in mice deficient in DNA repair and transcription." Science 296(5571): 1276-1279.

de Carli, C. M. D., Miller, B. L. M. D., Swan, G. E. P., Reed, T. P., Wolf, P. A. M. D. and Carmelli, D. P. (2001). "Cerebrovascular and Brain Morphologic Correlates of Mild Cognitive Impairment in the National Heart, Lung, and Blood Institute Twin Study." Archives of Neurology April 58(4): 643-647.

Department of Health and Inter-Departmental Working Group (1995). Sensible Drinking, Department Of Health.

Dillin, A., Crawford, D. K. and Kenyon, C. (2002). "Timing requirements for insulin/IGF-1 signaling in C-elegans." Science 298(5594): 830-834.

Doblhammer, G. (2000). "Reproductive history and mortality later in life: A comparative study of England and Wales and Austria." Population Studies 54(2): 169-176.

Doblhammer, G. and Vaupel, J. W. (2001). "Lifespan depends on month of birth." Proceedings of the National Academy of Sciences of the United States of America 98(5): 2934-2939.

Doblhammer, G. and Oeppen, J. (2003). "Reproduction and longevity among the British peerage: the effect of frailty and health selection." Proceedings of the Royal Society of London Series B-Biological Sciences 270(1524): 1541-1547.

Dong, L. M., Wilson, C., Wardell, M. R., Simmons, T., Mahley, R. W., Weisgraber, K. H. and Agard, D. A. (1994). "Human Apolipoprotein-E - Role of Arginine-61 in Mediating the Lipoprotein Preferences of the E3-Isoform and E4-Isoform." Journal of Biological Chemistry 269(35): 22358-22365.

Eliasson, B., Mero, N., Taskinen, M. R. and Smith, U. (1997). "The insulin resistance syndrome and postprandial lipid intolerance in smokers." Atherosclerosis 129(1): 79-88.

Engels, F. (1845). The Conditions Of The Working Class In England, Leipzig.

Engstrom, G., Liljedahl, L. E., Rasmuson, M. and Bjorklund, T. (1989). "Expression of Genetic and Environmental Variation During Aging .1. Estimation of Variance-Components for Number of Adult Offspring in Drosophila-Melanogaster." Theoretical and Applied Genetics 77(1): 119-122.

Eriksson, J. G., Forsen, T., Tuomilehto, J., Osmond, C. and Barker, D. J. P. (2001). "Early growth and coronary heart disease in later life: longitudinal study." British Medical Journal 322(7292): 949-953.

Evans, J. D. and Wheeler, D. E. (1999). "Differential gene expression between developing queens and workers in the honey bee, Apis mellifera." Proceedings

of the National Academy of Sciences of the United States of America 96(10): 5575-5580.

Feldman, H. A., Johannes, C. B., Derby, C. A., Kleinman, I. C., Mohr, B. A., Araujo, A. B. and McKinlay, J. B. (2000). "Erectile dysfunction and coronary risk factors: Prospective results from the Massachusetts Male Aging Study." Preventive Medicine 30(4): 328-338.

Finch, C. E. (1990). Longevity, Senescence, and the Genome. The University of Chicago, Chicago.

Finch, C. E. and Tanzi, R. E. (1997). "Genetics of aging." Science 278(5337): 407-411.

Finch, C. E. and Sapolsky, R. M. (1999). "The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms." Neurobiology of Aging 20(4): 407-428.

Finch, C. E. and Stanford, C. B. (2004). "Meat-adaptive genes and the evolution of slower aging in humans." Quarterly Review of Biology 79(1): 3-50.

Fletcher, G. F. (1997). "How to implement physical activity in primary and secondary prevention - A statement for healthcare professionals from the task force on risk reduction, American Heart Association." Circulation 96(1): 355-357.

Flurkey, K., Papaconstantinou, J., Miller, R. A. and Harrison, D. E. (2001). "Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production." PNAS 98(12): 6736-6741.

Flurkey, K., Papaconstantinou, J. and Harrison, D. E. (2002). "The Snell dwarf mutation Pit1dw can increase life span in mice." Mechanisms of Ageing and Development 123(2-3): 121-130.

Friedman, D. B. and Johnson, T. E. (1988). "A Mutation in the Age-1 Gene in Caenorhabditis-Elegans Lengthens Life and Reduces Hermaphrodite Fertility." Genetics 118(1): 75-86.

Frisoni, G. B., Louhija, J., Geroldi, C. and Trabucchi, M. (2001). "Longevity and the epsilon 2 allele of apolipoprotein E: The Finnish Centenarians Study." Journals of Gerontology Series A - Biological Sciences and Medical Sciences 56(2): M75-M78.

Fullerton, S. M., Clark, A. G., Weiss, K. M., Nickerson, D. A., Taylor, S. L., Stengard, J. H., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E. and Sing, C. F. (2000). "Apolipoprotein E variation at the sequence haplotype level: Implications for the origin and maintenance of a major human polymorphism." American Journal of Human Genetics 67(4): 881-900.

Galinsky, D., Tysoe, C., Brayne, C. E., Easton, D. F., Huppert, F. A., Dening, T. R., Paykel, E. S. and Rubinsztein, D. C. (1997). "Analysis of the apo E apo C-I, angiotensin converting enzyme and methylenetetrahydrofolate reductase genes as candidates affecting human longevity." Atherosclerosis 129(2): 177-183.

Gavrilov, L. A. and Gavrilova, N. S. (1999). "Is there a reproductive cost for human longevity?" Journal of Anti-aging Medicine 2(2): 121-123.

Gavrilov, L. A. and Gavrilova, N. S. (2002). "Evolutionary theories of aging and longevity." ScientificWorldJournal 2: 339-56.

Geesaman, B. J., Benson, E., Brewster, S. J., Kunkel, L. M., Blanche, H., Thomas, G., Perls, T. T., Daly, M. J. and Puca, A. A. (2003). "Haplotype-based identification of a microsomal transfer protein marker associated with the

human lifespan." Proceedings of the National Academy of Sciences of the United States of America 100(24): 14115-14120.

Gems, D. and Riddle, D. L. (1996). "Longevity in Caenorhabditis elegans reduced by mating but not gamete production." Nature 379(6567): 723-725.

Gems, D., Sutton, A. J., Sundermeyer, M. L., Albert, P. S., King, K. V., Edgley, M. L., Larsen, P. L. and Riddle, D. L. (1998). "Two pleiotropic classes of daf-2 mutation affect larval arrest, adult behavior, reproduction and longevity in Caenorhabditis elegans." Genetics 150(1): 129-155.

Gems, D. and Partridge, L. (2001). "Insulin/IGF signalling and ageing: seeing the bigger picture." Current Opinion in Genetics & Development 11(3): 287-292.

Gerdes, L. U., Gerdes, C., Hansen, P. S., Klausen, I. C. and Faergeman, O. (1996a). "Are men carrying the apolipoprotein epsilon 4- or epsilon 2 allele less fertile than epsilon 3 epsilon 3 genotypes?" Human Genetics 98(2): 239-242.

Gerdes, L. U., Gerdes, C., Hansen, P. S., Klausen, I. C., Faergeman, O. and Dyerberg, J. (1996b). "The apolipoprotein E polymorphism in Greenland Inuit in its global perspective." Human Genetics 98(5): 546-550.

Gerdes, L. U., Jeune, B., Ranberg, K. A., Nybo, H. and Vaupel, J. W. (2000). "Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: Apolipoprotein E gene is a "frailty gene", not a "longevity gene"." Genetic Epidemiology 19(3): 202-210.

Gerhard, G. S. (2003). "Comparative aspects of zebrafish (Danio rerio) as a model for aging research." Experimental Gerontology 38(11-12): 1333-1341.

Ghalambor, C. K. and Martin, T. E. (2001). "Fecundity-survival trade-offs and parental risk-taking in birds." Science 292(5516): 494-497.

Goddard, M. and Smith, P. (2001). "Equity of access to health care services: Theory and evidence from the UK." Social Science & Medicine 53(9): 1149-1162.

Godfrey, K. M. (1998). "Maternal regulation of fetal development and health in adult life." European Journal of Obstetrics Gynecology and Reproductive Biology 78(2): 141-150.

Gompertz, B. (1825). "On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies." Philosophical Transactions of the Royal Society of London 115: 513-585.

Gonos, E. S. (2000). "Genetics of aging: lessons from centenarians." Experimental Gerontology 35(1): 15-21.

Gouveia, N., Hajat, S. and Armstrong, B. (2003). "Socioeconomic differentials in the temperature-mortality relationship in Sao Paulo, Brazil." International Journal of Epidemiology 32(3): 390-397.

Guarente, L. and Kenyon, C. (2000). "Genetic pathways that regulate ageing in model organisms." Nature 408(6809): 255-262.

Gulliford, M. (2003). "Commentary: Epidemiological transition and socioeconomic inequalities in blood pressure in Jamaica." International Journal of Epidemiology 32(3): 408-409.

Hagberg, J. M., Wilund, K. R. and Ferrell, R. E. (2000). "APO E gene and gene-environment effects on plasma lipoprotein- lipid levels." Physiological Genomics 4(2): 101-108.

Hamilton, W. D. (1966). "The moulding of senescence by natural selection." Journal of Theoretical Biology 12(1): 12-45.

Hanlon, C. S. and Rubinsztein, D. C. (1995). "Arginine Residues at Codon-112 and Codon-158 in the Apolipoprotein-E Gene Correspond to the Ancestral State in Humans." Atherosclerosis 112(1): 85-90.

Hardy, R. and Kuh, D. (2002). "Does early growth influence timing of the menopause? Evidence from a British birth cohort." Human Reproduction 17(9): 2474-2479.

Hasty, P. and Vijg, J. (2002). "Aging - Genomic priorities in aging." Science 296(5571): 1250-1251.

Hasty, P., Campisi, J., Hoeijmakers, J., van Steeg, H. and Vijg, J. (2003). "Aging and genome maintenance: Lessons from the mouse?" Science 299(5611): 1355-1359.

Hawkes, K., O'Connell, J. F., Jones, N. G. B., Alvarez, H. and Charnov, E. L. (1998). "Grandmothering, menopause, and the evolution of human life histories." Proceedings of the National Academy of Sciences of the United States of America 95(3): 1336-1339.

Hegele, R. A. (2003). "Drawing the line in progeria syndromes." Lancet 362(9382): 416-417.

Hekimi, S., Lakowski, B., Barnes, T. M. and Ewbank, J. J. (1998). "Molecular genetics of life span in C-elegans: how much does it teach us?" Trends in Genetics 14(1): 14-20.

Hekimi, S. and Guarente, L. (2003). "Genetics and the specificity of the aging process." Science 299(5611): 1351-1354.

Helfand, S. L. and Rogina, B. (2003). "Molecular genetics of aging in the fly: is this the end of the beginning?" Bioessays 25(2): 134-141.

Helle, S., Kaar, P. and Jokela, J. (2002). "Human longevity and early reproduction in pre-industrial Sami populations." Journal of Evolutionary Biology 15(5): 803-807.

Hitchcock, C. and Sober, E. (2004). "Prediction versus accommodation and the risk of overfitting." British Journal for the Philosophy of Science 55(1): 1-34.

Hockett, B. and Haws, J. (2003). "Nutritional ecology and diachronic trends in Paleolithic diet and health." Evolutionary Anthropology 12: 211-216.

Holmes, D. J. and Ottinger, M. A. (2003). "Birds as long-lived animal models for the study of aging." Experimental Gerontology 38(11-12): 1365-1375.

Holzenberger, M., Dupont, J., Ducos, B., Leneuve, P., Geloen, A., Even, P. C., Cervera, P. and Le Bouc, Y. (2003). "IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice." Nature 421(6919): 182-187.

Horsburgh, K., McCarron, M. O., White, F. and Nicoll, J. A. R. (2000). "The role of apolipoprotein E in Alzheimer's disease, acute brain injury and cerebrovascular disease: evidence of common mechanisms and utility of animal models." Neurobiology of Aging 21(2): 245-255.

Houle, D., Hughes, K. A., Hoffmaster, D. K., Ihara, J., Assimacopoulos, S., Canada, D. and Charlesworth, B. (1994). "The effects of spontaneous mutation on quantitative traits .1. Variances and covariances of life-history traits." Genetics 138(3): 773-785.

Hsin, H. and Kenyon, C. (1999). "Signals from the reproductive system regulate the lifespan of C-elegans." Nature 399(6734): 362-366.

Hughes, K. A. and Charlesworth, B. (1994). "A genetic-analysis of senescence in Drosophila." Nature 367(6458): 64-66.

Hughes, K. A. (1995). "The evolutionary genetics of male life-history characters in Drosophila-melanogaster." Evolution 49(3): 521-537.

Hughes, K. A., Alipaz, J. A., Drnevich, J. M. and Reynolds, R. M. (2002). "A test of evolutionary theories of aging." Proceedings of the National Academy of Sciences of the United States of America 99(22): 14286-14291.

Hughes, W. O. H., Sumner, S., Van Borm, S. and Boomsma, J. J. (2003). "Worker caste polymorphism has a genetic basis in Acromyrmex leaf-cutting ants." Proceedings of the National Academy of Sciences of the United States of America 100(16): 9394-9397.

Humphries, K. H. and Gill, S. (2003). "Risks and benefits of hormone replacement therapy: The evidence speaks." Canadian Medical Association Journal 168(8): 1001-1010.

Humphries, S. E., Talmud, P. J., Hawe, E., Bolla, M., Day, I. N. M. and Miller, G. J. (2001). "Apolipoprotein E4 and coronary heart disease in middle-aged men who smoke: a prospective study." Lancet 358(9276): 115-119.

Huppert, F. A., Johnson, T. and Nickson, J. (2000). "High prevalence of prospective memory impairment in the elderly and in early-stage dementia: Findings from a population-based study." Applied Cognitive Psychology 14: S63-S81.

Ivanova, R., Henon, N., Lepage, V., Charron, D., Vicaut, E. and Schachter, F. (1998). "HLA-DR alleles display sex-dependent effects on survival and discriminate between individual and familial longevity." Human Molecular Genetics 7(2): 187-194.

Jazwinski, S. M. (1996). "Longevity, genes, and aging." Science 273(5271): 54-59.

Jazwinski, S. M. (1998). "Genetics of longevity." Experimental Gerontology 33(7-8): 773-783.

Jazwinski, S. M. (1999a). "Molecular mechanisms of yeast longevity." Trends in Microbiology 7(6): 247-252.

Jazwinski, S. M. (1999b). "Longevity, genes, and aging: A view provided by a genetic model system." Experimental Gerontology 34(1): 1-6.

Jazwinski, S. M. (1999c). "The RAS genes: a homeostatic device in Saccharomyces cerevisiae longevity." Neurobiology of Aging 20(5): 471-478.

Jensen, J., Lendorf, A., Stimpel, H., Frost, J., Ibsen, H. and Rosenkilde, P. (1999). "The prevalence and etiology of impotence in 101 male hypertensive outpatients." American Journal of Hypertension 12(3): 271-275.

Jessop, E. G. (2000). "Abolishing disease: the best way to tackle inequality." Journal of Public Health Medicine 22(2): 127-128.

Jordan, B. D., Relkin, N. R., Ravdin, L. D., Jacobs, A. R., Bennett, A. and Gandy, S. (1997). "Apolipoprotein E epsilon 4 associated with chronic traumatic brain injury in boxing." Jama-Journal of the American Medical Association 278(2): 136-140.

Kaplan, H., Hill, K., Lancaster, J. and Hurtado, A. M. (2000). "A theory of human life history evolution: Diet, intelligence, and longevity." Evolutionary Anthropology 9(4): 156-185.

Kaplan, J. R., Manuck, S. B., Fontenot, M. B. and Mann, J. J. (2002). "Central nervous system monoamine correlates of social dominance in cynomolgus monkeys (Macaca fascicularis)." Neuropsychopharmacology 26(4): 431-443.

Kaprio, J., Kujala, U. M., Koskenvuo, M. and Sarna, S. (2000). "Physical activity and other risk factors in male twin-pairs discordant for coronary heart disease." Atherosclerosis 150(1): 193-200.

Kauma, H., Savolainen, M. J., Rantala, A. O., Lilja, M., Kervinen, K., Reunanen, A. and Kesaniemi, Y. A. (1998). "Apolipoprotein E phenotype determines the

effect of alcohol on blood pressure in middle-aged men." American Journal of Hypertension 11(11): 1334-1343.

Keatinge, W. R., Donaldson, G. C., Bucher, K., Jendritsky, G., Cordioli, E., Martinelli, M., Dardanoni, L., Katsouyanni, K., Kunst, A. E., Mackenbach, J. P., McDonald, C., Nayha, S. and Vuori, I. (1997). "Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe." Lancet 349(9062): 1341-1346.

Keatinge, W. R. (2003). "Commentary: Mortality from environmental factors, but which ones?" International Journal of Epidemiology 32(3): 398-399.

Keavney, B., Parish, S., Palmer, A., Clark, S., Youngman, L., Danesh, J., McKenzie, C., Delepine, M., Lathrop, M., Peto, R. and Collins, R. (2003). "Large-scale evidence that the cardiotoxicity of smoking is not significantly modified by the apolipoprotein E epsilon 2/epsilon 3/epsilon 4 genotype." Lancet 361(9355): 396-398.

Keller, L. and Genoud, M. (1997). "Extraordinary lifespans in ants: a test of evolutionary theories of ageing." Nature 389(6654): 958-960.

Kemnitz, J. W., Sapolsky, R. M., Altmann, J., Muruthi, P., Mott, G. E. and Stefanick, M. L. (2002). "Effects of food availability on serum insulin and lipid concentrations in free-ranging baboons." American Journal of Primatology 57(1): 13-19.

Kennedy, G. E. (2003). "Palaeolithic grandmothers? Life history theory and early Homo." Journal of the Royal Anthropological Institute 9(3): 549-572.

Keskimaki, I. (2003). "How did Finland's economic recession in the early 1990s affect socio-economic equity in the use of hospital care?" Social Science & Medicine 56(7): 1517-1530.

Kirkland, J. L. (1989). "Evolution and aging." Genome 31(1): 398-405.

Kirkwood, T. B. (1977). "Evolution of ageing." Nature 270(5635): 301-4.

Kirkwood, T. B. L. and Cremer, T. (1982). "Cytogerontology since 1881 - a reappraisal of August Weismann and a review of modern progress." Human Genetics 60(2): 101-121.

Kirkwood, T. B. L. (1985). Comparative and evolutionary aspects of longevity.In: *Handbook of the Biology of Aging*. C. E. Finch and E. L. Schneider. Van Nostrand Reinhold. New York: 27-44.

Kirkwood, T. B. L. and Holliday, R. (1986). Ageing as a consequence of natural selection.In: *The Biology of Human Ageing*. K. J. Collins and A. J. Bittles. Cambridge University Press. Cambridge: 1-16.

Kirkwood, T. B. L. and Rose, M. R. (1991). "Evolution of senescence - Late survival sacrificed for reproduction." Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 332(1262): 15-24.

Kirkwood, T. B. L. (1992). "Comparative life spans of species - Why do species have the life spans they do." American Journal of Clinical Nutrition 55(6): S1191-S1195.

Kirkwood, T. B. L. (1993). "The disposable soma theory - Evidence and implications." Netherlands Journal of Zoology 43(3-4): 359-363.

Kirkwood, T. B. L. (1997). "The origins of human ageing." Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 352(1363): 1765-1772.

Kirkwood, T. B. L. and Austad, S. N. (2000). "Why do we age?" Nature 408(6809): 233-238.

Kirkwood, T. B. L. (2002). "Evolution of ageing." Mechanisms of Ageing and Development 123(7): 737-745.

Klatsky, A. L. (2001). "Commentary: Could abstinence from alcohol be hazardous to your health?" International Journal of Epidemiology 30(4): 739-742.

Koenig, W. M. D., Sund, M. M. S., Doring, A. M. D. and Ernst, E. M. D. (1997). "Leisure-Time Physical Activity but Not Work-Related Physical Activity Is Associated With Decreased Plasma Viscosity: Results From a Large Population Sample." Circulation 95(2): 335 - 341.

Korpelainen, H. (2000). "Fitness, reproduction and longevity among European aristocratic and rural Finnish families in the 1700s and 1800s." Proceedings of the Royal Society of London Series B-Biological Sciences 267(1454): 1765-1770.

Kotz, S., and Nadarajah S (2000). Extreme Value Distributions: Theory and Applications. Imperial College Press, London.

Kris-Etherton, P. M., Hecker, K. D., Bonanome, A., Coval, S. M., Binkoski, A. E., Hilpert, K. F., Griel, A. E. and Etherton, T. D. (2002). "Bioactive compounds

in foods: Their role in the prevention of cardiovascular disease and cancer." American Journal of Medicine 113: 71-88.

Kuusisto, J., Mykkanen, L., Kervinen, K., Kesaniemi, Y. A. and Laakso, M. (1995). "Apolipoprotein E4 Phenotype Is Not an Important Risk Factor for Coronary Heart-Disease or Stroke in Elderly Subjects." Arteriosclerosis Thrombosis and Vascular Biology 15(9): 1280-1286.

Lahoz, C., Schaefer, E. J., Cupples, L. A., Wilson, P. W. F., Levy, D., Osgood, D., Parpos, S., Pedro-Botet, J., Daly, J. A. and Ordovas, J. M. (2001). "Apolipoprotein E genotype and cardiovascular disease in the Framingham Heart Study." Atherosclerosis 154(3): 529-537.

Lane, M. A. (2000). "Nonhuman primate models in biogerontology." Experimental Gerontology 35(5): 533-541.

Lawlor, D. A., Harvey, D. and Dews, H. G. (2000). "Investigation of the association between excess winter mortality and socio-economic deprivation." Journal of Public Health Medicine 22(2): 176-181.

Le Bourg, E., Thon, B., Legare, J., Desjardins, B. and Charbonneau, H. (1993). "Reproductive life of French-Canadians in the 17-18th centuries: a search for a trade-off between early fecundity and longevity." Exp Gerontol 28(3): 217-32.

L'Ecuyer, P. (1998). Random Number Generation.In: *Handbook on Simulation.* J. Banks. Wiley. 93 - 137.

Lee, J. H., Tang, M. X., Schupf, N., Stern, Y., Jacobs, D. M., Tycko, B. and Mayeux, R. (2001). "Mortality and apolipoprotein E in Hispanic, African-American, and Caucasian elders." American Journal of Medical Genetics 103(2): 121-127.

Lee, R. D. (2003). "Rethinking the evolutionary theory of aging: Transfers, not births, shape social species." Proceedings of the National Academy of Sciences of the United States of America 100(16): 9637-9642.

Leroi, A. M. (2001). "Molecular signals versus the Loi de Balancement." Trends in Ecology & Evolution 16(1): 24-29.

Lewontin, R. C. (1968). Biological models.In: *Dictionary of the History of Ideas.* P. P. Wiener. Scribner's. New York, 1: 243-247.

Liang, H. Y., Masoro, E. J., Nelson, J. F., Strong, R., McMahon, C. A. and Richardson, A. (2003). "Genetic mouse models of extended lifespan." Experimental Gerontology 38(11-12): 1353-1364.

Lin, Y. J., Seroude, L. and Benzer, S. (1998). "Extended life-span and stress resistance in the Drosophila mutant methuselah." Science 282(5390): 943-946.

Lindstrom, J. (1999). "Early development and fitness in birds and mammals." Trends in Ecology & Evolution 14(9): 343-348.

Lio, D., Pes, G. M., Carru, C., Listi, F., Ferlazzo, V., Candore, G., Colonna-Romano, G., Ferrucci, L., Deiana, L., Baggio, G., Franceschi, C. and Caruso, C. (2003). "Association between the HLA-DR alleles and longevity: a study in Sardinian population." Experimental Gerontology 38(3): 313-317.

Liu, S. M., Ma, J., Ridker, P. M., Breslow, J. L. and Stampfer, M. J. (2003). "A prospective study of the association between APOE genotype and the risk of myocardial infarction among apparently healthy men." Atherosclerosis 166(2): 323-329.

Ljungquist, B., Berg, S., Lanke, J., McClearn, G. E. and Pedersen, N. L. (1998). "The effect of genetic factors for longevity: A comparison of identical and fraternal

twins in the Swedish twin registry." Journals of Gerontology Series a-Biological Sciences and Medical Sciences 53(6): M441-M446.

Lloyd-Jones, D. M., Larson, M. G., Leip, E. P., Beiser, A., D'Agostino, R. B., Kannel, W. B., Murabito, J. M., Vasan, R. S., Benjamin, E. J. and Levy, D. (2002). "Lifetime risk for developing congestive heart failure - The Framingham Heart Study." Circulation 106(24): 3068-3072.

Lokkegaard, E., Pedersen, A. T., Heitmann, B. L., Jovanovic, Z., Keiding, N., Hundrup, Y. A., Obel, E. B. and Ottesen, B. (2003). "Relation between hormone replacement therapy and ischaemic heart disease in women: prospective observational study." British Medical Journal 326(7386): 426-428A.

Lopez, A. D., Ahmad, O. B., Guillot, M., Inoue, M., Ferguson, B. D., Salomon, J. A. (2001). Life tables for 191 countries for 2000: Data, methods, results (GPE discussion paper no. 40). Geneva, World Health Organization.

Lopez-Azpiazu, I., Sanchez-Villegas, A., Johansson, L., Petkeviciene, J., Prattala, R. and Martinez-Gonzalez, M. A. (2003). "Disparities in food habits in Europe:

systematic review of educational and occupational differences in the intake of fat." Journal of Human Nutrition and Dietetics 16(5): 349-364.

Lopuhaa, C. E., Roseboom, T. J., Osmond, C., Barker, D. J. P., Ravelli, A. C. J., Bleker, O. P., van der Zee, J. S. and van der Meulen, J. H. P. (2000). "Atopy, lung function, and obstructive airways disease after prenatal exposure to famine." Thorax 55(7): 555-561.

Lucas, A. (2000). "Programming not metabolic imprinting." American Journal of Clinical Nutrition 71(2): 602-602.

Luckinbill, L. S., Arking, R., Clare, M. J., Cirocco, W. C. and Buck, S. A. (1984). "Selection for delayed senescence in Drosophila-Melanogaster." Evolution 38(5): 996-1003.

Lumey, L. H. and Stein, A. D. (1997). "In utero exposure to famine and subsequent fertility: The Dutch Famine Birth Cohort Study." American Journal of Public Health 87(12): 1962-1966.

Lummaa, V., Lemmetyinen, R., Haukioja, E. and Pikkola, M. (1998). "Seasonality of births in Homo sapiens in pre-industrial Finland: Maximisation of offspring survivorship?" Journal of Evolutionary Biology 11(2): 147-157.

Lummaa, V. and Clutton-Brock, T. (2002). "Early development, survival and reproduction in humans." Trends in Ecology & Evolution 17(3): 141-147.

Lummaa, V. (2003). "Early developmental conditions and reproductive success in humans: Downstream effects of prenatal famine, birthweight, and timing of birth." American Journal of Human Biology 15(3): 370-379.

Lummaa, V. and Tremblay, M. (2003). "Month of birth predicted reproductive success and fitness in pre-modern Canadian women." Proceedings of the Royal Society of London Series B-Biological Sciences 270(1531): 2355-2361.

Lycett, J. E., Dunbar, R. I. M. and Voland, E. (2000). "Longevity and the costs of reproduction in a historical human population." Proceedings of the Royal Society of London Series B-Biological Sciences 267(1438): 31-35.

Mackenbach, J. P., Kunst, A. E., Cavelaars, A. E. J. M., Groenhof, F. and Geurts, J. J. M. (1997). "Socioeconomic inequalities in morbidity and mortality in western Europe." The Lancet 349(9066): 1655-1659.

Mackenbach, J. P., Cavelaars, A., Kunst, A. E. and Groenhof, F. (2000). "Socioeconomic inequalities in cardiovascular disease mortality - An international study." European Heart Journal 21(14): 1141-1151.

Mahley, R. W. and Rall, S. C. (1999). "Is epsilon 4 the ancestral human apoE allele?" Neurobiology of Aging 20(4): 429-430.

Marmot, M. and Bobak, M. (2000). "International comparators and poverty and health in Europe." British Medical Journal 321(7269): 1124-1128.

Marmot, M. G. (2001a). "Alcohol and coronary heart disease." International Journal of Epidemiology 30(4): 724-729.

Marmot, M. G. (2001b). "Commentary: Reflections on alcohol and coronary heart disease." International Journal of Epidemiology 30(4): 729-734.

Marmot, M. G. (2003). "Understanding social inequalities in health." Perspectives in Biology and Medicine 46(3): S9-S23.

Martikainen, P. T. and Marmot, M. G. (1999). "Socioeconomic differences in weight gain and determinants and consequences of coronary risk factors." American Journal of Clinical Nutrition 69(4): 719-726.

Martin, G. M. (1999). "APOE alleles and lipophylic pathogens." Neurobiology of Aging 20(4): 441-443.

Martin, G. M. and Oshima, J. (2000). "Lessons from human progeroid syndromes." Nature 408(6809): 263-266.

Martin, T. E. (2002). "A new view of avian life-history evolution tested on an incubation paradox." Proceedings of the Royal Society of London Series B-Biological Sciences 269(1488): 309-316.

Marz, W., Hoffmann, M. M., Scharnagl, H., Fisher, E., Chen, M., Nauck, M., Feussner, G. and Wieland, H. (1998). "Apolipoprotein E2 (Arg136 -> Cys) mutation in the receptor binding domain of apoE is not associated with dominant type III hyperlipoproteinemia." Journal of Lipid Research 39(3): 658-669.

May, R. M. (2004). "Uses and abuses of mathematics in biology." Science 303(5659): 790-793.

McDade, T. W., Beck, M. A., Kuzawa, C. W. and Adair, L. S. (2001). "Prenatal undernutrition and postnatal growth are associated with adolescent thymic function." Journal of Nutrition 131(4): 1225-1231.

McDavid, K., Tucker, T. C., Sloggett, A. and Coleman, M. P. (2003). "Cancer survival in Kentucky and health insurance coverage." Archives of Internal Medicine 163(18): 2135-2144.

McNutt, L. A., Wu, C. T., Xue, X. N. and Hafner, J. P. (2003). "Estimating the relative risk in cohort studies and clinical trials of common outcomes." American Journal of Epidemiology 157(10): 940-943.

Medawar, P. B. (1952). An Unsolved Problem in Biology. Lewis London.

Medawar, P. B. (1960). Immunological tolerance.In: *Nobel Lectures, including Presentation Speeches and Laureates' Biographies.* Elsevier Publishing Co. Amsterdam.

Mendez, M. A., Cooper, R., Wilks, R., Luke, A. and Forrester, T. (2003). "Income, education, and blood pressure in adults in Jamaica, a middle-income developing country." International Journal of Epidemiology 32(3): 400-408.

Mercer, J. B. (2003). "Cold - an underrated risk factor for health." Environmental Research 92(1): 8-13.

Merry, B. J. (2004). "Oxidative stress and mitochondrial function with aging - the effects of calorie restriction." Aging Cell 3(1): 7-12.

Migliaccio, E., Giorgio, M., Mele, S., Pelicci, G., Reboidl, P., Pandolfi, P. P., Lanfrancone, L. and Pelicci, P. G. (1999). "The p66(shc) adaptor protein

controls oxidative stress response and life span in mammals." Nature 402(6759): 309-313.

Millen, B. E., Quatromoni, P. A., Copenhafer, D. L., Demissie, S., O'Horo, C. E. and D'Agostino, R. B. (2001). "Validation of a dietary pattern approach for evaluating nutritional risk: The Framingham Nutrition Studies." Journal of the American Dietetic Association 101(2): 187-194.

Millen, B. E., Quatromoni, P. A., Nam, B. H., O'Horo, C. E., Polak, J. F. and D'Agostino, R. B. (2002). "Dietary patterns and the odds of carotid atherosclerosis in women: The framingham nutrition studies." Preventive Medicine 35(6): 540-547.

Mitchell, B. D., Hsueh, W. C., King, T. M., Pollin, T. I., Sorkin, J., Agarwala, R., Schaffer, A. A. and Shuldiner, A. R. (2001). "Heritability of life span in the old order Amish." American Journal of Medical Genetics 102(4): 346-352.

Montooth, K. L., Marden, J. H. and Clark, A. G. (2003). "Mapping determinants of variation in energy metabolism, respiration and flight in Drosophila." Genetics 165(2): 623-635.

Moore, S. E., Cole, T. J., Poskitt, E. M. E., Sonko, B. J., Whitehead, R. G., McGregor, I. A. and Prentice, A. M. (1997). "Season of birth predicts mortality in rural Gambia." Nature 388(6641): 434-434.

Moore, S. E., Cole, T. J., Collinson, A. C., Poskitt, E. M. E., McGregor, I. A. and Prentice, A. M. (1999). "Prenatal or early postnatal events predict infectious deaths in young adulthood in rural Africa." International Journal of Epidemiology 28(6): 1088-1095.

Moore, V. and Davies, M. (2002). "Nutrition before birth, programming and the perpetuation of social inequalities in health." Asia Pacific Journal of Clinical Nutrition 11: S529-S536.

Morris, J. Z., Tissenbaum, H. A. and Ruvkun, G. (1996). "A phosphatidylinositol-3-OH kinase family member regulating longevity and diapause in Caenorhabditis elegans." Nature 382(6591): 536-539.

Mousseau, T. A. and Fox, C. W. (1998). "The adaptive significance of maternal effects." Trends in Ecology & Evolution 13(10): 403-407.

Mulder, V. B. (2000). "Optimizing offspring: the quantity-quality tradeoff in agropastoral Kipsigis." Evolution and Human Behavior 21(6): 391-410.

Murphy, C. T., McCarroll, S. A., Bargmann, C. I., Fraser, A., Kamath, R. S., Ahringer, J., Li, H. and Kenyon, C. (2003). "Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans." Nature 424(6946): 277-284.

Nager, R. G., Monaghan, P. and Houston, D. C. (2001). "The cost of egg production: increased egg production reduces future fitness in gulls." Journal of Avian Biology 32(2): 159-166.

Nakata, Y., Katsuya, T., Rakugi, H., Takami, S., Ohishi, M., Kamino, K., Higaki, J., Tabuchi, Y., Kumahara, Y., Miki, T. and Ogihara, T. (1996). "Polymorphism of the apolipoprotein E and angiotensin-converting enzyme genes in Japanese subjects with silent myocardial ischemia." Hypertension 27(6): 1205-1209.

Nanchahal, K., Ashton, W. D. and Wood, D. A. (2000). "Alcohol consumption, metabolic cardiovascular risk factors and hypertension in women." International Journal of Epidemiology 29(1): 57-64.

Napoli, C., Martin-Padura, I., de Nigris, F., Giorgio, M., Mansueto, G., Somma, P., Condorelli, M., Sica, G., De Rosa, G. and Pelicci, P. (2003). "Deletion of the p66(Shc) longevity gene reduces systemic and tissue oxidative stress, vascular

cell apoptosis, and early atherogenesis in mice fed a high-fat diet." Proceedings of the National Academy of Sciences of the United States of America 100(4): 2112-2116.

Nietzsche, F. (1996). Human, All Too Human: A Book for Free Spirits. Cambridge University Press, Cambridge.

Nilsson, P. M., Engberg, M., Nilsson, J. A., Karlsmose, B. and Lauritzen, T. (2003a). "Adverse social factors predict early ageing in middle-aged men and women: the Ebeltoft Health Study, Denmark." Scandinavian Journal of Public Health 31(4): 255-260.

Nilsson, P. M., Nilsson, E., Svanberg, L. and Samsioe, G. (2003b). "Longevity after early surgical menopause-the long-term effect of a permanent cessation of reproductive function and female sex hormone loss." Eur J Obstet Gynecol Reprod Biol 110(1): 63-5.

Opresko, P. L., von Kobbe, C., Laine, J.-P., Harrigan, J., Hickson, I. D. and Bohr, V. A. (2002). "Telomere-binding Protein TRF2 Binds to and Stimulates the Werner and Bloom Syndrome Helicases." Journal of Biological Chemistry 277(43): 41110-41119.

Ordovas, J. M., LopezMiranda, J., Mata, P., PerezJimenez, F., Lichtenstein, A. H. and Schaefer, E. J. (1995). "Gene-diet interaction in determining plasma lipid response to dietary intervention." Atherosclerosis 118: S11-S27.

Ordovas, J. M. (1999). "The genetics of serum lipid responsiveness to dietary interventions." Proceedings of the Nutrition Society 58(1): 171-187.

Orr, W. C. and Sohal, R. S. (1994). "Extension of life-span by overexpression of superoxide- dismutase and catalase in Drosophila-melanogaster." Science 263(5150): 1128-1130.

Osmond, C., Barker, D. J. P., Eriksson, J. G. and Forsen, T. (2001). "Early growth and coronary heart disease in later life - Reply." British Medical Journal 323(7312): 573-573.

Page, R. E. and Peng, C. Y. S. (2001). "Aging and development in social insects with emphasis on the honey bee, Apis mellifera L." Experimental Gerontology 36(4-6): 695-711.

Panagiotakos, D. B., Pitsavos, C., Chrysohoou, C., Stefanadis, C. and Toutouzas, P. (2002). "Risk stratification of coronary heart disease in Greece: Final results

from the CARDIO2000 epidemiological study." Preventive Medicine 35(6): 548-556.

Papiha, S. S. (1996). "Genetic variation in India." Human Biology 68(5): 607-628.

Partridge, L. and Barton, N. H. (1993). "Optimality, mutation and the evolution of aging." Nature 362(6418): 305-311.

Partridge, L., Prowse, N. and Pignatelli, P. (1999). "Another set of responses and correlated responses to selection on age at reproduction in Drosophila melanogaster." Proceedings of the Royal Society of London Series B-Biological Sciences 266(1416): 255-261.

Pepine, C. J. (1997). "Potential role of angiotensin-converting enzyme inhibition in myocardial ischemia and current clinical trials." Clinical Cardiology 20(11): 58-64.

Petersen, S., Rayner M. (2002). Coronary Heart Statistics. Department of Public Health, London, UK.

Phillips, D. I. W. (1998). "Birth weight and the future development of diabetes - A review of the evidence." Diabetes Care 21: B150-B155.

Piano, M. R. (2002). "Alcoholic cardiomyopathy - Incidence, clinical characteristics, and pathophysiology." Chest 121(5): 1638-1650.

Piccinni, M. P. (2002). "T-cell cytokines in pregnancy." American Journal of Reproductive Immunology 47(5): 289-294.

Pletcher, S. D., Houle, D. and Curtsinger, J. W. (1998). "Age-specific properties of spontaneous mutations affecting mortality in Drosophila melanogaster." Genetics 148(1): 287-303.

Prattala, R., Karisto, A. and Berg, M. A. (1994). "Consistency and Variation in Unhealthy Behavior among Finnish Men, 1982-1990." Social Science & Medicine 39(1): 115-122.

Press, V., Freestone, I. and George, C. F. (2003). "Physical activity: the evidence of benefit in the prevention of coronary heart disease." Qjm-an International Journal of Medicine 96(4): 245-251.

Promislow, D. E. L., Tatar, M., Khazaeli, A. A. and Curtsinger, J. W. (1996). "Age-specific patterns of genetic variance in Drosophila melanogaster .1. Mortality." Genetics 143(2): 839-848.

Promislow, D. E. L. and Tatar, M. (1998). "Mutation and senescence: where genetics and demography meet." Genetica 103: 299-314.

Prowse, N. and Partridge, L. (1997). "The effects of reproduction on longevity and fertility in male Drosophila melanogaster." Journal of Insect Physiology 43(6): 501-512.

Puca, A. A., Daly, M. J., Brewster, S. J., Matise, T. C., Barrett, J., Shea-Drinkwater, M., Kang, S., Joyce, E., Nicoli, J., Benson, E., Kunkel, L. M. and Perls, T. (2001). "A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4." Proceedings of the National Academy of Sciences of the United States of America 98(18): 10505-10508.

Puddey, I. B., Rakic, V., Dimmitt, S. B. and Beilin, L. J. (1999a). "Influence of pattern of drinking on cardiovascular disease and cardiovascular risk factors - a review." Addiction 94(5): 649-663.

Puddey, I. B., Rakic, V., Dimmitt, S. B., Burke, V., Beilin, L. J. and Van Bockxmeer, F. (1999b). "Apolipoprotein E genotype and the blood pressure raising effect of alcohol." American Journal of Hypertension 12(9): 946-947.

Puddey, I. B., Zilkens, R. R., Croft, K. D. and Beilin, L. J. (2001). "Alcohol and endothelial function: A brief review." Clinical and Experimental Pharmacology and Physiology 28(12): 1020-1024.

Rall, S. C., Jr., Weisgraber, K. H. and Mahley, R. W. (1982). "Human apolipoprotein E. The complete amino acid sequence." Journal of Biological Chemistry. 257(8): 4171-8.

Ramaraj, P., Kessler, S. P., Colmenares, C. and Sen, G. C. (1998). "Selective restoration of male fertility in mice lacking angiotensin-converting enzymes by sperm-specific expression of the testicular isozyme." The Journal of Clinical Investigation 102(2): 371-378.

Rapin, I., Lindenbaum, Y., Dickson, D. W., Kraemer, K. H. and Robbins, J. H. (2000). "Cockayne syndrome and xeroderma pigmentosum: DNA repair disorders with overlaps and paradoxes." Neurology 55(10): 1442-1449.

Rea, I. M., Mc Dowell, I., McMaster, D., Smye, M., Stout, R. and Evans, A. (2001). "Apolipoprotein E alleles in nonagenarian subjects in the Belfast Elderly Longitudinal Free-living Ageing Study (BELFAST)." Mechanisms of Ageing and Development 122(13): 1367-1372.

Reznick, D., Buckwalter, G., Groff, J. and Elder, D. (2001). "The evolution of senescence in natural populations of guppies (Poecilia reticulata): a comparative approach." Experimental Gerontology 36(4-6): 791-812.

Richards, M. P. (2002). "A brief review of the archaeological evidence for Palaeolithic and Neolithic subsistence." European Journal of Clinical Nutrition 56(12): U12-U20.

Riddle, D. L. (1999). "Ageing - A message from the gonads." Nature 399(6734): 308-309.

Rimm, E. (2001). "Commentary: Alcohol and coronary heart disease - laying the foundation for future work." International Journal of Epidemiology 30(4): 738-739.

Rimm, E. B., Williams, P., Fosher, K., Criqui, M. and Stampfer, M. J. (1999). "Moderate alcohol intake and lower risk of coronary heart disease: meta-analysis of effects on lipids and haemostatic factors." British Medical Journal 319(7224): 1523-1528D.

Robbins, A. S., Chao, S. Y. and Fonseca, V. P. (2002). "What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes." Annals of Epidemiology 12(7): 452-454.

Rogina, B., Reenan, R. A., Nilsen, S. P. and Helfand, S. L. (2000). "Extended life-span conferred by cotransporter gene mutations in Drosophila." Science 290(5499): 2137-2140.

Rooney, C., McMichael, A. J., Kovats, R. S. and Coleman, M. P. (1998). "Excess mortality in England and Wales, and in Greater London, during the 1995 heatwave." Journal of Epidemiology and Community Health 52(8): 482-486.

Roos, G., Johansson, L., Kasmel, A., Klumbiene, J. and Prattala, R. (2001). "Disparities in vegetable and fruit consumption: European cases from the north to the south." Public Health Nutrition 4(1): 35-43.

Rose, M. R. and Charlesworth, B. (1981). "Genetics of life-history in Drosophila-Melanogaster .2. Exploratory selection experiments." Genetics 97(1): 187-196.

Rose, M. R. (1991). Evolutionary Biology of Aging. Oxford University Press.

Rose, M. R. (1999). "Genetics of aging in Drosophila." Experimental Gerontology 34(5): 577-585.

Roses, A. D. (1997). "Apolipoprotein E, a gene with complex biological interactions in the aging brain." Neurobiology of Disease 4(3-4): 170-185.

Rothenbacher, D., Hoffmeister, A., Brenner, H. and Koenig, W. (2003). "Physical activity, coronary heart disease, and inflammatory response." Archives of Internal Medicine 163(10): 1200-1205.

Roux, A. V. D., Chambless, L., Markin, S. S., Arnett, D., Eigenbrodt, M., Nieto, F. J., Szklo, M. and Sorlie, P. (2002). "Socioeconomic disadvantage associated and change in blood pressure with aging." Circulation 106(6): 703-710.

Rudel, L. L., Parks, J. S., Hedrick, C. C., Thomas, M. and Williford, K. (1998). "Lipoprotein and cholesterol metabolism in diet-induced coronary artery atherosclerosis in primates, role of cholesterol and fatty acids." Progress in Lipid Research 37(6): 353-370.

Sacks, F. M. and Katan, M. (2002). "Randomized clinical trials on the effects of dietary fat and carbohydrate on plasma lipoproteins and cardiovascular disease." American Journal of Medicine 113: 13-24.

Sapolsky, R. M. and Finch, C. E. (2000). Alzheimer's disease and some speculations about the evolution of its modifiers.In: *Alzheimer's Disease: A Compendium of Current Theories*. **924:** 99-103.

Sayer, A. A., Cooper, C., Evans, J. R., Rauf, A., Wormald, R. P. L., Osmond, C. and Barker, D. J. P. (1998). "Are rates of ageing determined in utero?" Age and Ageing 27(5): 579-583.

Schachter, F., Faure-delanef, L., Guenot, F., Rouger, H., Froguel, P., Lesueur-ginot, L. and Cohen, D. (1994). "Genetic associations with human longevity at the Apoe and Ace loci." Nature Genetics 6(1): 29-32.

Schmitz, K. H., Schreiner, P. J., Jacobs, D. R., Leon, A. S., Liu, K., Howard, B. and Sternfeld, B. (2001). "Independent and interactive effects of apolipoprotein E phenotype and cardiorespiratory fitness on plasma lipids." Annals of Epidemiology 11(2): 94-103.

Schunkert, H., Hense, H. W., Holmer, S. R., Stender, M., Perz, S., Keil, U., Lorell, B. H. and Riegger, G. A. (1994). "Association between a deletion polymorphism of the angiotensin-converting-enzyme gene and left ventricular hypertrophy." New England Journal of Medicine 330(23): 1634-8.

Scupin, R. (1992). Cultural Anthropology: A Global Perspective. Prentice Hall, Englewood Cliffs, NJ.

Semenza, J. C., Rubin, C. H., Falter, K. H., Selanikio, J. D., Flanders, W. D., Howe, H. L. and Wilhelm, J. L. (1996). "Heat-related deaths during the July 1995 heat wave in Chicago." New England Journal of Medicine 335(2): 84-90.

Service, P. M., Hutchinson, E. W. and Rose, M. R. (1988). "Multiple genetic mechanisms for the evolution of senescence in Drosophila-melanogaster." Evolution 42(4): 708-716.

Sgro, C. M. and Partridge, L. (1999). "A delayed wave of death from reproduction in Drosophila." Science 286(5449): 2521-2524.

Sgro, C. M., Geddes, G., Fowler, K. and Partridge, L. (2000). "Selection on age at reproduction in Drosophila melanogaster: Female mating frequency as a correlated response." Evolution 54(6): 2152-2155.

Shanley, D. P. and Kirkwood, T. B. L. (2000). "Calorie restriction and aging: A life-history analysis." Evolution 54(3): 740-750.

Shiell, A. W., Campbell-Brown, M., Haselden, S., Robinson, S., Godfrey, K. M. and Barker, D. J. P. (2001). "High-meat, low-carbohydrate diet in pregnancy -

Relation to adult blood pressure in the offspring." Hypertension 38(6): 1282-1288.

Shih, S. J., Allan, C., Grehan, S., Tse, E., Moran, C. and Taylor, J. M. (2000). "Duplicated downstream enhancers control expression of the human apolipoprotein E gene in macrophages and adipose tissue." Journal of Biological Chemistry 275(41): 31567-31572.

Silbermann, R. and Tatar, M. (2000). "Reproductive costs of heat shock protein in transgenic Drosophila melanogaster." Evolution 54(6): 2038-2045.

Sinclair, D. A. and Guarente, L. (1997). "Extrachromosomal rDNA circles - A cause of aging in yeast." Cell 91(7): 1033-1042.

Smith, J. D. (2002). "Apolipoproteins and aging: emerging mechanisms." Ageing Research Reviews 1(3): 345-365.

Snoke, M. S. and Promislow, D. E. L. (2003). "Quantitative genetic tests of recent senescence theory: age- specific mortality and male fertility in Drosophila melanogaster." Heredity 91(6): 546-556.

Sober, E. (1993). "Mathematics and indispensability + Scientific realism and empirical explanation." Philosophical Review 102(1): 35-57.

Sober, E. (2002). "Instrumentalism, parsimony, and the Akaike framework." Philosophy of Science 69(3): S112-S123.

Sophocles (1889). The Oedipus at Colonus of Sophocles. Cambridge University Press, Cambridge.

Spencer, C. C., Howell, C. E., Wright, A. R. and Promislow, D. E. L. (2003). "Testing an 'aging gene' in long-lived Drosophila strains: increased longevity depends on sex and genetic background." Aging Cell 2(2): 123-130.

Stampfer, M. J., Hu, F. B., Manson, J. E., Rimm, E. B. and Willett, W. C. (2000). "Primary prevention of coronary heart disease in women through diet and lifestyle." New England Journal of Medicine 343(1): 16-22.

Stansfeld, S. A., Fuhrer, R., Shipley, M. J. and Marmot, M. G. (2002). "Psychological distress as a risk factor for coronary heart disease in the Whitehall II Study." International Journal of Epidemiology 31(1): 248-255.

Stearns, S. C. (1992). The Evolution of Life Histories. Oxford University Press.

Stearns, S. C., Ackermann, M., Doebeli, M. and Kaiser, M. (2000). "Experimental evolution of aging, growth, and reproduction in fruitflies." Proceedings of the

National Academy of Sciences of the United States of America 97(7): 3309-3313.

Stein, C. E., Fall, C. H. D., Kumaran, K., Osmond, C., Cox, V. and Barker, D. J. P. (1996). "Fetal growth and coronary heart disease in South India." The Lancet 348(9037): 1269-1273.

Stengard, J. H., Pekkanen, J., Ehnholm, C., Nissinen, A. and Sing, C. F. (1996). "Genotypes with the apolipoprotein epsilon 4 allele are predictors of coronary heart disease mortality in a longitudinal study of elderly Finnish men." Human Genetics 97(5): 677-684.

Stengard, J. H., Weiss, K. M. and Sing, C. F. (1998). "An ecological study of association between coronary heart disease mortality rates in men and the relative frequencies of common allelic variations in the gene coding for apolipoprotein E." Human Genetics 103(2): 234-241.

Stevens, R. F. (1999). "The history of haemophilia in the Royal Families of Europe." British Journal of Haematology 105(1): 25-32.

Sun, J. and Tower, J. (1999). "FLP recombinase-mediated induction of Cu/Zn-superoxide dismutase transgene expression can extend the life span of adult

Drosophila melanogaster flies." Molecular and Cellular Biology 19(1): 216-228.

Tanasescu, N., Leitzmann, M. F., Rimm, E. B., Willett, W. C., Stampfer, M. J. and Hu, F. B. (2002). "Exercise type and intensity in relation to coronary heart disease in men." Jama-Journal of the American Medical Association 288(16): 1994-2000.

Tang, M., Chen, Y. and Krewski, D. (2003). "Gender-related differences in the association between socioeconomic status and self-reported diabetes." International Journal of Epidemiology 32(3): 381-385.

Tang, M. X., Jacobs, D., Stern, Y., Marder, K., Schofield, P., Gurland, B., Andrews, H. and Mayeux, R. (1996). "Effect of oestrogen during menopause on risk and age at onset of Alzheimer's disease." Lancet 348(9025): 429-432.

Tatar, M., Promislow, D. E. I., Khazaeli, A. A. and Curtsinger, J. W. (1996). "Age-specific patterns of genetic variance in Drosophila melanogaster .2. Fecundity and its genetic covariance with age- specific mortality." Genetics 143(2): 849-858.

Tatar, M., Kopelman, A., Epstein, D., Tu, M. P., Yin, C. M. and Garofalo, R. S. (2001). "A mutant Drosophila insulin receptor homolog that extends life- span and impairs neuroendocrine function." Science 292(5514): 107-110.

Teasdale, G. M., Nicoll, J. A. R., Murray, G. and Fiddes, M. (1997). "Association of apolipoprotein E polymorphism with outcome after head injury." Lancet 350(9084): 1069-1071.

Thomas, F., Teriokhin, A. T., Renaud, F., De Meeus, T. and Guegan, J. F. (2000). "Human longevity at the cost of reproductive success: evidence from global data." Journal of Evolutionary Biology 13(3): 409-414.

Thompson, P. D. and Rader, D. J. (2001). "Does exercise increase HDL cholesterol in those who need it the most?" Arteriosclerosis Thrombosis and Vascular Biology 21(7): 1097-1098.

Tissenbaum, H. A. and Ruvkun, G. (1998). "An insulin-like signaling pathway affects both longevity and reproduction in Caenorhabditis elegans." Genetics 148(2): 703-718.

Turic, D., Fisher, P. J., Plomin, R. and Owen, M. J. (2001). "No association between apolipoprotein E polymorphisms and general cognitive ability in children." Neuroscience Letters 299(1-2): 97-100.

Twisk, J. W. R., Kemper, H. C. G., van Mechelen, W. and Post, G. B. (2001). "Clustering of risk factors for coronary heart disease: The longitudinal relationship with lifestyle." Annals of Epidemiology 11(3): 157-165.

van Bockxmeer, F. M. (1994). "Apoe and Ace genes - Impact on human longevity." Nature Genetics 6(1): 4-5.

van der Meer, I. M. and Witteman, J. C. M. (2002). "Apolipoprotein E genotype, smoking and cardiovascular disease." Journal of Hypertension 20(12): 2327-2329.

van Doorslaer, E., Wagstaff, A., van der Burg, H., Christiansen, T., De Graeve, D., Duchesne, I., Gerdtham, U. G., Gerfin, M., Geurts, J., Gross, L., Hakkinen, U., John, J., Klavus, J., Leu, R. E., Nolan, B., O'Donnell, O., Propper, C., Puffer, F., Schellhorn, M., Sundberg, G. and Winkelhake, O. (2000). "Equity in the delivery of health care in Europe and the US." Journal of Health Economics 19(5): 553-583.

Walsh, T. J. and Brake, J. (1999). "Effects of feeding program and crude protein intake during rearing on fertility of broiler breeder females." Poultry Science 78(6): 827-832.

Wannamethee, S. G., Whincup, P. H., Shaper, G. and Walker, M. (1996). "Influence of fathers' social class on cardiovascular disease in middle-aged men." The Lancet 348(9037): 1259-1263.

Waterland, R. A. and Garza, C. (2002a). "Early postnatal nutrition determines adult pancreatic glucose- responsive insulin secretion and islet gene expression in rats (vol 132, pg 357, 2002)." Journal of Nutrition 132(5): 1055-1055.

Waterland, R. A. and Garza, C. (2002b). "Early postnatal nutrition determines adult pancreatic glucose- responsive insulin secretion and islet gene expression in rats." Journal of Nutrition 132(3): 357-364.

Weggemans, R. M., Zock, P. L., Ordovas, J. M., Pedro-Botet, J. and Katan, M. B. (2001). "Apoprotein E genotype and the response of serum cholesterol to dietary fat, cholesterol and cafestol." Atherosclerosis 154(3): 547-555.

Weismann, A. (1881). The Duration of Life.In: *Essays upon heredity and kinred biological problems, vol I.* Clarendon Press. Oxford.

Westendorp, R. G. and Kirkwood, T. B. L. (1999a). "Human longevity and reproductive success: Repsonse to Gacrilov and Gavrilova." Journal of Anti-aging Medicine 2(2): 125-126.

Westendorp, R. G. J. and Kirkwood, T. B. L. (1998). "Human longevity at the cost of reproductive success." Nature 396(6713): 743-746.

Westendorp, R. G. J. and Kirkwood, T. B. L. (1999b). "Longevity - does family size matter? Reply." Nature 399(6736): 522-522.

Westendorp, R. G. J., van Dunne, G. M., Kirkwood, T. B. L., Helmerhorst, F. M. and Huizinga, T. W. J. (2001). "Optimizing human fertility and survival." Nature Medicine 7(8): 873-873.

Williams, G. C. (1957). "Pleiotropy, natural selection, and the evolution of senescence." Evolution 11: 398-411.

Wilson, P. W. F., Schaefer, E. J., Larson, M. G. and Ordovas, J. M. (1996). "Apolipoprotein E alleles and risk of coronary disease - A meta- analysis." Arteriosclerosis Thrombosis and Vascular Biology 16(10): 1250-1255.

Wilson, P. W. F. M. D., Myers, R. H. P., Larson, M. G. S., Ordovas, J. M. P., Wolf, P. A. M. D. and Schaefer, E. J. M. D. (1994). "Apolipoprotein E Alleles,

Dyslipidemia, and Coronary Heart Disease: The Framingham Offspring Study." Jama-Journal of the American Medical Association 272(21): 1666-1671.

Wolfram, S. (1999). The Mathematica Book, 4th ed. Wolfram Media/ Cambridge University Press.

Woo, J., Ho, S. C. and Yu, A. L. M. (2000). "The influence of income on morbidity, mortality and dependency in elderly Hong Kong Chinese." Archives of Gerontology and Geriatrics 30(1): 55-61.

Wrigley, E., Davies, R., Oeppen, J. and Schofield, R. (1997). English population history from family reconstitution 1580-1837, Cambridge University Press.

Yu, V. W. Y., Lin, C. H., Chen, S. P., Hong, C. J. and Tsai, S. J. (2000). "Intelligence and event-related potentials for young female human volunteer apolipoprotein E epsilon 4 and non-epsilon 4 carriers." Neuroscience Letters 294(3): 179-181.

Yu, Z. J., Nissinen, A., Vartiainen, E., Hu, G., Tian, H. G. and Guo, Z. Y. (2002). "Socio-economic status and serum lipids: A cross-sectional study in a Chinese urban population." Journal of Clinical Epidemiology 55(2): 143-149.

Zakhari, S. (1997). "Alcohol and the cardiovascular system - Molecular mechanisms for beneficial and harmful action." Alcohol Health & Research World 21(1): 21-29.

Zerba, K. E., Ferrell, R. E. and Sing, C. F. (1996). "Genotype-environment interaction: Apolipoprotein E (ApoE) gene effects and age as an index of time and spatial context in the human." Genetics 143(1): 463-478.

Zhang, J. and Yu, K. F. (1998). "What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes." Jama-Journal of the American Medical Association 280(19): 1690-1691.

Zmuda, J. M., Yurgalevitch, S. M., Flynn, M. M., Bausserman, L. L., Saratelli, A., Spannaus-Martin, D. J., Herbert, P. N. and Thompson, P. D. (1998). "Exercise training has little effect on HDL levels and metabolism in men with initially low HDL cholesterol." Atherosclerosis 137(1): 215-221.

Zwaan, B. J., (1993). *Genetical and Environmental Aspects of Ageing in Drosophila melanogaster* Thesis (PhD), University of Groningen

Zwaan, B. J. (1999). "The evolutionary genetics of ageing and longevity." Heredity 82: 589-597.

BLANK PAGE IN ORIGINAL

Appendix 1. *Mathematica* notebook for the implementation of the disposable soma theory of ageing.

The disposable soma theory is based on the framework of physiological ecology and its main target is to describe the optimal balance between reproduction and maintenance in a given organism under specific conditions.

To do that, ecological models can be used. Fitness can be defined in terms of the Malthusian parameter r, of the Euler-Lotka equation:

$$\int e^{-r\,x}\, l[x;\, s]\, m[x;\, s]\, dx = 1,$$

where survival l[x;s] and fecundity m[x;s] are written as functions depending on age x and investment in maintenance s. From the above, the effect of different values of maintenance on survivorship and fecundity can be seen, and most importantly curves relating r to s can be obtained.

The equations for l[x; s] and m[x; s] derived from the Gompertz model are

$$l[x;\, s] = (1 - V)\, \text{Exp}[-\alpha\,(e^{\beta x} - e^{\beta a})/\beta - \gamma\,(x - a)]$$

$$m[x;\, s] = h\, \text{Exp}[-\alpha\,(e^{\beta x} - e^{\beta a})/\beta],\ \text{for}\ x \geq a$$

where
V = total juvenile mortality
α = basal vulnerability
β = rate of ageing
γ = extrinsic age independent adult mortality as described in the Gompertz – Makeham equation
$\mu = \alpha\, e^{\beta x} + \gamma$ Gompertz – Makeham equation
a = age of maturity and
h = peak rate of reproduction at age a

The dependence on s is introduced in β, a, and h as follows

$\beta = \beta o\left(\frac{so}{s} - 1\right)$ for s < so and
$\beta = 0$ for s ≥ so
so is the level of investment in maintenance required for no senescence to occur

$$a = \frac{ao}{1 - s}$$

ao = the age of maturity when all energy is invested in reproduction

$h = ho\,(1 - s)$
ho = reproductive rate when all the energy is invested in reproduction

The original model was specific for the life history of the mouse where menopause does not occur. In the present context, where humans are considered, menopause shortens the reproductive life span much earlier than the actual longevity of the organism and so menopause must be included in the model. Here we have set the age of menopause at

50, independent of investment. In a normal human population we might expect that to be able to change with increasing s. Due to the parameters chosen in this case, the individuals of this population are especially short-lived, mirroring primitive conditions of human evolution, so that the time of menopause is of little importance as we shall see later.

The following parameters can be given values according to our current knowledge about the life history of primitive human populations.

Age of first reproduction in years.

```
a = 16;
```

Reproductive rate in birth per woman per year

```
h = 0.5;
```

α and β rates of ageing are

```
α := 0.001
```

$\beta = 0.1305$

% juvenile mortality in a moderately protected environment can be

```
V := 0.4
```

And adult age independent mortality

```
γ := 0.01
```

Considering that the level of investment required to succesfully maintain the organism free of damage and that minimum possible age of reproduction are

```
ao := 8
so := 0.8
```

the values of ho and βo can be calculated as

```
βo := 0.2175
ho := 1
```

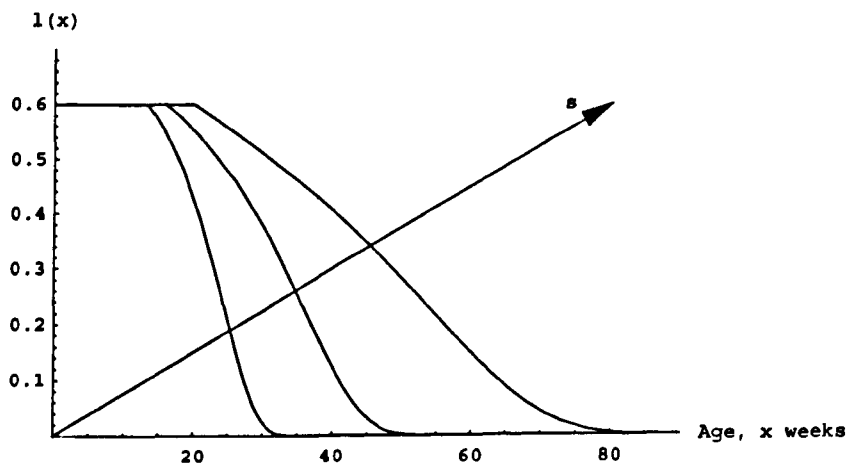
Then the equations below can be solved

```
<< Graphics'MultipleListPlot'
<< Graphics'Arrow'

β[s_] /; (s < so) := βo (so / s - 1)
β[s_] /; (s ≥ so) := 0

a[s_] := ao / (1 - s)

h[s_] := ho (1 - s)

l[x_, s_] /; x ≥ a[s] := (1 - V) Exp[(-α (e^(β[s] x) - e^(β[s] a[s])) / β[s]) - (γ (x - a[s]))]
l[x_, s_] /; x < a[s] := 1 - V
Clear[m]
menopause = 50;
m[x_, s_] /; menopause ≥ x ≥ a[s] := h[s] Exp[-α (e^(β[s] x) - e^(β[s] a[s])) / β[s]]
m[x_, s_] /; x < a[s] := 0
m[x_, s_] /; x ≥ menopause := 0
```

The dependance of survivorship and fecundity on the investment in maintenance can be graphically represented.

Survival

```
Plot[{l[x, 0.4], l[x, 0.5], l[x, 0.6]}, {x, 0, 125}, PlotRange -> {{0, 90}, {0, 0.7}},
  AxesLabel → {"Age, x weeks", "l(x)"}, Epilog → {Arrow[{0, 0}, {80, 0.6}], Text["s", {74, .6}]}]
```

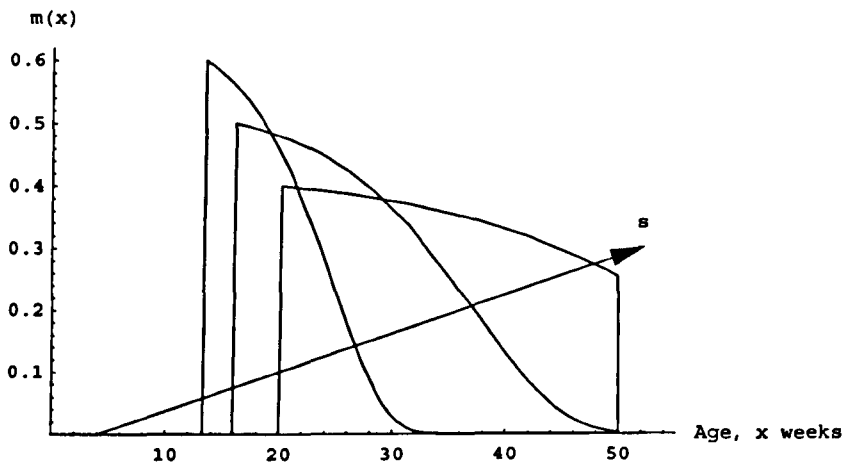



```
- Graphics -
```

The three lines have an s of 0.4, 0.5 and 0.6 from left to right. So, higher investment in maintenance increases the survival of the organism.

Fecundity

```
Plot[{m[x, 0.4], m[x, 0.5], m[x, 0.6]}, {x, 0, 125}, PlotRange -> {{0, 55}, {0, 0.62}},
  AxesLabel → {"Age, x weeks", "m(x)"}, Epilog → {Arrow[{4, 0}, {52, 0.3}], Text["s", {52, .34}]}]
```

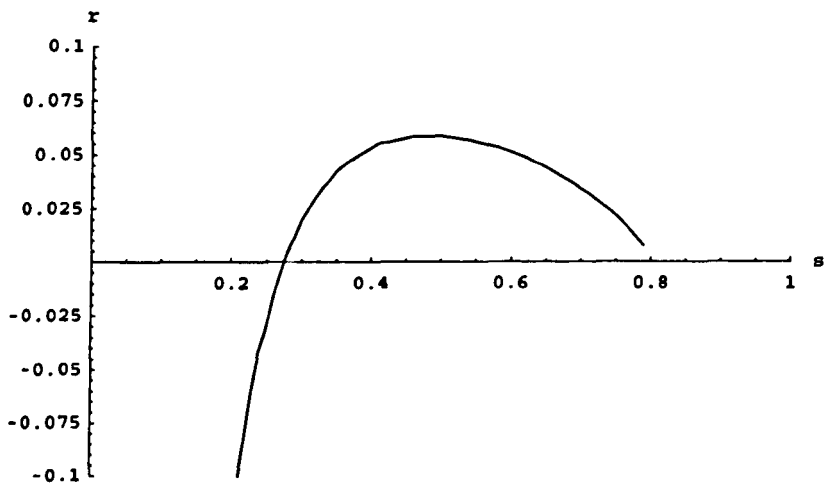



```
- Graphics -
```

The three lines with **s** = 0.4, 0.5, and 0.6 from left to right show that higher investment in maintenance results in longer reproduction with a lower starting point and a less steep decline later in life.

To find the optimal strategy for the organism the above equations must be combined to give the dependence of **r** on **s**

```
Timing[
 ListPlot[Join[Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]},
    {s, 0.01, 0.41, .01}],
   Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]},
    {s, 0.42, 0.62, 0.04}],
   Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]},
    {s, 0.62, 0.79, .01}]], PlotJoined → True,
  AxesLabel → {"s", "r"}, PlotRange -> {{0, 1}, {-0.1, 0.1}}]]
```



The highest point of the curve describes the optimal strategy for humans under the specific conditions considered. The optimum investment in maintenance is close to $s = 0.5$, which gives a 5% annual increase of the population.

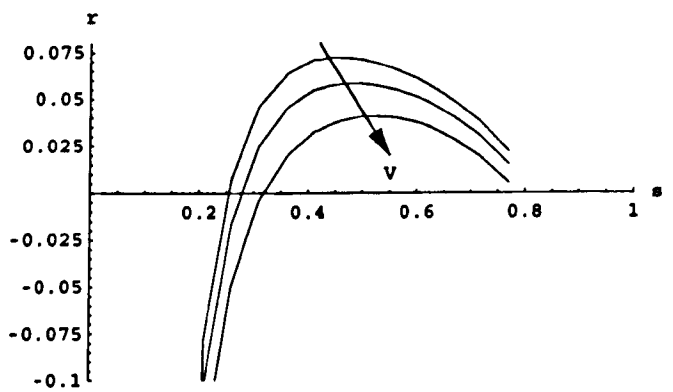
Effects of changing the parameters of the population

The following commands turn off a number of messages that the program prints to let us know that some of the calculations exceed the precision depth of the application. Despite this loss of precision our results have been verified as accurate.

```
Off[General::unfl]
Off[NIntegrate::ploss]
Off[FindRoot::frsec]
Off[NIntegrate::ncvb]
Off[NIntegrate::slwcon]
Off[Graphics::gprim]
```

Effect of changing total juvenile mortality

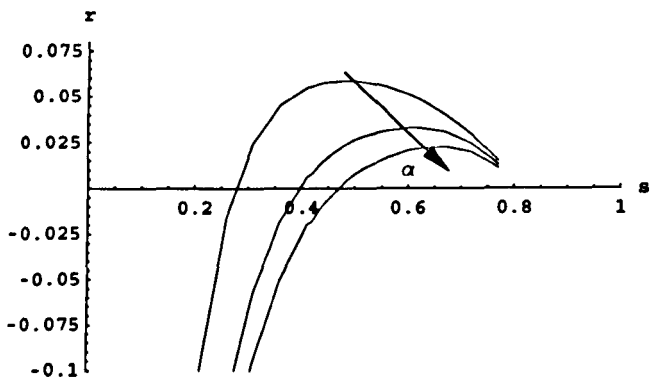
```
Clear[V];
Timing[MultipleListPlot[
  Table[Join[Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.01, 0.41, 0.05}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.42, 0.62, 0.02}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.62, 0.79, 0.05}]],
   {V, 0.2, 0.6, 0.2}], PlotJoined → True, SymbolShape → {None}, PlotStyle -> {Line},
  AxesLabel → {"s", "r"}, PlotRange -> {{0, 1}, {-0.1, 0.08}},
  Epilog → {Arrow[{0.42, 0.08}, {0.55, 0.02}], Text["V", {0.55, .01}]}]]
```



The increase of juvenile mortality exerts its effects both in the maximum value of **r** reached by the population and also on the optimal investment in maintenance required. As very early mortality is increased, the survival of the adult organism also has to increase to achieve its optimal strategy.
Values for V→ 0.2, 0.4, 0.6

Effect of changing rate of ageing α

```
Clear[α];
Timing[MultipleListPlot[
  Table[Join[Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.01, 0.41, 0.05}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.42, 0.62, 0.02}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.62, 0.79, 0.05}]],
   {α, 0.001, 0.01, 0.0045}], PlotJoined → True, SymbolShape → {None},
  PlotStyle -> {Line}, AxesLabel → {"s", "r"}, PlotRange -> {{0, 1}, {-0.1, 0.08}},
  Epilog → {Arrow[{0.48, 0.063}, {0.675, 0.01}], Text["α", {0.6, .01}]}]]
```



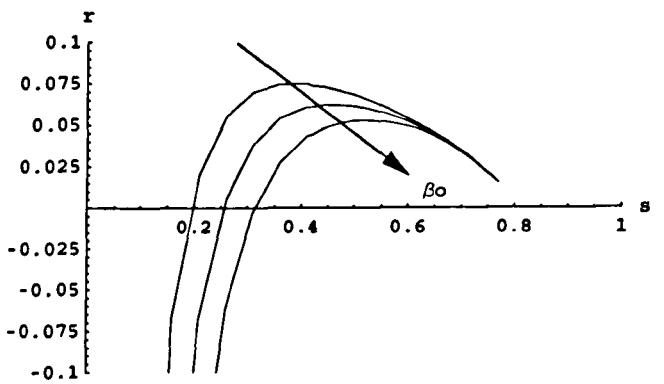

```
{29.344 Second, - Graphics -}
```

Since α is the basal vulnerability of the organism it is expected that it would have an impact on the balance between r and s required to reach the point of maximum population increase. This case is similar to what we saw when V changed. Again a frail population will have to invest more in maintenance to be able to reach its optimum reproductive potential.

Values of $\alpha \rightarrow$ 0.001, 0.0055, 0.01

Effect of changing rate of ageing β

```
Clear[βo];
Timing[MultipleListPlot[
  Table[Join[Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.01, 0.41, 0.05}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.42, 0.62, 0.02}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.62, 0.79, 0.05}]],
   {βo, 0.15, 0.25, 0.05}], PlotJoined → True, SymbolShape→ {None},
  PlotStyle -> {Line}, AxesLabel → {"s", "r"}, PlotRange -> {{0, 1}, {-0.1, 0.1}},
  Epilog → {Arrow[{0.28, 0.099}, {0.6, 0.02}], Text["βo", {0.65, .01}]}]]
```

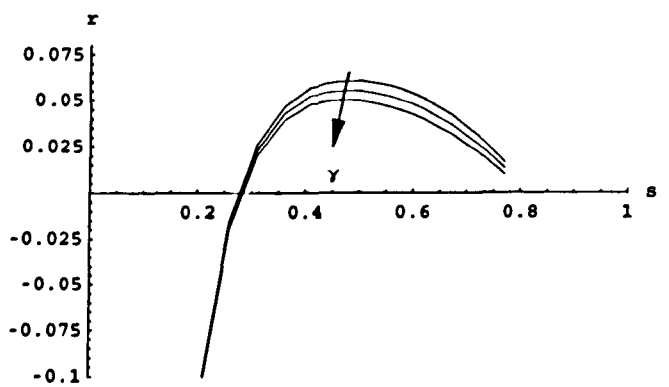



```
{30.547 Second, - Graphics -}
```

When the individuals of a population are ageing faster (higher βo) those with a higher investment in maintenance will be able to do better since they will able to live for longer and at the same time be more fecund.
Values of βo → 0.15, 0.20, 0.25

Effects of changing extrinsic age independent adult mortality γ

```
Clear[γ];
Timing[MultipleListPlot[
  Table[Join[Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.01, 0.41, 0.05}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.42, 0.62, 0.02}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.62, 0.79, 0.05}]],
   {γ, 0.02, 0.06, 0.02}], PlotJoined → True, SymbolShape → {None},
  PlotStyle -> {Line}, AxesLabel → {"s", "r"}, PlotRange -> {{0, 1}, {-0.1, 0.08}},
  Epilog → {Arrow[{0.48, 0.065}, {0.45, 0.025}], Text["γ", {0.45, .01}]}]]
```

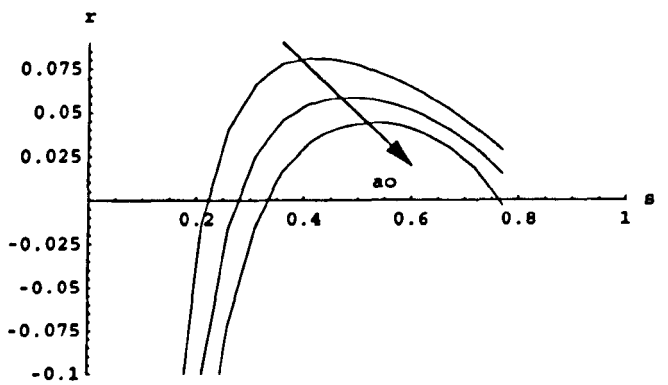



```
{74.953 Second, - Graphics -}
```

As γ increases the amount in investment needed for the optimal strategy decreases. The optimum strategy is thus to invest more and more in reproduction than maintenance since the expected life span of the organism is reduced due to a higher probability of death by external random causes. This is one of the central aspects of the disposable soma theory under which the differences in lifespan between organisms is attributed to expected longevity in the wild.
Values for $\gamma \rightarrow$ 0.0025, 0.2125, 0.04

Effects of changing minimum age at maturity ao

```
Clear[ao];
Timing[MultipleListPlot[
  Table[Join[Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.01, 0.41, 0.05}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.42, 0.62, 0.02}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.62, 0.79, 0.05}]], {ao, 6, 10, 2}],
  PlotJoined → True, SymbolShape → {None}, PlotStyle -> {Line}, AxesLabel → {"s", "r"},
  PlotRange -> {{0, 1}, {-0.1, 0.09}}, Epilog → {Arrow[{0.36, 0.09}, {0.6, 0.02}], Text["ao", {0.55, .01}]}]]
```

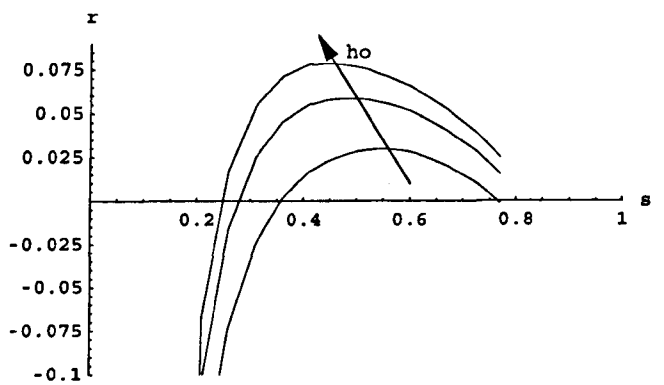



```
{69.843 Second, - Graphics -}
```

As the age of maturity increases investment in maintenance also increases, ensuring that the organism is in good condition to start reproducing later in life than previously.
Values of ao → 6, 8, 10

Effects of change of maximum rate of reproduction ho

```
Clear[ho]
Timing[MultipleListPlot[
  Table[Join[Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.01, 0.41, 0.05}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.42, 0.62, 0.02}],
    Table[{s, r /. FindRoot[∫_0^100 e^(-r x) l[x, s] m[x, s] dx == 1, {r, -1, 1}]}, {s, 0.62, 0.79, 0.05}]],
   {ho, 0.5, 1.5, 0.5}], PlotJoined → True, SymbolShape → {None},
  PlotStyle -> {Line}, AxesLabel → {"s", "r"}, PlotRange -> {{0, 1}, {-0.1, 0.09}},
  Epilog → {Arrow[{0.6, 0.01}, {0.425, 0.095}], Text["ho", {0.5, .085}]}]]
```




```
{30.625 Second, - Graphics -}
```

As the maximum rate of reproduction increases, less and less investment in maintenance is required to keep the population at its maximum rate of increase.

Values for ho → 0.5, 1.0, 1.5

Appendix 2. *Mathematica* notebook for the implementation of the immunogenetic model for the trade-offs between human fertility and lifespan.

To illustrate our model we are going to use a very simple example. We will first consider the optimum strategies required in two very different populations, one from The Gambia where the infection load is large and one from Denmark where death due to infections is almost non-existent. Let us then imagine a woman from The Gambia, where she has to face the challenges of infections, moving to Denmark. We will follow her progeny and the change of the allele frequencies under the new conditions.

To do that we will use the Denmark life tables to calculate the Gompertz equation parameters and then we will consider that the difference between the Danish and Gambian mortality curves are mainly due to infections and can affect the fitness of η.
The first step is to consider that there is no external mortality in Denmark, thus the Gompertz parameters can be calculated directly by the Danish life tables with the use of a non linear fitting algorithm. The ages used are between 9 and 94 since they are the range over which the Gompertz curve gives the best fit.

Loading Packages

```
<< Statistics'NonlinearFit'
```

Denmark

Data (WHO)

```
DNK = {{9, 0.000452}, {14, 0.000736}, {19, 0.000922},
    {24, 0.001293}, {29, 0.001596}, {34, 0.00265}, {39, 0.004688}, {44, 0.008466},
    {49, 0.012113}, {54, 0.019533}, {59, 0.0333}, {64, 0.057348}, {69, 0.094244},
    {74, 0.139863}, {79, 0.205224}, {84, 0.312648}, {89, 0.456124}, {94, 0.595709}};
```

Fitting an equation

```
theorDNK =
  NonlinearFit[DNK, alpha Exp[beta x], {x}, {{alpha, 0.0002}, {beta, 0.08}}]
```

$$0.000625216\, e^{0.0733576\, x}$$

```
NonlinearRegress[DNK, alpha Exp[beta x], {x}, {{alpha, 0.0002}, {beta, 0.01}}]
```

$$\{BestFitParameters \to \{alpha \to 0.000625251,\ beta \to 0.073357\},\ ParameterCITable \to$$

	Estimate	*Asymptotic SE*	*CI*
alpha	*0.000625251*	*0.000117851*	*{0.000375419, 0.000875084}*,
beta	*0.073357*	*0.00210295*	*{0.0688989, 0.077815}*

$$EstimatedVariance \to 0.000144161,$$

$$ANOVATable \to$$

	DF	*SumOfSq*	*MeanSq*
Model	*2*	*0.733953*	*0.366977*
Error	*16*	*0.00230657*	*0.000144161*,
Uncorrected Total	*18*	*0.73626*	
Corrected Total	*17*	*0.525679*	

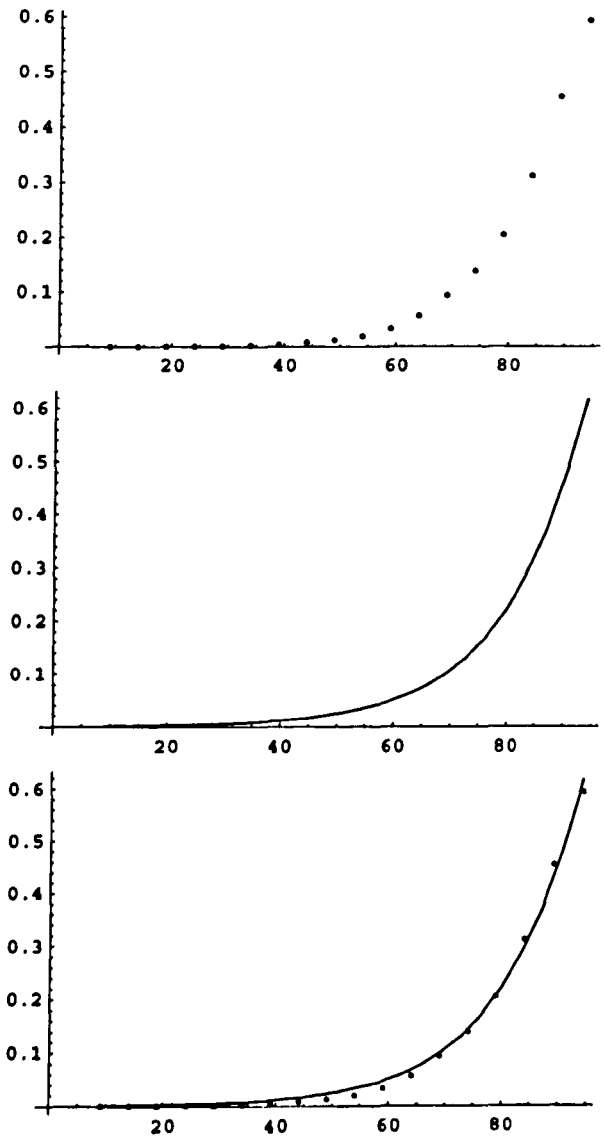
$$AsymptoticCorrelationMatrix \to \begin{pmatrix} 1. & -0.997232 \\ -0.997232 & 1. \end{pmatrix},$$

$$FitCurvatureTable \to$$

	Curvature
Max Intrinsic	*0.0394309*
Max Parameter-Effects	*3.5654*
95. % Confidence Region	*0.524595*

Plots

Show[ListPlot[DNK], Plot[theorDNK, {x, 9, 94}]]



- Graphics -

We will consider that β is a characteristic of the species so we can use the same value for both populations. We define βo as the value of β when the optimum investment in maintenance is half the one required to have no age dependent mortality.

Gambia

Data (WHO)

```
GMB = {{9, 0.00935198}, {14, 0.00704255}, {19, 0.01312712},
    {24, 0.02180081}, {29, 0.02999855}, {34, 0.03581704}, {39, 0.03885223},
    {44, 0.04203231}, {49, 0.04908543}, {54, 0.06282653}, {59, 0.08236681},
    {64, 0.11697598}, {69, 0.17011485}, {74, 0.25559405}, {79, 0.36989701},
    {84, 0.51931764}, {89, 0.68798683}, {94, 0.79402981}};
```

Fitting an equation

```
theorGMB = NonlinearFit[GMB,
  alpha Exp[0.0733576 x] + gamma, {x}, {{alpha, 0.0006}, {gamma, 0.1}}]
```

$$0.0269311 + 0.000878084\, e^{0.0733576\, x}$$

```
NonlinearRegress[GMB, alpha Exp[0.0733576 x] + gamma,
  {x}, {{alpha, .0002}, {gamma, 0.1}}]
```

$$\{BestFitParameters \to \{alpha \to 0.000878084,\ gamma \to 0.0269311\},\ ParameterCITable \to$$

	Estimate	*Asymptotic SE*	*CI*
alpha	*0.000878084*	*0.0000345661*	*{0.000804807, 0.000951361},*
gamma	*0.0269311*	*0.011164*	*{0.00326461, 0.0505977}*

$$EstimatedVariance \to 0.00155808,$$

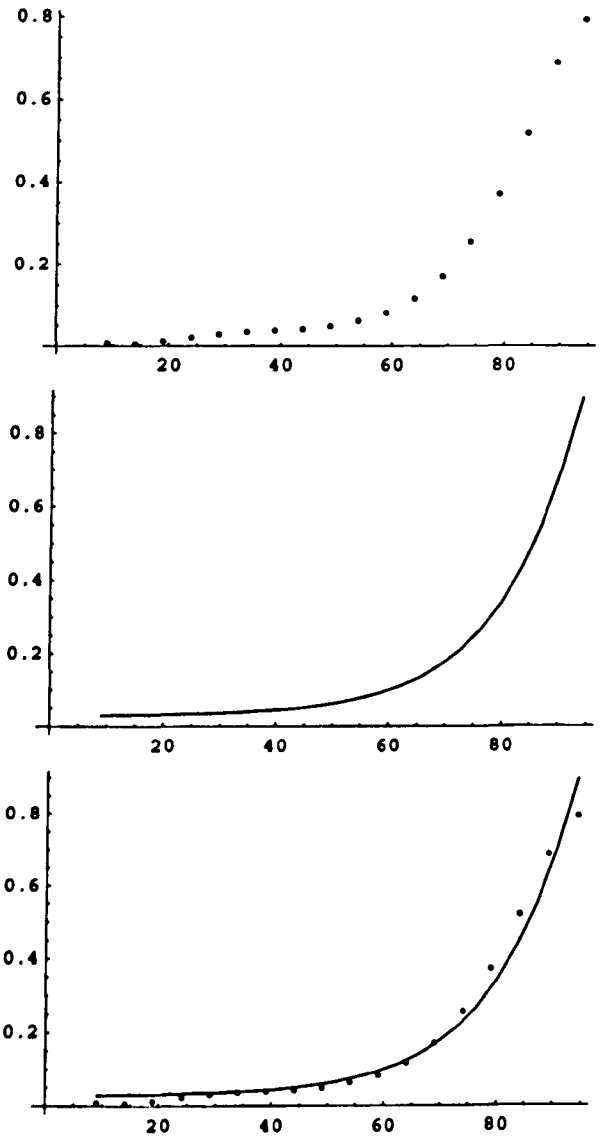
		DF	*SumOfSq*	*MeanSq*
	Model	*2*	*1.61273*	*0.806365*
ANOVATable →	*Error*	*16*	*0.0249292*	*0.00155808,*
	Uncorrected Total	*18*	*1.63766*	
	Corrected Total	*17*	*1.03038*	

$$AsymptoticCorrelationMatrix \to \begin{pmatrix} 1. & -0.55271 \\ -0.55271 & 1. \end{pmatrix},$$

		Curvature
FitCurvatureTable →	*Max Intrinsic*	*0*
	Max Parameter–Effects	*0*
	95. % Confidence Region	*0.524595*

Plots

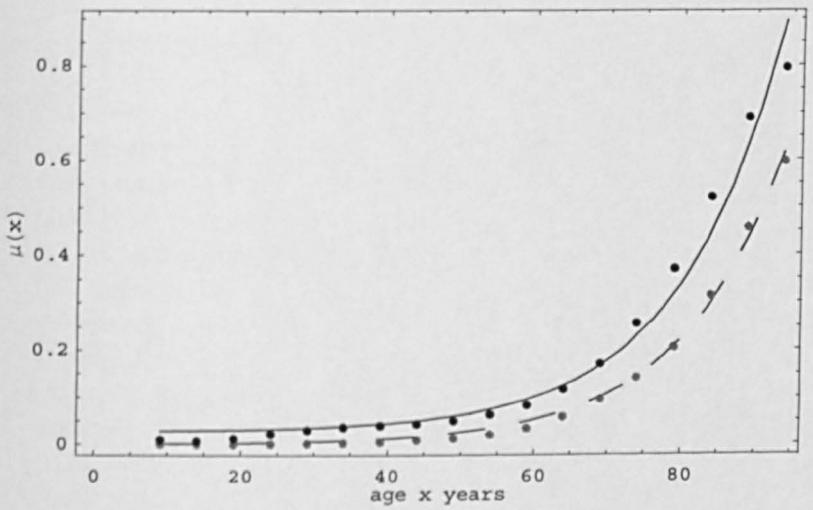
Show[ListPlot[GMB], Plot[theorGMB, {x, 9, 94}]]



- Graphics -

Now we can compare the WHO data values (points) and their fitted models (lines), for Denmark (dashed line and grey points) and the Gambia (solid line and black points) for the increase of the force of mortality with age

```
Show[Plot[theorDNK, {x, 9, 94}, PlotStyle → Dashing[{0.05, 0.05}]],
  Plot[theorGMB, {x, 9, 94}],
  ListPlot[DNK, PlotStyle → {AbsolutePointSize[4], GrayLevel[0.5]}],
  ListPlot[GMB, PlotStyle → AbsolutePointSize[4]], Frame → True,
  FrameLabel → {"age x years", "μ(x)"}, ImageSize → 350]
```



- Graphics -

Having found the values of the Gompertz parameters we can find the contour plots for each set of conditions. From the data we can calculate the weighted mean of r for each order of η assuming that negative values of r do not exist in the population. The indeterminate points are just those for which r is negative.

For Denmark

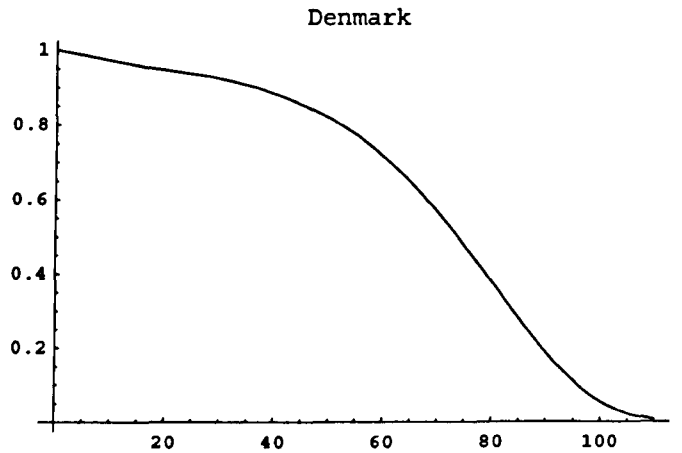
- Turning off messages

```
Off[General::spell]
Off[General::unfl]
Off[NIntegrate::ploss]
Off[FindRoot::frsec]
Off[NIntegrate::ncvb]
Off[NIntegrate::slwcon]
Off[Graphics::gprim]
Off[Syntax::com]
```

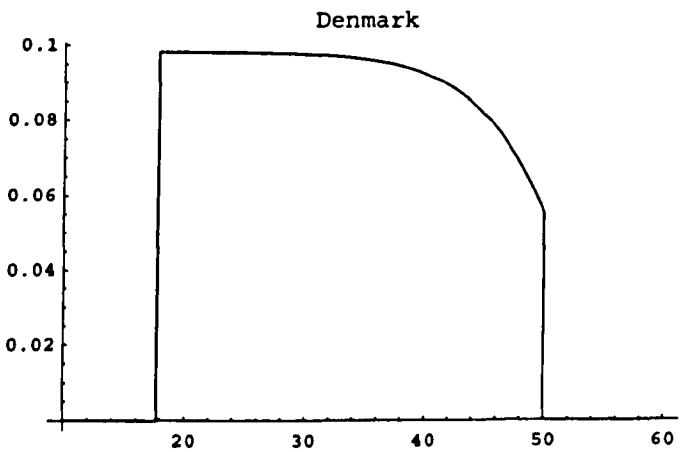
- Equations and contour plot

```
Clear[κ]; κ := 0
Clear[γo]; γo := 0
Clear[ao]; ao := 8
Clear[βo]; βo := 0.122167
Clear[αo]; αo := 0.000625216
Clear[λ]; λ := 0
Clear[Vo]; V := 0.05
Clear[ho]; ho := 0.24
Clear[so]; so := 0.8
Clear[β];
Clear[m];
Clear[mo];
Clear[l];

β[s_] /; (s < so) := βo (so / s - 1)
β[s_] /; (s ≥ so) := 0

a[s_] := ao / (1 - s)

h[s_] := ho (1 - s)

l[x_, η_, s_] /; x ≥ a[s] := (1 - V) Exp[-(αo + λ η) (e^(β[s] x) - e^(β[s] a[s])) / β[s] - (x - a[s]) (κ η + γo)]
l[x_, η_, s_] /; x < a[s] := 1 - V - (x - a[s]) V / a[s]


mo[x_, s_, η_] /; 50 ≥ x ≥ a[s] := h[s] Exp[-(αo + λ η) (e^(4 β[s] x) - e^(4 β[s] a[s])) / 4 β[s]]
mo[x_, s_, η_] /; x < a[s] := 0
mo[x_, s_, η_] /; x > 50 := 0
m[x_, η_, s_] := mo[x, s, η] (1 - 1 / (η + 1))
```

```
Plot[l[x, 10, 0.55], {x, 0, 110}, ImageSize → 350,
  PlotLabel → StyleForm["Denmark", FontSize → 12]]
Plot[m[x, 10, 0.55], {x, 10, 60}, ImageSize → 350,
  PlotLabel → StyleForm["Denmark", FontSize → 12]]
```



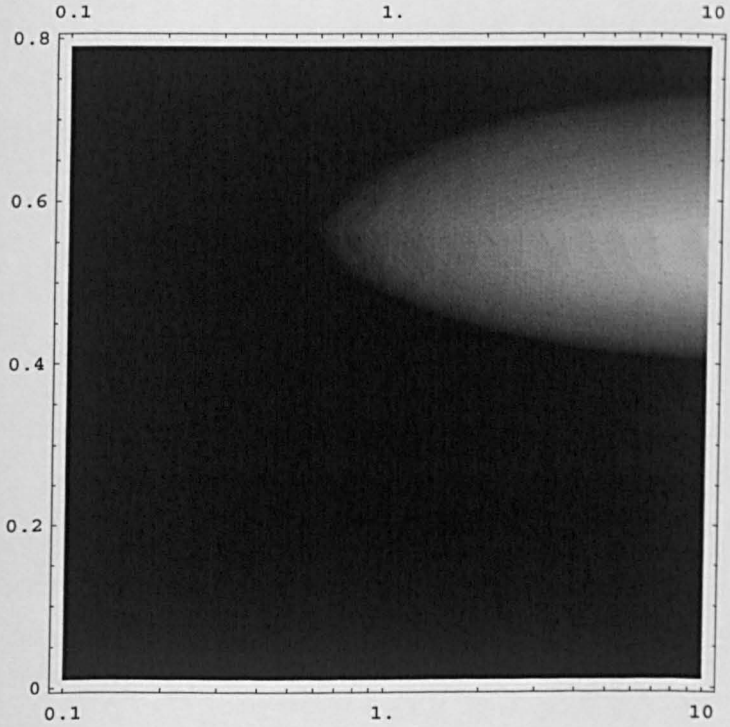
- Graphics -



- Graphics -

$$\mathbf{r[\eta\_, s\_] := r /. FindRoot\left[\int_9^{94} e^{-r\,x}\, l[x, \eta, s]\, m[x, \eta, s]\, dx == 1, \{r, -0.05, 0.05\}\right]}$$

```
tickpos = Join[Table[Log[10, η], {η, 0.1, 1, 0.1}], Table[Log[10, η], {η, 2, 10, 1}]];
ticklabels = Map[If[Chop[Mod[#, 1]] == 0, 10^#, ""] &, tickpos];

cp1 = ContourPlot[r[10^η, s], {η, -1, 1},
    {s, 0.01, 0.79}, Contours → 30, ContourLines → False,
    FrameTicks → {Transpose[{tickpos, ticklabels}], Automatic},
    PlotRange → {Automatic, Automatic, {0, Automatic}}, ImageSize → 350]
```



- ContourGraphics -

```
r[10, 0.542]
```

0.0313587

```
Mean[Cases[Flatten[Table[r[i, z], {i, 0.2, 0.5, 0.1}, {z, 0.35, 0.75, 0.05}]], _?Positive]]
Mean[Cases[Flatten[Join[Table[r[i, z], {i, 0.8, 1, 0.1}, {z, 0.35, 0.75, 0.05}],
      Table[r[i, z], {i, 2, 3, 1}, {z, 0.35, 0.75, 0.05}]]], _?Positive]]
Mean[Cases[Flatten[Table[r[i, z], {i, 5, 10, 1}, {z, 0.35, 0.75, 0.05}]], _?Positive]]
```

Mean[{}]

0.00951127

0.0205863

For Gambia

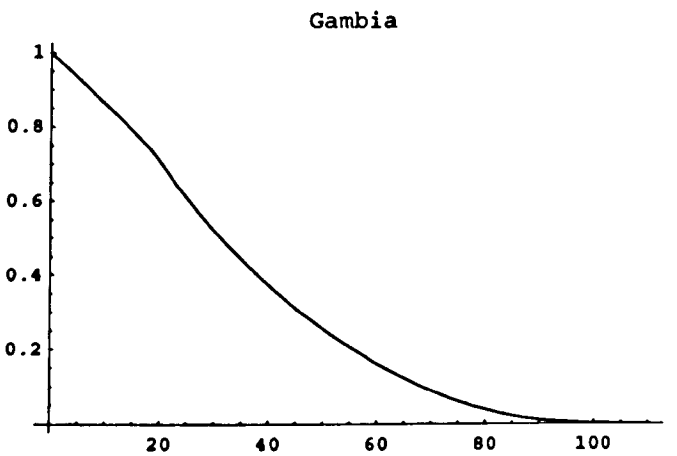
- Turning off messages

```
Off[General::spell]
Off[General::unfl]
Off[General::ovfl]
Off[NIntegrate::ploss]
Off[FindRoot::frsec]
Off[FindRoot::frnum]
Off[NIntegrate::ncvb]
Off[NIntegrate::slwcon]
Off[NIntegrate::inum]
Off[NIntegrate::inovf]
Off[Graphics::gprim]
Off[ReplaceAll::reps]
```

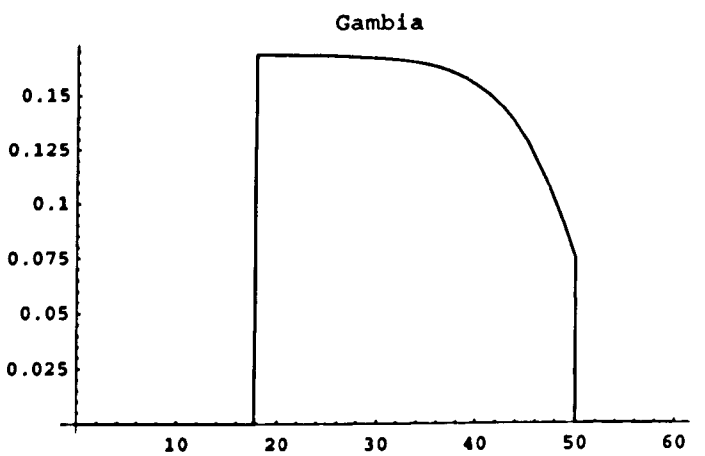
- Equations and contour plot

```
Clear[κ]; κ := 0.0269311
Clear[γo]; γo := 0
Clear[ao]; ao := 8
Clear[βo]; βo := 0.122167
Clear[αo]; αo := 0.000625216
Clear[λ]; λ := 0.000252868
Clear[Vo]; V := 0.25
Clear[ho]; ho := 0.75
Clear[so]; so := 0.8
Clear[β];
Clear[m];
Clear[mo];
Clear[I];

β[s_] /; (s < so) := βo (so/s - 1)
β[s_] /; (s ≥ so) := 0

a[s_] := ao/(1 - s)

h[s_] := ho (1 - s)

I[x_, η_, s_] /; x ≥ a[s] := (1 - V) Exp[-(αo + λ η) (e^(β[s] x) - e^(β[s] a[s]))/β[s] - (x - a[s]) (κ η + γo)]
I[x_, η_, s_] /; x < a[s] := 1 - V - (x - a[s]) V/a[s]


mo[x_, s_, η_] /; 50 ≥ x ≥ a[s] := h[s] Exp[-(αo + λ η) (e^(4 β[s] x) - e^(4 β[s] a[s]))/4 β[s]]
mo[x_, s_, η_] /; x < a[s] := 0
mo[x_, s_, η_] /; x > 50 := 0
m[x_, η_, s_] := mo[x, s, η] (1 - 1/(η + 1))
```

```
Plot[l[x, 1, 0.55], {x, 0, 110}, ImageSize → 350,
  PlotLabel → StyleForm["Gambia", FontSize → 12]]
Plot[m[x, 1, 0.55], {x, 0, 60}, ImageSize → 350,
  PlotLabel → StyleForm["Gambia", FontSize → 12]]
```



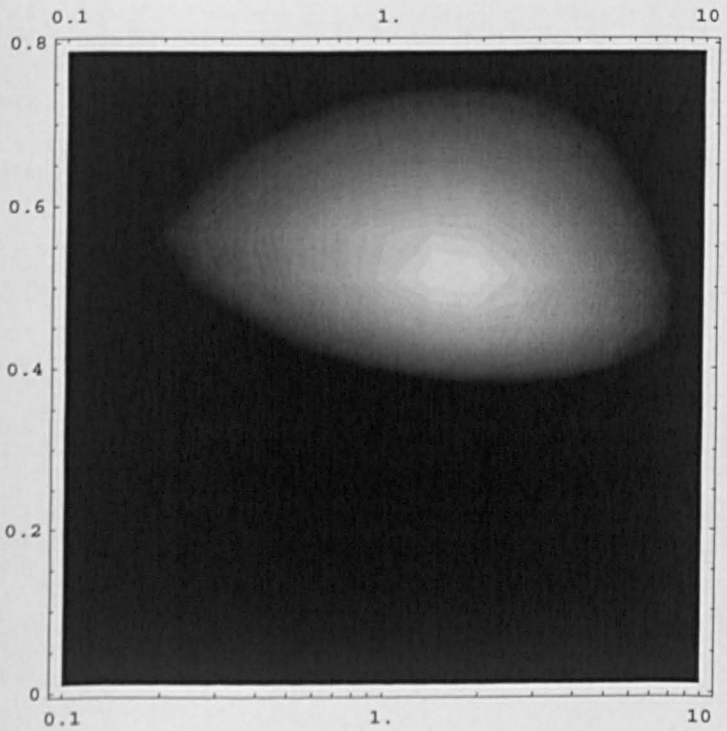
- Graphics -



- Graphics -

$$\mathbf{r[\eta\_,\ s\_] := r\ /.\ FindRoot\Big[\int_9^{94} e^{-r\,x}\ l[x,\ \eta,\ s]\ m[x,\ \eta,\ s]\ dx == 1,\ \{r,\ -0.05,\ 0.05\}\Big]}$$

```
tickpos = Join[Table[Log[10, η], {η, 0.1, 1, 0.1}], Table[Log[10, η], {η, 2, 10, 1}]];
ticklabels = Map[If[Chop[Mod[#, 1]] == 0, 10^#, ""] &, tickpos];

cp2 = ContourPlot[r[10^η, s], {η, -1, 1},
    {s, 0.01, 0.79}, Contours → 30, ContourLines → False,
    FrameTicks → {Transpose[{tickpos, ticklabels}], Automatic},
    PlotRange → {Automatic, Automatic, {0, Automatic}}, ImageSize → 350]
```



- ContourGraphics -

```
r[2, 0.5265]
```

0.0314954

```
Mean[Cases[Flatten[Table[r[i, z], {i, 0.2, 0.5, 0.1}, {z, 0.35, 0.75, 0.05}]], _?Positive]]
Mean[Cases[Flatten[Join[Table[r[i, z], {i, 0.8, 1, 0.1}, {z, 0.35, 0.75, 0.05}],
      Table[r[i, z], {i, 2, 3, 1}, {z, 0.35, 0.75, 0.05}]]], _?Positive]]
Mean[Cases[Flatten[Table[r[i, z], {i, 5, 10, 1}, {z, 0.35, 0.75, 0.05}]], _?Positive]]
```

0.0102003

0.0193559

0.00763002

The last step is to find out the gene frequencies of η in The Gambia and then move them to Denmark to examine the predicted change. To do that we use the relative fitness of each genotype, where we divide the values of r by their highest value so that the best genotype has relative fitness 1. And we make the simplifying assumption that only one gene is responsible for the balance between Th1 and Th2.

For The Gambia

If we call the two alleles of the immune reaction G and g then we can say that G homozygotes have an η of the order of 10 while g homozygotes have η of the order 0.1 and the heterozygotes 1. Then in The Gambia the value of r will be:

GG	Gg	gg	
0.0102003	0.0193559	0.00763002	r
0.5269866	1	0.39419608	relative frequency
$1\text{-}s_1$	1	$1\text{-}s_2$	
s_1 =0.473013		s_2 = 0.605804	coefficient of selection

We have an overdominance case when it comes to selection i.e the fitness of the heterozygotes is higher than either of the homozygotes and thus we expect that a point of balance will exist between the two alleles.

The point of balance between the two allele frequencies can be calculated easily with the use of a standard formula

$$p_G = \frac{s_2}{s_1+s_2}$$

```
Clear[s1]; s1 = 0.473013;
Clear[s2]; s2 = 0.605804;
Clear[pG];
Clear[pg];
```

$$\text{Print}\left["p_G = ", pG = \frac{s2}{s1 + s2}\right]$$

```
Print["p_g = ", pg = 1 - pG]
```

$p_G = 0.561545$

$p_g = 0.438455$

So using the allele frequencies in The Gambia as a starting point, we can investigate the change of these frequencies when a female population is moved to Denmark (or where the infectious load of a population is decreased significantly during a single generation).

The first generation in Denmark will be:

GG	Gg	gg	
0.315333	0.492424	0.192243	1st generation
0	0.009511	0.020586	r
0	0.462013	1	relative frequency
1-s	1-h s		
s=1	h = 0.462013		coefficient of selection

In this case we expect that in the end the g allele will become extinct from the population. To find the number of generations necessary for this and the rate of change of the g allele's frequency in each passing generation we can use again a standard equation for selection under partial dominance.

```
Clear[PartialDominance];
PartialDominance[G0_, g0_, h_, s_] := Module[{x, g, G, i = 1, dQ},

  g = Table[0, {i, 1, 1000}];
  G = Table[0, {i, 1, 1000}];

  g[[1]] = g0;
  G[[1]] = G0;

  While[-(s g[[i]] G[[i]] (G[[i]] + h (g[[i]] - G[[i]]))) / (1 - 2 h s g[[i]] G[[i]] - s G[[i]]^2) ≤ -10^-6,

    G[[i + 1]] = (G[[i]] - h s g[[i]] G[[i]] - s G[[i]]^2) / (1 - 2 h s g[[i]] G[[i]] - s G[[i]]^2);

    g[[i + 1]] = (1 - G[[i + 1]]);

    Print[i, " ", "pG = ", G[[i]], " ", "pg = ", g[[i]], " ", "dQ = ",
      -(s g[[i - 1]] G[[i - 1]] (G[[i - 1]] + h (g[[i - 1]] - G[[i - 1]]))) / (1 - 2 h s g[[i - 1]] G[[i - 1]] - s G[[i - 1]]^2)];

    Print["GG", " ", G[[i]]^2, " ", "Gg", " ", 2 G[[i]] g[[i]], " ", "gg", " ", g[[i]]^2];
    i++]
]
```

```
PartialDominance[0.561545, 0.438455, 0.462013, 1]
```

1 p_G = 0.561545 p_g = 0.438455 dQ = $-\frac{List^3}{1 - 1.92403\, List^2}$

GG 0.315333 Gg 0.492424 gg 0.192243

2 p_G = 0.289743 p_g = 0.710257 dQ = -0.271802

GG 0.0839508 Gg 0.411584 gg 0.504465

3 p_G = 0.15252 p_g = 0.84748 dQ = -0.137222

GG 0.0232625 Gg 0.258516 gg 0.718222

4 p_G = 0.0811141 p_g = 0.918886 dQ = -0.0714063

GG 0.00657949 Gg 0.149069 gg 0.844351

5 p_G = 0.043371 p_g = 0.956629 dQ = -0.037743

GG 0.00188105 Gg 0.08298 gg 0.915139

6 p_G = 0.0232564 p_g = 0.976744 dQ = -0.0201146

GG 0.000540861 Gg 0.0454311 gg 0.954028

7 $p_G = 0.0124896$ $p_g = 0.98751$ $dQ = -0.0107668$

GG 0.00015599 Gg 0.0246672 gg 0.975177

8 $p_G = 0.00671286$ $p_g = 0.993287$ $dQ = -0.00577672$

GG 0.0000450625 Gg 0.0133356 gg 0.986619

9 $p_G = 0.00360959$ $p_g = 0.99639$ $dQ = -0.00310327$

GG 0.0000130292 Gg 0.00719313 gg 0.992794

10 $p_G = 0.00194138$ $p_g = 0.998059$ $dQ = -0.00166821$

GG 3.76896×10^{-6} *Gg 0.00387523 gg 0.996121*

11 $p_G = 0.00104428$ $p_g = 0.998956$ $dQ = -0.000897098$

GG 1.09053×10^{-6} *Gg 0.00208639 gg 0.997913*

12 $p_G = 0.000561767$ $p_g = 0.999438$ $dQ = -0.000482517$

GG 3.15582×10^{-7} *Gg 0.0011229 gg 0.998877*

13 $p_G = 0.00030221$ $p_g = 0.999698$ $dQ = -0.000259556$

GG 9.1331×10^{-8} *Gg 0.000604238 gg 0.999396*

14 $p_G = 0.000162581$ $p_g = 0.999837$ $dQ = -0.000139629$

GG 2.64327×10^{-8} *Gg 0.00032511 gg 0.999675*

15 $p_G = 0.0000874656$ $p_g = 0.999913$ $dQ = -0.0000751158$

GG 7.65024×10^{-9} *Gg 0.000174916 gg 0.999825*

16 $p_G = 0.0000470551$ $p_g = 0.999953$ $dQ = -0.0000404106$

GG 2.21418×10^{-9} *Gg 0.0000941057 gg 0.999906*

17 $p_G = 0.0000253149$ $p_g = 0.999975$ $dQ = -0.0000217401$

GG 6.40845×10^{-10} *Gg 0.0000506286 gg 0.999949*

18 $p_G = 0.0000136191$ $p_g = 0.999986$ $dQ = -0.0000116958$

GG 1.85479×10^{-10} *Gg 0.0000272378 gg 0.999973*

19 $p_G = 7.32688 \times 10^{-6}$ $p_g = 0.999993$ $dQ = -6.2922 \times 10^{-6}$

GG 5.36831×10^{-11} *Gg 0.0000146536 gg 0.999985*

20 $p_G = 3.94176 \times 10^{-6}$ $p_g = 0.999996$ $dQ = -3.38511 \times 10^{-6}$

GG 1.55375×10^{-11} *Gg* 7.88349×10^{-6} *gg 0.999992*

So we can see that initially the rate of change is very big but it tends to get smaller and smaller. We can say the allele is virtually absent from the population when its frequency falls below 1 in 1000000.

Appendix 3. *Mathematica* notebook for the calculation of the allelic effect

To calculate the allelic effect of ϵ2 and ϵ4 relative to the ϵ3ϵ3 genotype we imagine a population of individuals all carrying the ϵ3ϵ3 genotype. The allele effect will be the difference in the mean lifespan if we change the genotype of the population from ϵ3ϵ3 to ϵ3ϵx where ϵx can be either ϵ2 or ϵ4.

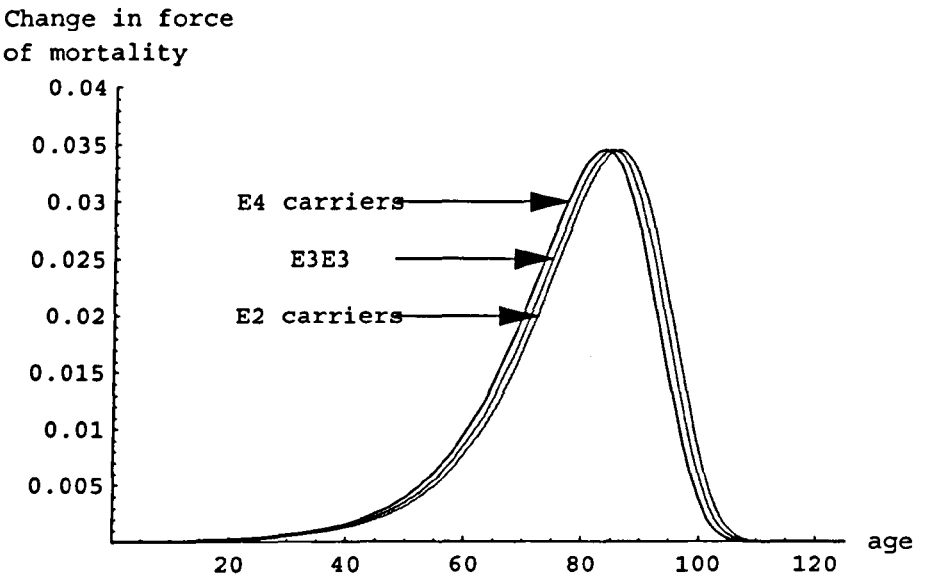
Plotting the Gompertz equation in a form representing the number of deaths at each age and applying the relative risk factors for the ϵ2 and ϵ3 carriers we can calculate the change of the mean lifespan in each case.

ϵ3o, ϵ2o and ϵ4o specify the number of carriers of each allele in a cohort of 100,000

```
<< Statistics'ContinuousDistributions'
<< Statistics'NormalDistribution'
<< Statistics'DescriptiveStatistics'
<< Graphics'Arrow'
<< Graphics'Colors'
```

```
ϵ2o := 13000
ϵ3o := 58000
ϵ4o := 29000

α := 3.51364 10^-5
β := 0.093832

μ3[x_] := α e^(β x)
μ2[x_] := 0.9 α e^(β x)
μ4[x_] := 1.13 α e^(β x)

meanLifeSpanϵ3[x_] = Sum_{x=0}^{122} x (e^(-∫_1^x μ3[y-1] dy) - e^(-∫_0^x μ3[y] dy));

meanLifeSpanϵ2[x_] = Sum_{x=0}^{122} x (e^(-∫_1^x μ2[y-1] dy) - e^(-∫_0^x μ2[y] dy));

meanLifeSpanϵ4[x_] = Sum_{x=0}^{122} x (e^(-∫_1^x μ4[y-1] dy) - e^(-∫_0^x μ4[y] dy));

Gompertzϵ3[x_] = e^(-∫_1^x μ3[y-1] dy) - e^(-∫_0^x μ3[y] dy);
Gompertzϵ2[x_] = e^(-∫_1^x μ2[y-1] dy) - e^(-∫_0^x μ2[y] dy);
Gompertzϵ4[x_] = e^(-∫_1^x μ4[y-1] dy) - e^(-∫_0^x μ4[y] dy);

Print["Gompertz Curve"]
Plot[{Gompertzϵ3[x], Gompertzϵ2[x], Gompertzϵ4[x]}, {x, 0, 125},
  AxesLabel → {"age", "Change in force\nof mortality"},
  PlotRange -> {{0, 125}, {0, 0.04}},
  AxesOrigin → {0, 0}, PlotStyle → {Red, BlueViolet, Blue},
  Epilog → {Arrow[{48, 0.03}, {77, 0.03}], Text["E4 carriers", {35, .03}],
      Arrow[{48, 0.025}, {74.5, 0.025}], Text["E3E3", {35, .025}],
      Arrow[{48, 0.02}, {72, 0.02}], Text["E2 carriers", {35, .02}]}]

Print["Mean Life Span for ϵ3 ->", meanLifeSpanϵ3[x]]
Print["Mean Life Span for ϵ2 carriers ->", meanLifeSpanϵ2[x]]
Print["Mean Life Span for ϵ4 carriers ->", meanLifeSpanϵ4[x]]

Print["Change in Mean Life Span between ϵ3ϵ3 and ϵ2 carriers-> ",
  meanLifeSpanϵ2[x] - meanLifeSpanϵ3[x]]
Print["Change in Mean Life Span between ϵ3ϵ3 and ϵ4 carriers-> ",
  meanLifeSpanϵ4[x] - meanLifeSpanϵ3[x]]
```

Gompertz Curve



- Graphics -

Mean Life Span for ϵ3 ->78.4683

Mean Life Span for ϵ2 carriers ->79.5882

Mean Life Span for ϵ4 carriers ->77.1695

Change in Mean Life Span between ϵ3ϵ3 and ϵ2 carriers-> 1.11992

Change in Mean Life Span between ϵ3ϵ3 and ϵ4 carriers-> -1.29875

Assuming that the alleles have an additive effect on lifespan we can calculate the difference caused by each genotype relative to ϵ3ϵ3 individuals.

$$\alpha\epsilon 2 := \alpha 2 \,/.\, \text{Solve}\left[79.5882 - 78.4683 == \frac{11.5\,\alpha 2}{13} + \frac{2\,1.5\,\alpha 2}{13}, \alpha 2\right][[1]]$$

$$\alpha\epsilon 4 := \alpha 4 \,/.\, \text{Solve}\left[77.1695 - 78.4683 == \frac{25\,\alpha 4}{28} + \frac{2\,4\,\alpha 4}{28}, \alpha 4\right][[1]]$$

```
Print["The ϵ2 allele effect is ", αϵ2]
Print["The ϵ4 allele effect is ", αϵ4]
```

The ϵ2 allele effect is 1.00405

The ϵ4 allele effect is -1.10201

So now based again on the assumption of the apo E additive effects we can easily calculate the relative change expected by each genotype relative to ϵ3ϵ3 and ϵ2ϵ4 genotypes:

```
Print["For ϵ3ϵ3 and ϵ2ϵ4 ->", 0]
Print["For ϵ2ϵ3 ->", αϵ2]
Print["For ϵ2ϵ2 ->", 2 αϵ2]
Print["For ϵ4ϵ3 ->", αϵ4]
Print["For ϵ4ϵ4 ->", 2 αϵ4]
```

For ϵ3ϵ3 and ϵ2ϵ4 ->0

For ϵ2ϵ3 ->1.00405

For ϵ2ϵ2 ->2.0081

For ϵ4ϵ3 ->-1.10201

For ϵ4ϵ4 ->-2.20402

From this values we have to go back and calculate the relative risk factor associated with each genotype.

```
RiskFactor[GT_] :=
  Module[{α = 3.51364 10^-5, β = 0.093832, μR, μ3, RF},
         μR[x_] := R α e^(β x);
         μ3[x_] := α e^(β x);
    RF = R /. FindRoot[Sum_{x=0}^{122} x (e^(-∫_1^x μ3[y-1] dy) - e^(-∫_0^x μ3[y] dy)) + GT ==
         Sum_{x=0}^{122} x (e^(-∫_1^x μR[y-1] dy) - e^(-∫_0^x μR[y] dy)), {R, 1}]]

Print["For ϵ3ϵ3 and ϵ2ϵ4 ->", 1]
Print["For ϵ2ϵ3 ->", RiskFactor[αϵ2]]
Print["For ϵ2ϵ2 ->", RiskFactor[2 αϵ2]]
Print["For ϵ4ϵ3 ->", RiskFactor[αϵ4]]
Print["For ϵ4ϵ4 ->", RiskFactor[2 αϵ4]]
```

For ϵ3ϵ3 and ϵ2ϵ4 ->1

For ϵ2ϵ3 ->0.909864

For ϵ2ϵ2 ->0.827868

For ϵ4ϵ3 ->1.10927

For ϵ4ϵ4 ->1.23051

Appendix 4. *Mathematica* notebook for the calculation of the relative risk from the odds ratio

Many health studies use logistic regression to analyse the data obtained. However the resulting odds ratio may overestimate the true effect of the risk factor of interest (Robbins et al., 2002). The problem is that logistic regression and the odds ratio are accurate for rare events (<10%) but are often used to study common outcomes (McNutt et al., 2003). Since we are interested in the relative risk of each lifestyle parameter, we will the methodology of Zhang and Yu (1998) to approximate the relative risk from the adjusted odds ratio which is commonly used in CVD studies.

For example, to calculate the relative risk associated with gender we are going to use the odds ratio provided by Stevens et al (2001) and the incidence described by Panagiotakos et al (2002). Using the formula:

$$RR = \frac{OR}{(1 - P_O) + (P_O \times OR)}$$

where RR is the relative risk, P_O is the incidence of the outcome in the non-exposed population and OR is the odds ratio. Substituting the incidence with 20% and the OR with 0.525 we can calculate the relative risk as 0.58. Using the Gompertz model for the number of deaths per year of age we can see that the difference produced between males and females is almost 6 years which is within the limits of the expected lag of cardiovascular disease between men and women.

We are going to use the same method to estimate the relative risk associated with each dietary pattern group of the Framingham study based on the odds ratio provided by Millen et al (2002). So the Heart Healthy group is used as the reference with risk ratio 1.0, the Light Eating risk is corrected to 1.17, the Wine and Moderate is changed to 1.32 which is very close to the original odds ratio mainly due to the small representation of the group in the population, the relative risk for a High Fat diet is 1.36 and finally the Empty Calorie diet risk factor was estimated to 2.07.

To check the validity of the CVD risk factor associated with gender we are going to use the set of equations described previously to calculate the difference of age at peak mortality and compare it to the observed 5 - 15 years lag seen between men and women

```
<< Statistics`ContinuousDistributions`
<< Statistics`NormalDistribution`
<< Statistics`DescriptiveStatistics`
<< Graphics`Arrow`
```

$$\alpha := 3.51364\,10^{-5}$$
$$\beta := 0.093832$$

$$\mu\text{male}[x\_] := \alpha\, e^{\beta x}$$
$$\mu\text{female}[x\_] := 0.58\,\alpha\, e^{\beta x}$$

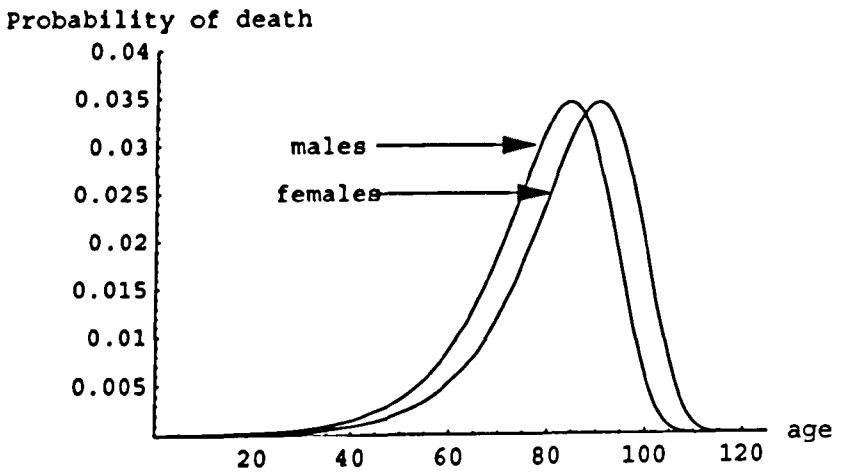
$$\text{meanLifeSpanmale}[x\_] = \sum_{x=0}^{122} x\left(e^{-\int_1^{\mu\text{male}[y-1]\,dy}} - e^{-\int_0^{\mu\text{male}[y]\,dy}}\right);$$

$$\text{meanLifeSpanfemale}[x\_] = \sum_{x=0}^{122} x\left(e^{-\int_1^{\mu\text{female}[y-1]\,dy}} - e^{-\int_0^{\mu\text{female}[y]\,dy}}\right);$$

$$\text{Gompertzmale}[x\_] = e^{-\int_1^{\mu\text{male}[y-1]\,dy}} - e^{-\int_0^{\mu\text{male}[y]\,dy}};$$
$$\text{Gompertzfemale}[x\_] = e^{-\int_1^{\mu\text{female}[y-1]\,dy}} - e^{-\int_0^{\mu\text{female}[y]\,dy}};$$

```
Print["Gompertz Curve"]
Plot[{Gompertzmale[x], Gompertzfemale[x]}, {x, 0, 125},
  AxesLabel → {"age", "Probability of death"}, PlotRange -> {{0, 125}, {0, 0.04}},
  AxesOrigin → {0, 0}, Epilog → {Arrow[{45, 0.03}, {77, 0.03}], Text["males", {35, .03}],
     Arrow[{45, 0.025}, {80, 0.025}], Text["females", {35, .025}]}]

Print["Mean Life Span for males ->", meanLifeSpanmale[x]]
Print["Mean Life Span for females ->", meanLifeSpanfemale[x]]

Print["Change in Mean Life Span between males and females -> ",
  meanLifeSpanfemale[x] - meanLifeSpanmale[x]]
```

Gompertz Curve



- Graphics -

Mean Life Span for males ->78.4683

```
Mean Life Span for females ->84.261

Change in Mean Life Span between males and females -> 5.79267
```

Using the relative risk of 0.58 we observe a lag of 5.7 years at the peak mortality age.

Appendix 5. *Mathematica* notebook for the calculation of the genotype environment interaction

Due to lack of data on the quantitative link between apo E and the environmental risk parameters, we are going to calculate theoretically these associations based on current knowledge of their characteristics.

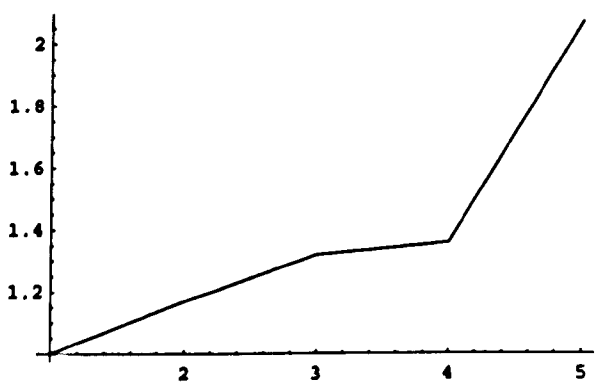
■ apo E × Diet

In our discussion of the interaction between apo E and diet we assumed that a certain diet will exist such that no distinction between the different genotypes can be seen, and that the changes in diet will have a higher impact on $\epsilon 4$ carriers relative to those having the $\epsilon 3$ and $\epsilon 2$ alleles, in line with published data. Furthermore we can fit a number of linear and non-linear model curves through the 5 categories given by Millen et al (2002) to consider the distinct dietary patterns as a continuum.
Based on the above we can find six curves that describe the relationship between diet and apo E genotype. We will assume that the $\epsilon 3\epsilon 3$ genotype curve corresponds to the risk factors measured in the population by Millen et al (2002). This is a good approximation and will simplify the problem significantly since apo $\epsilon 3\epsilon 3$ is the most numerous group (more than 60%), and in terms of risk factors, it lies approximately in the middle between $\epsilon 2$ and $\epsilon 4$ carriers' risk.

The first step is to fit a curve through the data points for the dietary patterns. We will use a hybrid set of data between the age and multivariate adjusted odds ratio. The reason is that although we require the risk to be adjusted for other confounding factors, when this is applied at the level of HDL cholesterol for a high fat diet a big part of the association with the apo E is lost. So we will use the multivariate-adjusted risk factors except in the case of the high fat diet where we will use the age-adjusted one.

If we plot the five dietary patterns in a diet risk graph we can see that the data can easily be considered as being continuous

```
datapoints = ListPlot[{1.0, 1.17, 1.32, 1.36, 2.07}, PlotJoined → True]
```



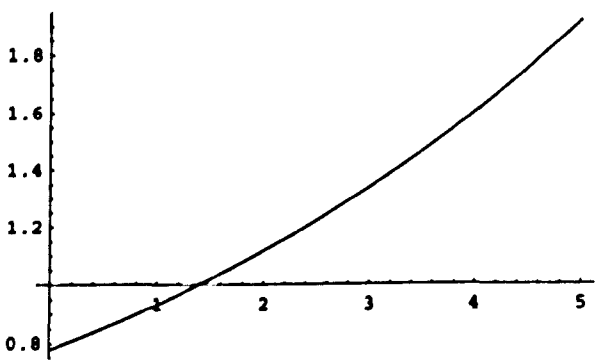

```
- Graphics -
```

We will use two different models to fit a curve through the points. A non-linear exponential with the equation $y(x) = \alpha e^{\beta x}$ and a linear quadratic described by $y(x) = \alpha + \beta x^2$.

```
<< Statistics`NonlinearFit`

NonlinearRegress[{1.0, 1.17, 1.32, 1.36, 2.07},
 alpha Exp[beta x], {x}, {{alpha, 0.8}, {beta, 0.05}}]
```

```
{BestFitParameters → {alpha → 0.777504, beta → 0.180799}, ParameterCITable →
            Estimate     Asymptotic SE      CI
  alpha     0.777504     0.126102           {0.37619, 1.17882}    ,
  beta      0.180799     0.0414913          {0.0487549, 0.312843}
 EstimatedVariance → 0.0297182,
                                  DF     SumOfSq        MeanSq
              Model               2      10.1566        5.07832
 ANOVATable → Error               3      0.0891546      0.0297182,
              Uncorrected Total   5      10.2458
              Corrected Total     4      0.66852
 AsymptoticCorrelationMatrix → ( 1.        -0.942745 ),
                               ( -0.942745  1.       )
                                                      Curvature
 FitCurvatureTable → Max Intrinsic                    0.0698095}
                     Max Parameter-Effects            0.618342
                     95. % Confidence Region          0.323557
```

```
fitcurve = Plot[0.7775 e^(0.1808 x), {x, 0, 5}]
```

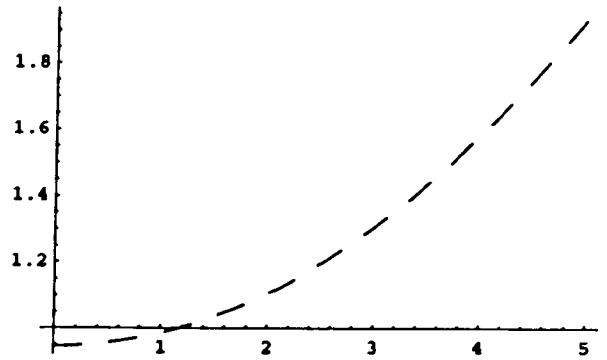



```
- Graphics -
```

```
<< Statistics`LinearRegression`

Regress[{1.0, 1.17, 1.32, 1.36, 2.07}, {1, x^2}, x]
```

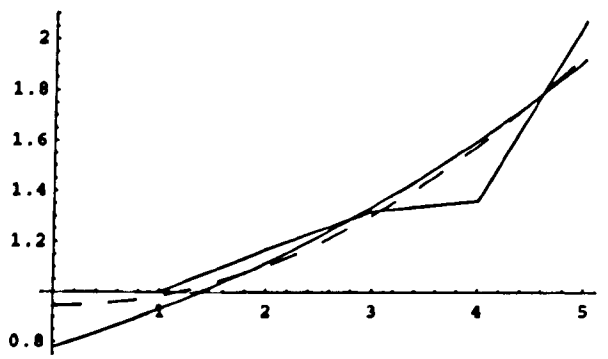
```
                        Estimate     SE           TStat      PValue
{ParameterTable → 1     0.944294     0.111248     8.48817    0.00343356,
                  x^2   0.0399733    0.00795037   5.02785    0.0151597
 RSquared → 0.893915, AdjustedRSquared → 0.858554,
 EstimatedVariance → 0.0236399, ANOVATable →
          DF     SumOfSq       MeanSq        FRatio     PValue
 Model    1      0.5976        0.5976        25.2793    0.0151597}
 Error    3      0.0709197     0.0236399
 Total    4      0.66852
```

```
fitlinear = Plot[0.9442 + 0.04 x^2, {x, 0, 5}, PlotStyle → {{Dashing[{0.05, 0.05}]}}]
```




```
- Graphics -
```

```
Show[datapoints, fitcurve, fitlinear]
data = {1.0, 1.17, 1.32, 1.36, 2.07}
datanonlinear = Table[0.7775 e^(0.1808 x), {x, 1, 5}]
datalinear = Table[0.9442 + 0.04 x^2, {x, 1, 5}]
```




```
- Graphics -
```

```
{1., 1.17, 1.32, 1.36, 2.07}
```

```
{0.931581, 1.1162, 1.3374, 1.60244, 1.92001}
```

```
{0.9842, 1.1042, 1.3042, 1.5842, 1.9442}
```

We can see that both models give lines of good fit with the data. Although the differences between the fitted lines are small we will use the linear model as having a marginally better fit and being as the simpler of the two.

A point we have yet not explained is what x=0 means in this model. As discussed earlier, the differences between the apo E genotypes are caused by the action of the environment. So the point of x=0 is a diet that produces the exact same CVD risk factor for all six genotypes.

To account for the differences between the genotypes we will consider that the relative risks between them are the relative difference between their curves. Thus having set the $\epsilon 3\epsilon 3$ curve equal to the population mean we could easily find the β parameter for all six genotypes. To do that we will divide the product of each curve with the one from the reference curve and make it equal to the relative risk between the corresponding genotype and the $\epsilon 3\epsilon 3$ one.

For the $\epsilon 2\epsilon 2$ genotype

```
Clear[β22];
Solve[∏x=1^5 (0.9442 + β22 x^2) / ∏x=1^5 (0.9442 + 0.04 x^2) == 0.8279, β22]
β22 = Select[Table[β22 /. %], Positive]
```

```
{{β22 → -0.943672}, {β22 → -0.277625}, {β22 → -0.0976552 - 0.111175 i},
 {β22 → -0.0976552 + 0.111175 i}, {β22 → 0.0346665}}

{0.0346665}
```

We require only the real and positive solutions so in this case the only acceptable solution is for $\beta 22 = 0.0345$.

Similar for the $\epsilon 2\epsilon 3$ genotype

```
Clear[β23];
Solve[∏x=1^5 (0.9442 + β23 x^2) / ∏x=1^5 (0.9442 + 0.04 x^2) == 0.9099, β23]
β23 = Select[Table[β23 /. %], Positive]
```

```
{{β23 → -0.94362}, {β23 → -0.280229}, {β23 → -0.0976954 - 0.114627 i},
 {β23 → -0.0976954 + 0.114627 i}, {β23 → 0.0372985}}

{0.0372985}
```

The $\epsilon 2\epsilon 4$ genotypes and the $\epsilon 3\epsilon 3$ genotypes were considered equal so they also share the same slope

```
β = 0.04;
```

For the $\epsilon 3\epsilon 4$ genotype

```
Clear[β34];
Solve[∏x=1^5 (0.9442 + β34 x^2) / ∏x=1^5 (0.9442 + 0.04 x^2) == 1.1093, β34]
β34 = Select[Table[β34 /. %], Positive]
```

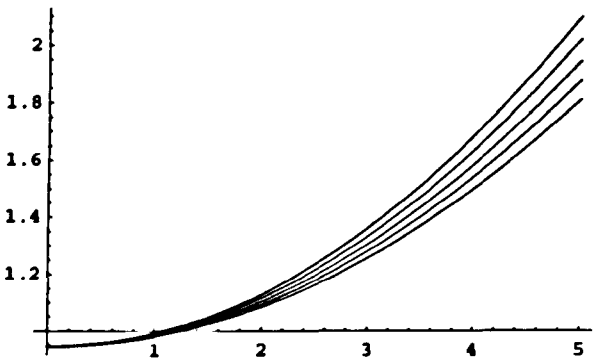
```
{{β34 → -0.943492}, {β34 → -0.286076}, {β34 → -0.0977118 - 0.122101 i},
 {β34 → -0.0977118 + 0.122101 i}, {β34 → 0.0430503}}

{0.0430503}
```

And finally for the $\epsilon 4\epsilon 4$ genotype

```
Clear[β44];
Solve[∏x=1^5 (0.9442 + β44 x^2) / ∏x=1^5 (0.9442 + 0.04 x^2) == 1.2305, β44]
β44 = Select[Table[β44 /. %], Positive]
```

```
{{β44 → -0.943415}, {β44 → -0.289348}, {β44 → -0.0976829 - 0.12614 i},
 {β44 → -0.0976829 + 0.12614 i}, {β44 → 0.0461875}}

{0.0461875}
```

```
allgenot = Plot[{0.9442 + β22 x^2, 0.9442 + β23 x^2,
   0.9442 + β x^2, 0.9442 + β34 x^2, 0.9442 + β44 x^2}, {x, 0, 5}]
TableForm[Table[{0.9442 + β22 x^2, 0.9442 + β23 x^2, 0.9442 + β x^2,
   0.9442 + β34 x^2, 0.9442 + β44 x^2}, {x, 0, 5}],
 TableHeadings -> {None, {"ε2ε2", "ε2ε3", "ε3ε3/ε2ε4", "ε3ε4", "ε4ε4"}}]
```



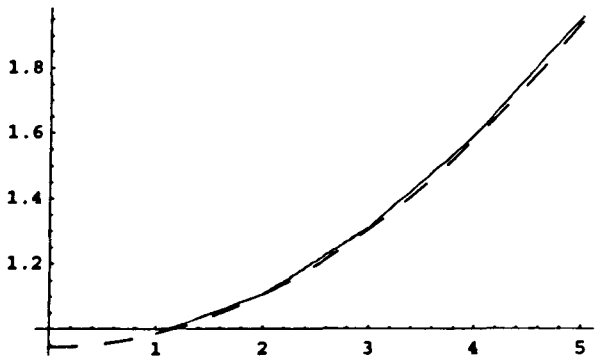

```
- Graphics -
```

ε2ε2	*ε2ε3*	*ε3ε3/ε2ε4*	*ε3ε4*	*ε4ε4*
0.9442	*0.9442*	*0.9442*	*0.9442*	*0.9442*
0.978866	*0.981499*	*0.9842*	*0.98725*	*0.990388*
1.08287	*1.09339*	*1.1042*	*1.1164*	*1.12895*
1.2562	*1.27989*	*1.3042*	*1.33165*	*1.35989*
1.49886	*1.54098*	*1.5842*	*1.633*	*1.6832*
1.81086	*1.87666*	*1.9442*	*2.02046*	*2.09889*

To see the impact of our initial assumption that the ε3ε3 curve was equal to the population mean we will calculate the overall weighted mean for the six curves and compare it to the one provided by the original data.

```
freq22 := 0.017;
freq23 := 0.116;
freq33 := 0.558;
freq24 := 0.019;
freq34 := 0.251;
freq44 := 0.039;

meancurve =
 Flatten[Table[freq22 (0.9442 + β22 x^2) + freq23 (0.9442 + β23 x^2) + freq33 (0.9442 + β x^2) +
    freq24 (0.9442 + β x^2) + freq34 (0.9442 + β34 x^2) + freq44 (0.9442 + β44 x^2), {x, 1, 5}]]

{0.984803, 1.10661, 1.30963, 1.59385, 1.95927}
```

```
Show[fitlinear, ListPlot[meancurve, PlotJoined → True, PlotStyle → Hue[.8]]]
```




```
- Graphics -
```

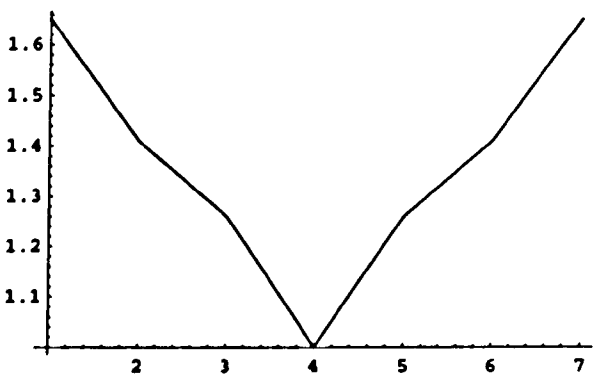
So we can see that the initial assumption has a very small impact on the values of the risk factors between the two curves.

■ apo E × Alcohol

As we mentioned earlier a number of studies have found a U- or J- shape association between alcohol consumption and the risk of CVD. In order to link the action of alcohol with the effect of the apo E genotype we will use a set of U shapes with parameters that correspond to the relative risk of each genotype. Again the first step will be to fit a curve through the data provided. Although Stampfer et al (2000) considered a J- curve we can extend that to describe a U-curve based on the fact that at the point of high consumption the risk factor for CVD increases and on the assumption that this point is a point of symmetry for the curve. In this case it is easy to describe the alcohol risk factor as points of a quadratic curve of the form $\alpha(x-\varphi)^2+\gamma$.

If we plot the points of the risk factors setting the final group as the point of symmetry then we can have an indication of the expected U-shaped curve.

```
datacurve = ListPlot[{1.65, 1.41, 1.26, 1.0, 1.26, 1.41, 1.65}, PlotJoined → True]
```

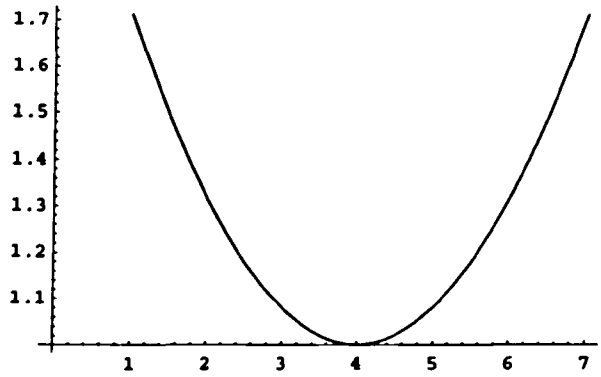



```
- Graphics -
```

To fit the quadratic curve then

```
<< Statistics`NonlinearFit`

Clear[alpha];
Clear[gamma];
NonlinearRegress[{1.65, 1.41, 1.26, 1.0, 1.26, 1.41, 1.65},
 alpha (x - 4)^2 + 1, {x}, {alpha}]

{BestFitParameters → {alpha → 0.0790816}, ParameterCITable →
            Estimate       Asymptotic SE      CI
  alpha     0.0790816      0.00877897         {0.0576003, 0.100563}'
 EstimatedVariance → 0.0151058,

                                  DF      SumOfSq        MeanSq
              Model               1       13.5058        13.5058
ANOVATable → Error                6       0.0906347      0.0151058,
              Uncorrected Total   7       13.5964
              Corrected Total     6       0.320743

AsymptoticCorrelationMatrix → {{1.}},

                                                  Curvature
                          Max Intrinsic           0
FitCurvatureTable →       Max Parameter-Effects   0          }
                          95. % Confidence Region 0.408678
```

```
fitcurve = Plot[0.0791 (x - 4)^2 + 1.0, {x, 1, 7}]
Table[0.0791 (x - 4)^2 + 1.0, {x, 0, 7}]
```

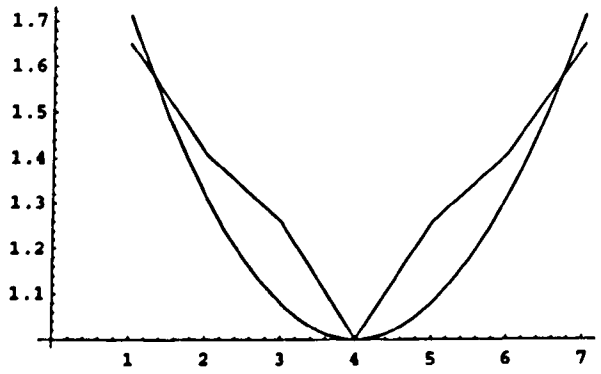



```
- Graphics -
```

```
{2.2656, 1.7119, 1.3164, 1.0791, 1., 1.0791, 1.3164, 1.7119}
```

To compare them we can just fit them in the same graph. We see that the fitted curve underestimates the risk of group 2 and 3 but their values still remain within the confidence intervals predicted by Stampfer et al (2000).

```
Show[datacurve, fitcurve]
```




```
- Graphics -
```

As in the case of diet we will consider that the fitted curve corresponds to the $\epsilon 3\epsilon 3$ (and $\epsilon 2\epsilon 4$) genotype and that all genotypes reach 1.0 as the minimum risk factor. Using the $\epsilon 3\epsilon 3$ curve and knowing the relative risk between the apo E alleles we can calculate the α of the curves corresponding to the rest of the genotypes.

```
Clear[α22];
Solve[ (∏_{x=1}^{7} (α22 (x - 4)^2 + 1.0)) / (∏_{x=1}^{7} (0.0791 (x - 4)^2 + 1.0)) == 0.8279, α22]
α22 = Select[Table[α22 /. %], Positive]
```

```
{{α22 → -1.07695}, {α22 → -0.868977}, {α22 → -0.561207},
 {α22 → -0.142079 - 0.250355 i}, {α22 → -0.142079 + 0.250355 i}, {α22 → 0.0690731}}
```

```
{0.0690731}
```

```
Clear[α23];
Solve[∏_{x=1}^{7} (α23 (x - 4)^2 + 1.0) / ∏_{x=1}^{7} (0.0791 (x - 4)^2 + 1.0) == 0.9099, α23]
α23 = Select[Table[α23 /. %], Positive]
```

```
{{α23 → -1.08011}, {α23 → -0.858185}, {α23 → -0.576961},
 {α23 → -0.140502 - 0.256189 i}, {α23 → -0.140502 + 0.256189 i}, {α23 → 0.0740343}}
```

```
{0.0740343}
```

```
Clear[α34];
Solve[∏_{x=1}^{7} (α34 (x - 4)^2 + 1.0) / ∏_{x=1}^{7} (0.0791 (x - 4)^2 + 1.0) == 1.1093, α34]
α34 = Select[Table[α34 /. %], Positive]
```

```
{{α34 → -1.08708}, {α34 → -0.828646}, {α34 → -0.617254},
 {α34 → -0.137013 - 0.26875 i}, {α34 → -0.137013 + 0.26875 i}, {α34 → 0.0847887}}
```

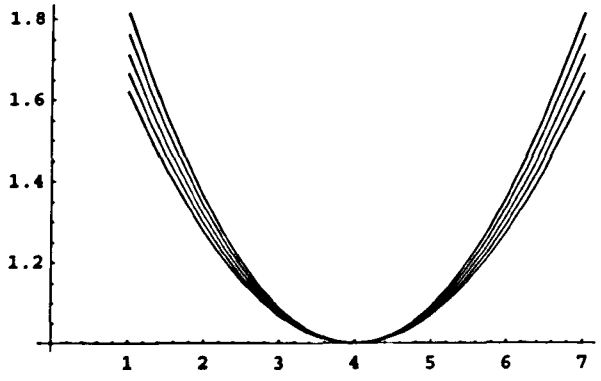
```
{0.0847887}
```

```
Clear[α23];
Solve[∏_{x=1}^{7} (α44 (x - 4)^2 + 1.0) / ∏_{x=1}^{7} (0.0791 (x - 4)^2 + 1.0) == 1.2305, α44]
α44 = Select[Table[α44 /. %], Positive]
```

```
{{α44 → -1.09094}, {α44 → -0.806151}, {α44 → -0.645567},
 {α44 → -0.135086 - 0.275499 i}, {α44 → -0.135086 + 0.275499 i}, {α44 → 0.0906069}}
```

```
{0.0906069}
```

```
allgenot = Plot[{α22 (x - 4)^2 + 1.0, α23 (x - 4)^2 + 1.0,
   0.0791 (x - 4)^2 + 1.0, α34 (x - 4)^2 + 1.0, α44 (x - 4)^2 + 1.0}, {x, 1, 7}]
TableForm[Table[{α22 (x - 4)^2 + 1.0, α23 (x - 4)^2 + 1.0, 0.0791 (x - 4)^2 + 1.0,
   α34 (x - 4)^2 + 1.0, α44 (x - 4)^2 + 1.0}, {x, 1, 7}],
 TableHeadings -> {None, {"ε2ε2", "ε2ε3", "ε3ε3/ε2ε4", "ε3ε4", "ε4ε4"}}]
```



- Graphics -

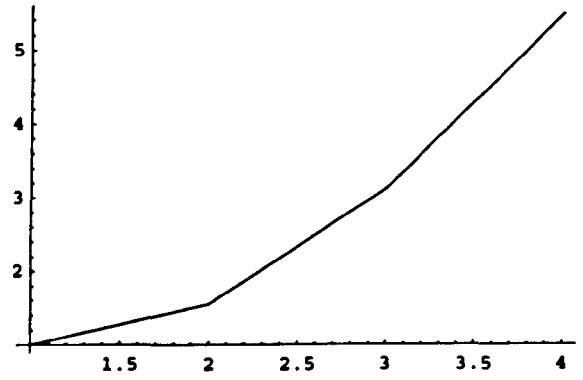
ε2ε2	ε2ε3	ε3ε3/ε2ε4	ε3ε4	ε4ε4
1.62166	1.66631	1.7119	1.7631	1.81546
1.27629	1.29614	1.3164	1.33915	1.36243
1.06907	1.07403	1.0791	1.08479	1.09061
1.	1.	1.	1.	1.
1.06907	1.07403	1.0791	1.08479	1.09061
1.27629	1.29614	1.3164	1.33915	1.36243
1.62166	1.66631	1.7119	1.7631	1.81546

The frequencies that will be used for each category are based on the UK National Statistics on alcohol consumption per week for both sexes.

■ apo E × Smoking

In the case of smoking where the precise form of the interaction with apo E is still largely unknown we will follow the same methodology as in the case of diet. Moreover we will use the same model to fit the data provided by Stampfer et al (2000).

```
datacurve = ListPlot[{1.0, 1.55, 3.12, 5.48}, PlotJoined → True]
```



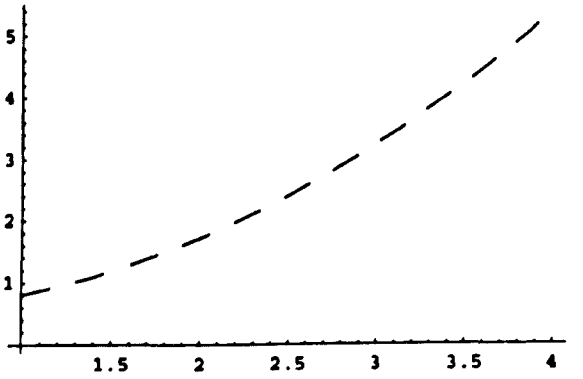

```
- Graphics -
```

```
<< Statistics`LinearRegression`
```

```
Regress[{1.0, 1.55, 3.12, 5.48}, {1, x^2}, x]
```

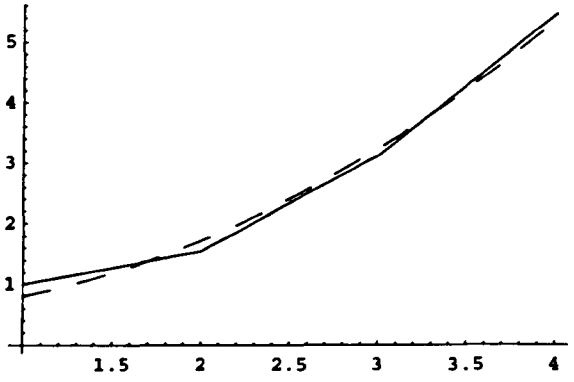
```
                          Estimate      SE            TStat        PValue
{ParameterTable → 1       0.500581      0.178176      2.80948      0.106781  ,
                  x^2     0.304922      0.0189399     16.0995      0.00383592

 RSquared → 0.992343, AdjustedRSquared → 0.988514,
 EstimatedVariance → 0.0462746, ANOVATable →

            DF      SumOfSq        MeanSq          FRatio       PValue
  Model     1       11.9941        11.9941         259.195      0.00383592}
  Error     2       0.0925492      0.0462746
  Total     3       12.0867
```

```
fitlinear = Plot[0.5006 + 0.3049 x^2, {x, 1, 4}, PlotStyle → {{Dashing[{0.05, 0.05}]}}]
```




```
- Graphics -
```

```
Show[datacurve, fitlinear]
data = {1.0, 1.55, 3.12, 5.48}
datalinear = Table[0.5006 + 0.3049 x^2, {x, 1, 4}]
```




```
- Graphics -

{1., 1.55, 3.12, 5.48}

{0.8055, 1.7202, 3.2447, 5.379}
```

We can see again that the points of the fitted curve are within the 95% confidence interval provided by Stampfer et al (2000).

Again we will use the same methodology used earlier in the association of apo E with diet to calculate the interaction of genotype with smoking.

For the $\epsilon 2\epsilon 2$ genotype

```
Clear[β22];
Solve[∏x=1^4 (0.5006 + β22 x^2) / ∏x=1^4 (0.5006 + 0.3049 x^2) == 0.8279, β22]
β22 = Select[Table[β22 /. %], Positive]

{{β22 → -0.666267}, {β22 → -0.165764 - 0.393595 i},
 {β22 → -0.165764 + 0.393595 i}, {β22 → 0.285135}}

{0.285135}
```

Similar for the $\epsilon 2\epsilon 3$ genotype

```
Clear[β23];
Solve[∏x=1^4 (0.5006 + β23 x^2) / ∏x=1^4 (0.5006 + 0.3049 x^2) == 0.9099, β23]
β23 = Select[Table[β23 /. %], Positive]

{{β23 → -0.674926}, {β23 → -0.166313 - 0.404642 i},
 {β23 → -0.166313 + 0.404642 i}, {β23 → 0.294892}}

{0.294892}
```

The $\epsilon 2\epsilon 4$ genotypes and the $\epsilon 3\epsilon 3$ genotypes were considered equal so they also share the same slope

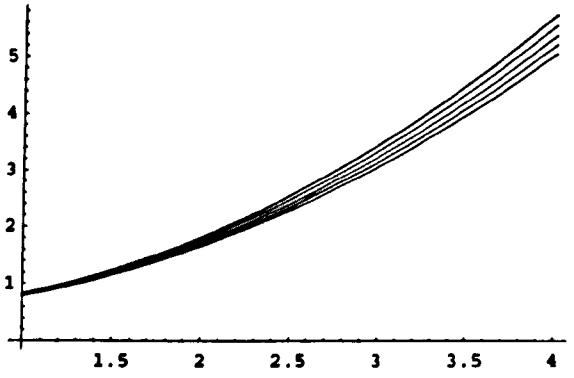
```
β = 0.3049;
```

For the ϵ3ϵ4 genotype

```
Clear[β34];
Solve[∏x=1^4 (0.5006 + β34 x^2) / ∏x=1^4 (0.5006 + 0.3049 x^2) == 1.1093, β34]
β34 = Select[Table[β34 /. %], Positive]

{{β34 → -0.694067}, {β34 → -0.167393 - 0.428609 i},
 {β34 → -0.167393 + 0.428609 i}, {β34 → 0.316193}}

{0.316193}
```

And finally for the ϵ4ϵ4 genotype

```
Clear[β44];
Solve[∏x=1^4 (0.5006 + β44 x^2) / ∏x=1^4 (0.5006 + 0.3049 x^2) == 1.2305, β44]
β44 = Select[Table[β44 /. %], Positive]

{{β44 → -0.704621}, {β44 → -0.167921 - 0.441591 i},
 {β44 → -0.167921 + 0.441591 i}, {β44 → 0.327803}}

{0.327803}

allgenot = Plot[{0.5006 + β22 x^2, 0.5006 + β23 x^2,
   0.5006 + β x^2, 0.5006 + β34 x^2, 0.5006 + β44 x^2}, {x, 1, 4}]
TableForm[Table[{0.5006 + β22 x^2, 0.5006 + β23 x^2, 0.5006 + β x^2,
   0.5006 + β34 x^2, 0.5006 + β44 x^2}, {x, 1, 4}],
 TableHeadings -> {None, {"ϵ2ϵ2", "ϵ2ϵ3", "ϵ3ϵ3/ϵ2ϵ4", "ϵ3ϵ4", "ϵ4ϵ4"}}]
```




```
- Graphics -
```

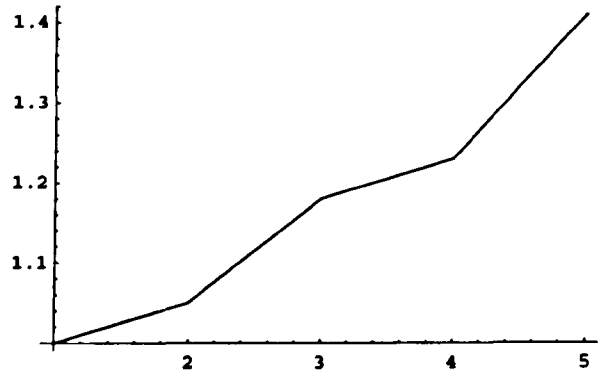
ϵ2ϵ2	ϵ2ϵ3	ϵ3ϵ3/ϵ2ϵ4	ϵ3ϵ4	ϵ4ϵ4
0.785735	0.795492	0.8055	0.816793	0.828403
1.64114	1.68017	1.7202	1.76537	1.81181
3.06682	3.15463	3.2447	3.34634	3.45083
5.06277	5.21888	5.379	5.55969	5.74545

We can see that by fitting the data we have lost the reference risk of 1.0. Since the fitting of the curve to the original data was good and the relative risk between the genotypes remains the same then this difference will not cause significant changes to the selection of apo E, in which we are interested.

■ apo E × Exercise

In the discussion on the way that exercise affects HDL and its association with Apo E, we concluded that the effect of exercise on blood lipoproteins was higher in individuals that exhibited initially high HDL associated relative risk. Thus that exercise will be more beneficial for the $\epsilon 4$ carriers than for those with the $\epsilon 2$. So in a similar way to diet, we will expect that there is a level of exercise in which there will be no differences between the genotypes, while for less physical activity the relative risk will be higher. As in the previous cases we will use the data by Stampfer et al (2000) with the addition of a point for x=0 that will stand for the point of common risk between the genotypes.

```
datacurve = ListPlot[{1.0, 1.05, 1.18, 1.23, 1.41}, PlotJoined → True]
```

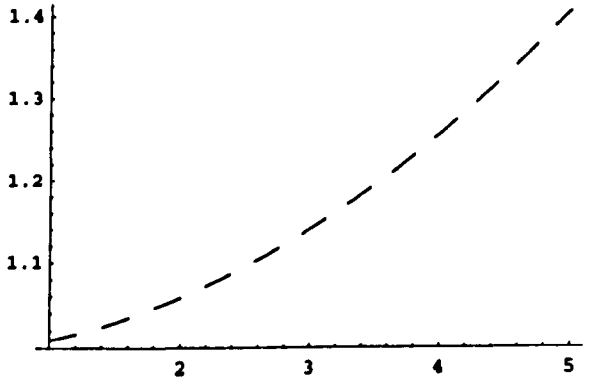



```
- Graphics -
```

```
<< Statistics`LinearRegression`

Regress[{1.0, 1.05, 1.18, 1.23, 1.41}, {1, x^2}, x]
```

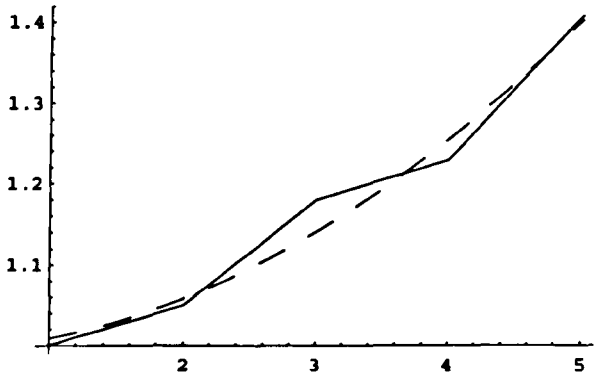
```
                          Estimate        SE            TStat        PValue
{ParameterTable → 1       0.992235        0.0204706     48.4711      0.0000193356,
                  x²      0.0165241       0.00146294    11.2951      0.00148825

 RSquared → 0.977026, AdjustedRSquared → 0.969367,
 EstimatedVariance → 0.000800428, ANOVATable →

           DF      SumOfSq         MeanSq          FRatio      PValue
  Model    1       0.102119        0.102119        127.58      0.00148825}
  Error    3       0.00240128      0.000800428
  Total    4       0.10452
```

```
fitlinear = Plot[0.9922 + 0.0165 x², {x, 1, 5}, PlotStyle → {{Dashing[{0.05, 0.05}]}}]
```




```
- Graphics -
```

```
Show[datacurve, fitlinear]
data = {1.0, 1.05, 1.18, 1.23, 1.41}
datalinear = Table[0.9922 + 0.0165 x^2, {x, 1, 5}]
```




```
- Graphics -

{1., 1.05, 1.18, 1.23, 1.41}

{1.0087, 1.0582, 1.1407, 1.2562, 1.4047}
```

We can see that the model fits the data very closely. Now we can make predictions for the risk factors for the rest of genotypes setting this curve as the one corresponding to the $\epsilon 3\epsilon 3$ genotype.

For the $\epsilon 2\epsilon 2$ genotype

```
Clear[β22];
Solve[∏x=1^5 (0.9922 + β22 x^2) / ∏x=1^5 (0.9922 + 0.0165 x^2) == 0.8279, β22]
β22 = Select[Table[β22 /. %], Positive]

{{β22 → -0.991987}, {β22 → -0.270305}, {β22 → -0.101076 - 0.0840309 i},
 {β22 → -0.101076 + 0.0840309 i}, {β22 → 0.0122493}}

{0.0122493}
```

We require only the real and positive solutions so in this case the only acceptable solution is for $\beta 22 = 0.0123$.

Similarly for the $\epsilon 2\epsilon 3$ genotype

```
Clear[β23];
Solve[∏x=1^5 (0.9922 + β23 x^2) / ∏x=1^5 (0.9922 + 0.0165 x^2) == 0.9099, β23]
β23 = Select[Table[β23 /. %], Positive]

{{β23 → -0.991966}, {β23 → -0.271954}, {β23 → -0.10131 - 0.0869627 i},
 {β23 → -0.10131 + 0.0869627 i}, {β23 → 0.0143447}}

{0.0143447}
```

The $\epsilon 2\epsilon 4$ genotypes and the $\epsilon 3\epsilon 3$ genotypes were considered equal so they also share the same slope

```
β = 0.0165;
```

For the ϵ3ϵ4 genotype

```
Clear[β34];
Solve[∏x=1^5 (0.9922 + β34 x^2) / ∏x=1^5 (0.9922 + 0.0165 x^2) == 1.1093, β34]
β34 = Select[Table[β34 /. %], Positive]
```

```
{{β34 → -0.991915}, {β34 → -0.275724}, {β34 → -0.101747 - 0.093324 i},
 {β34 → -0.101747 + 0.093324 i}, {β34 → 0.0189388}}
```

```
{0.0189388}
```

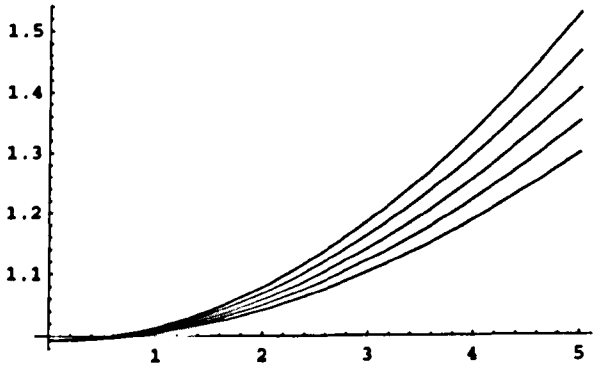
And finally for the ϵ4ϵ4 genotype

```
Clear[β44];
Solve[∏x=1^5 (0.9922 + β44 x^2) / ∏x=1^5 (0.9922 + 0.0165 x^2) == 1.2305, β44]
β44 = Select[Table[β44 /. %], Positive]
```

```
{{β44 → -0.991884}, {β44 → -0.277872}, {β44 → -0.101946 - 0.0967685 i},
 {β44 → -0.101946 + 0.0967685 i}, {β44 → 0.0214526}}
```

```
{0.0214526}
```

```
allgenot = Plot[{0.9922 + β22 x^2, 0.9922 + β23 x^2,
   0.9922 + β x^2, 0.9922 + β34 x^2, 0.9922 + β44 x^2}, {x, 0, 5}]
TableForm[Table[{0.9922 + β22 x^2, 0.9922 + β23 x^2, 0.9922 + β x^2,
   0.9922 + β34 x^2, 0.9922 + β44 x^2}, {x, 0, 5}],
 TableHeadings -> {None, {"ϵ2ϵ2", "ϵ2ϵ3", "ϵ3ϵ3/ϵ2ϵ4", "ϵ3ϵ4", "ϵ4ϵ4"}}]
```



- Graphics -

ϵ2ϵ2	ϵ2ϵ3	ϵ3ϵ3/ϵ2ϵ4	ϵ3ϵ4	ϵ4ϵ4
0.9922	0.9922	0.9922	0.9922	0.9922
1.00445	1.00654	1.0087	1.01114	1.01365
1.0412	1.04958	1.0582	1.06796	1.07801
1.10244	1.1213	1.1407	1.16265	1.18527
1.18819	1.22172	1.2562	1.29522	1.33544
1.29843	1.35082	1.4047	1.46567	1.52851

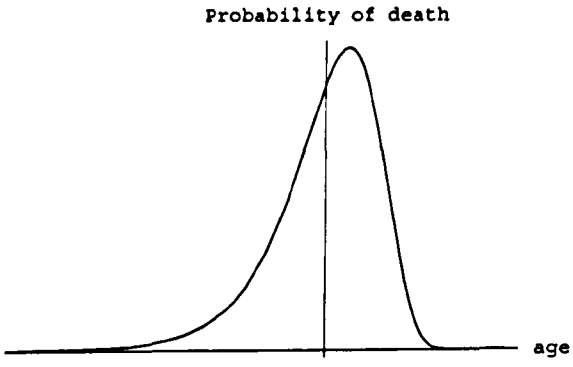
Appendix 6. *Mathematica* notebook for the comparison of statistical distributions and the Gompertz curve.

```
<< Statistics'ContinuousDistributions'
<< Statistics'NormalDistribution'
<< Statistics'DescriptiveStatistics'
<< Graphics'Arrow'
<< Graphics'Colors'
<< Statistics'LinearRegression'
```

First we construct a modified Gompertz curve describing the probability of death at each age for the population.

```
α := 3.51364 10^-5
β := 0.093832

μ[x_] := α e^(β x)

meanLifeSpan[x_] = Sum[x (e^-∫_1^x μ[y-1] dy - e^-∫_0^x μ[y] dy), {x, 0, 122}];

Gompertz[x_] = e^-∫_1^x μ[y-1] dy - e^-∫_0^x μ[y] dy;

Print["Gompertz Curve"]
GompertzPDF = Plot[Gompertz[x], {x, 0, 125},
    AxesLabel → {"age", "Probability of death"}, PlotRange → All,
    AxesOrigin → {meanLifeSpan[x], 0}, Ticks → None]

Print["Mean Life Span ->", meanLifeSpan[x]]
```

```
Gompertz Curve
```

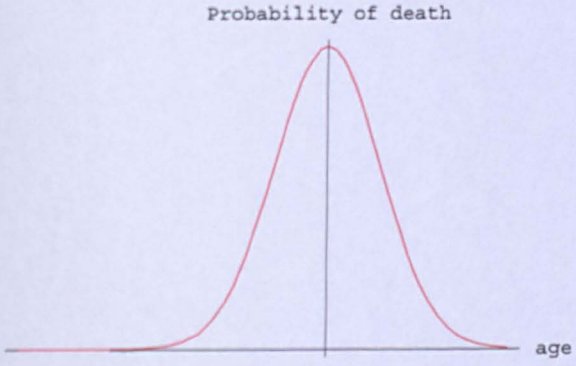



```
- Graphics -

Mean Life Span ->78.4683
```

The y-axis meets the x-axis at the point of the mean age of the distribution.

Then we can construct a probability density function (PDF) of a normal distribution with the same mean and range, calculated from the expected values for the Gompertz model.

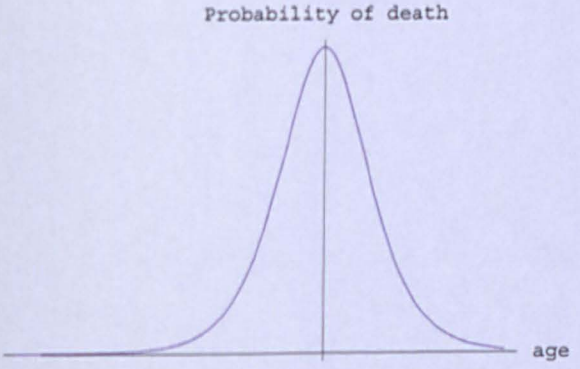
```
SimpleDistr = NormalDistribution[meanLifeSpan[x],
  Sqrt[Sum[(x - meanLifeSpan[x])^2 (E^-Integrate[μ[y-1] dy, {y, 1, x}] - E^-Integrate[μ[y] dy, {y, 0, x}]), {x, 0, 122}]]];

FisherMeanLong = Mean[FisherDistr];

NormalPDF = Plot[PDF[SimpleDistr, x], {x, 0, 125},
  AxesLabel → {"age", "Probability of death"}, PlotRange → All,
  AxesOrigin → {meanLifeSpan[x], 0}, Ticks → None, PlotStyle → Red]
```




```
- Graphics -
```

Another very popular distribution in demographic analysis is the Logistic distribution.

```
parameterLogist = 1/π (Sqrt[3 × Sum[(x - meanLifeSpan[x])^2 (E^-Integrate[μ[y-1] dy, {y, 1, x}] - E^-Integrate[μ[y] dy, {y, 0, x}]), {x, 0, 122}]]);

LogisticPDF = Plot[PDF[LogisticDistribution[meanLifeSpan[x], parameterLogist], x],
  {x, 0, 125}, AxesLabel → {"age", "Probability of death"}, PlotRange → All,
  AxesOrigin → {meanLifeSpan[x], 0}, Ticks → None, PlotStyle → BlueViolet]
```




```
- Graphics -
```

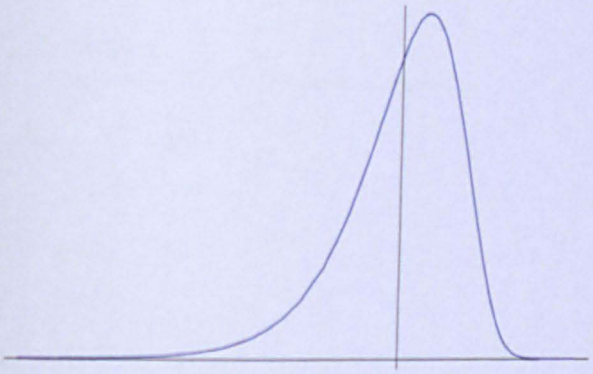
Finally we construct an Extreme Value distribution for the minimum using the parameters of the modified Gompertz curve.

$$\text{parameterExtreme} = \frac{1}{\pi}\left(\sqrt{\sum_{x=0}^{122}(x - \text{meanLifeSpan}[x])^2\left(e^{-\int_1^x \mu[y-1]\,dy} - e^{-\int_0^x \mu[y]\,dy}\right)} \times \sqrt{6}\right)$$

10.5822

```
PDF[ExtremeValueDistribution[a, b], x]
ExtrPDF = Plot[PDF[ExtremeValueDistribution[-85, parameterExtreme], -x], {x, -8, 120},
   PlotRange → All, AxesOrigin → {-Mean[ExtremeValueDistribution[-85, parameterExtreme]], 0},
   Ticks → None, PlotStyle → Blue]
```

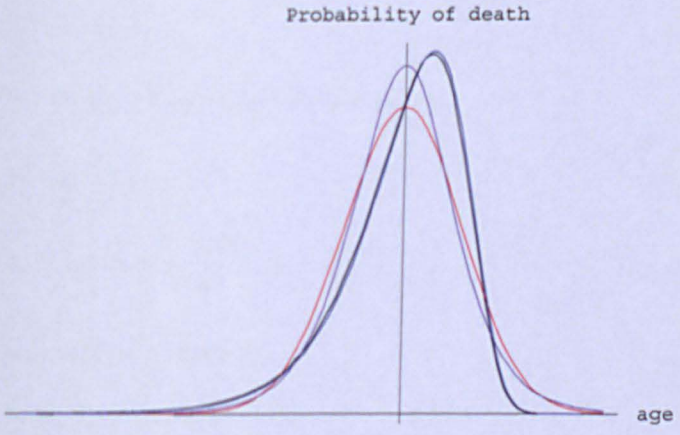
$$\frac{e^{-e^{\frac{a-x}{b}}+\frac{a-x}{b}}}{b}$$



```
- Graphics -
```

If we compare all of them together we can see how closely similar are the Extreme Value distribution for the minimum and our model.

```
Show[GompertzPDF, NormalPDF, LogisticPDF, ExtrPDF]
```



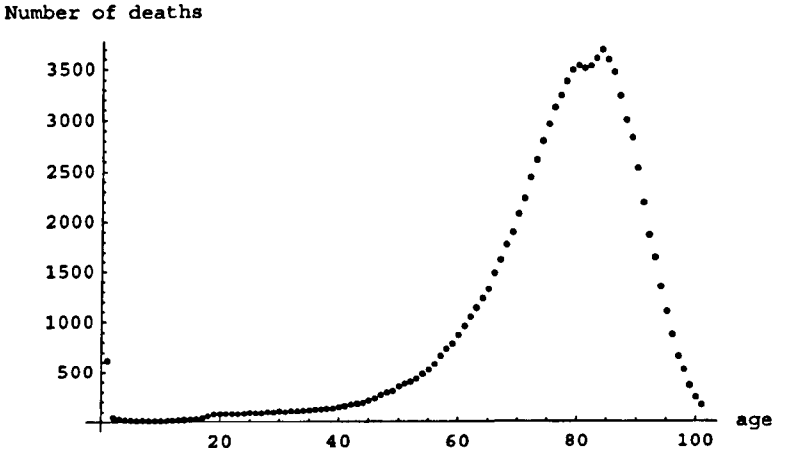

```
- Graphics -
```

Appendix 7. *Mathematica* notebook for the calculation of the Gompertz and Gumbel distribution parameters.

```
<< Statistics'LinearRegression'
<< Statistics'DescriptiveStatistics'
```

The data for the probability of death at each age and their graph can be seen below.

```
data = {615.9, 46.5, 28.2, 18.4, 15.5, 11.2, 14.4, 12.6, 11.1, 10.4, 12.7, 12.2, 16.1, 16.9, 22.0, 23.5, 36.7,
   58.1, 76.7, 77.7, 77.0, 79.9, 79.5, 82.1, 88.9, 85.3, 84.0, 93.6, 89.8, 100.9, 95.9, 102.7, 105.3, 109.5,
   114.8, 116.9, 121.6, 128.3, 134.2, 148.1, 159.7, 170.9, 180.1, 190.1, 209.8, 233.6, 265.6, 291.4, 308.4,
   353.9, 384.2, 400.7, 433.4, 481.1, 524.6, 581.7, 660.7, 731.1, 785.6, 871.9, 963.3, 1053.9, 1139.3,
   1232.7, 1326.3, 1481.5, 1613.8, 1775.9, 1894.2, 2079.4, 2237.8, 2437.5, 2611.8, 2796.0, 2957.9, 3126.5,
   3239.7, 3384.2, 3491.2, 3535.9, 3508.7, 3533.9, 3605.7, 3690.5, 3592.7, 3473.1, 3236.9, 2999.0,
   2828.7, 2531.3, 2190.0, 1868.0, 1632.7, 1347.0, 1098.9, 869.8, 655.2, 516.9, 359.3, 243.4, 166.5};

ListPlot[data,
  AxesLabel → {"age", "Number of deaths"}]
```




```
- Graphics -
```

Predicted by the given frequencies of deaths per age

$$\text{total} = \sum_{i=1}^{100} \text{data}[[i]];$$

$$\text{PDF} = \text{Table}\left[\frac{\text{data}[[x]]}{\text{total}}, \{x, 1, 100\}\right];$$

$$\text{meanvalue} = \sum_{x=1}^{100} \text{PDF}[[x]]\, x;$$

```
Print["Mean is ", meanvalue]
```

Mean is 75.9604

Predicted by a rounded frequency population

```
DataList = Flatten[Table[Table[x, {i, 1, Round[data[[x]]]}], {x, 1, 100}]];
```

```
meanSim = N[Mean[DataList]];
Print["Mean is ", meanSim]
varianceSim = N[Variance[DataList]];
Print["Variance is ", varianceSim]
SD = √N[Variance[DataList]] ;
Print["Standard deviation is ", SD]
```

Mean is 75.9592

Variance is 224.721

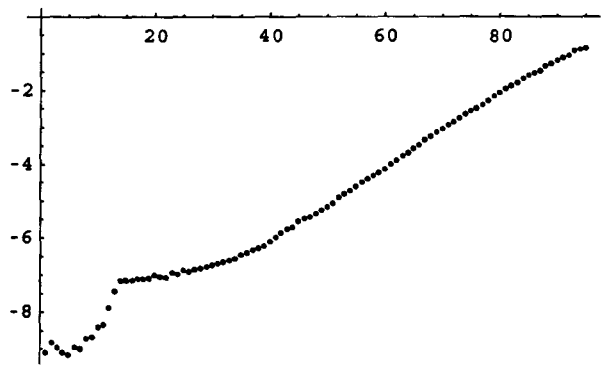
Standard deviation is 14.9907

It can be easily seen that the shape of the data corresponds to the prediction of the Gompertz law as seen earlier

The data for the natural logarithm of the mortality in the population together with their graph can be seen below. To accurate calculate the slope of the line we have to avoid child mortality, which is not accounted by the Gompertz model. The most accurately part of the curve to calculate the Gompertz parameters is between 6-99 years.

```
lndata = Drop[Drop[{-5.086760684, -7.667042262, -8.16653632, -8.595154733, -8.765654551, -9.088122739,
      -8.838776816, -8.971323472, -9.097011687, -9.171119659, -8.963480294, -9.003326203,
      -8.721760357, -8.679712121, -8.412833176, -8.347450417, -7.899308495, -7.440485738,
      -7.162647529, -7.148553766, -7.156216638, -7.119711641, -7.123426815, -7.090476916,
      -7.010896038, -7.050471581, -7.064409089, -6.955895654, -6.996586493, -6.879167822,
      -6.928978915, -6.858965115, -6.832647806, -6.793534135, -6.744786451, -6.725433722,
      -6.685412048, -6.630123542, -6.584223554, -6.483141352, -6.406374114, -6.336775732,
      -6.282886939, -6.227186881, -6.12613945, -6.016577248, -5.885304351, -5.789667214,
      -5.730023786, -5.588937083, -5.502884532, -5.456375671, -5.373687674, -5.264302488,
      -5.17238981, -5.063087154, -4.928824919, -4.820097583, -4.739730679, -4.626189553,
      -4.516061011, -4.414715129, -4.324210659, -4.231539801, -4.143072728, -4.015496414,
      -3.910973556, -3.793973339, -3.705984922, -3.586539496, -3.483948754, -3.365897808,
      -3.260427252, -3.151754704, -3.050441923, -2.945096195, -2.854493162, -2.750169912,
      -2.651915645, -2.565561545, -2.49362454, -2.399911304, -2.283684833, -2.151625213,
      -2.05657694, -1.956042306, -1.879574806, -1.796888601, -1.679683354, -1.59641804,
      -1.531778041, -1.467343909, -1.358616939, -1.283737773, -1.197662755, -1.116111511,
      -1.060958187, -0.927508415, -0.883022798, -0.846713365, -0.777888409}, 5], -1];
```

lndataPlot = ListPlot[lndata]



- Graphics -

Now we can fit a line to the data considering the equation:
$y(x) = \text{Ln}(\alpha) + \beta x$

Regress[lndata, {1, x}, x]

		Estimate	*SE*	*TStat*	*PValue*
ParameterTable →	*1*	*−9.25216*	*0.0584783*	*−158.215*	*0.*
	x	*0.0877779*	*0.00105784*	*82.9788*	*0.*

RSquared → *0.986673*, *AdjustedRSquared* → *0.98653*, *EstimatedVariance* → *0.0799424*,

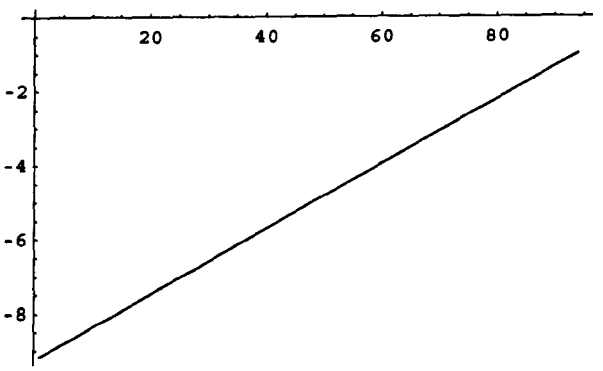
		DF	*SumOfSq*	*MeanSq*	*FRatio*	*PValue*
ANOVATable →	*Model*	*1*	*550.442*	*550.442*	*6885.48*	*0.*
	Error	*93*	*7.43465*	*0.0799424*		
	Total	*94*	*557.877*			

Solve[Log[α] == −9.252156, α]

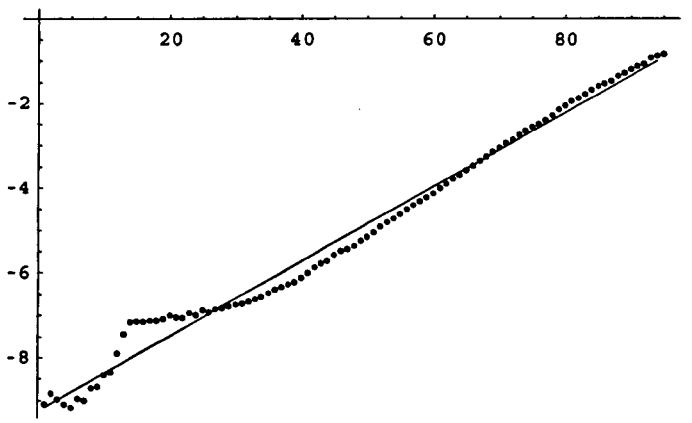
{{α → 0.0000959047}}

So the Gompertz parameters for ages 6-99 will be $\alpha = 0.0000959047$ and $\beta = 0.0877779$

lnregRessionPlot = Plot[−9.252156 + 0.0877779 x, {x, 1, 94}];



Show[lndataPlot, lnregRessionPlot]



- Graphics -

Appendix 8

```
//Fotios Drenos
;//26-08-2003
//Version 1.1
//An alternative way to work the Apo-E
//model for fast times using c++

//Header files used
#include <iostream>
//#include <process.h>
#include <time.h>
#include <gsl/gsl_rng.h>
#include <gsl/gsl_randist.h>
#include <math.h>
#include <sstream>
#include <string>
#include <fstream>
#include <cstdlib>
#include <ctime>
#include <sys/types.h>
#include <unistd.h>



using namespace std;

//Definitions of program variables

gsl_rng *r;

const gsl_rng_type *T;

long seed;

int female_genotype;
int male_genotype;

#define MIN(X,Y) ((X) < (Y) ?  (X) : (Y))

//Inline function for the construction of the curve sum
inline double temp_sum(int age, double total_risk) {

    double alpha = 3.51364 * pow(10.0, -5.0);
    double beta = 0.093832;
      double temp;

      temp = (pow(M_E, -(total_risk * alpha *
        (-(1 / beta) + (pow(M_E, -beta + age *beta) /
            beta)))) - pow(M_E, -(total_risk *
            alpha * (-(1 / beta) + (pow(M_E, age * beta) /
            beta)))));

      return (temp);
}

//Function for random number generator of an Extreme value
//distribution for the minimum. also known as log-Weibull and
//Gumbel type I distribution.

double gsl_ran_gumbelcustom (const gsl_rng * r, const double a, const
double b)
{
  double x = gsl_rng_uniform_pos (r);

  double z = a + b * log(log(1 / (1 - x)));

  return z;
}


//Female risk factor
double freq2, freq3, freq4;
double frq22, frq23, frq33, frq24, frq34, frq44;

double female (double frq22,double frq23,double frq33,
                    double frq24,double frq34,double frq44)
{
      int diet, alcohol, smoking, exercise;
      int sex = 1;

      //Random numbers definitions
      double random_genotype = gsl_rng_uniform(r);
      double random_diet = gsl_rng_uniform(r);
      double random_alcohol = gsl_rng_uniform(r);
    double random_smoking = gsl_rng_uniform(r);
    double random_exercise = gsl_rng_uniform(r);

      double random_genotype_temp = random_genotype;
      double random_diet_temp = random_diet;
    double random_alcohol_temp = random_alcohol;
    double random_smoking_temp = random_smoking;
    double random_exercise_temp = random_exercise;

      //Parameters of the normal curve
      double sex_risk, diet_risk, alcohol_risk, smoking_risk,
            exercise_risk, total_risk;

      int age1 = 0;
      int age2 = 0;
    double temp1 = 0;
      double temp2 = 0;
      double temp3 = 0;
      double temp4 = 0;
      double alpha = 3.51364 * pow(10.0, -5.0);
      double beta = 0.093832;
      double mean_of_distribution = 0;


      //Genotype declaration
      if (random_genotype_temp <= frq22) {
        female_genotype = 22;
        }
     else if (random_genotype_temp <= (frq22 + frq23)) {
        female_genotype = 23;
        }
     else if (random_genotype_temp <= (frq22 + frq23 + frq33)) {
            female_genotype = 33;
            }
```

```
    else if (random_genotype_temp <= (frq22 + frq23 + frq33 + frq24))
{
        female_genotype = 24;
            }
    else if (random_genotype_temp <=
      (frq22 + frq23 + frq33 + frq24 + frq34)) {
        female_genotype = 34;
        }
    else if (random_genotype_temp <=
      (frq22 + frq23 + frq33 + frq24 + frq34 + frq44)) {
        female_genotype = 44;
        }
      else {
            cout << "error in genotype declaration" << "\n";
      }

      //Diet group selected
      if (random_diet_temp <= 0.00) {
            diet = 0;
            }
       else if (random_diet_temp <= 0.200) {
         diet = 1;
            }
     else if (random_diet_temp <= 0.679) {
         diet = 2;
         }
     else if (random_diet_temp <= 0.715) {
         diet = 3;
         }
       else if (random_diet_temp <= 0.918) {
         diet = 4;
         }
     else if (random_diet_temp <= 1.00) {
         diet = 5;
         }
     else {
            cout << "error in diet selection" << "\n";
       }
       //diet = 5;

     //Alcohol consumption group selected
       if (random_alcohol_temp <= 0.33) {
         alcohol = 1;
         }
     else if (random_alcohol_temp <= 0.52) {
         alcohol = 2;
         }
     else if (random_alcohol_temp <= 0.66) {
         alcohol = 3;
         }
     else if (random_alcohol_temp <= 0.76) {
         alcohol = 4;
         }
     else if (random_alcohol_temp <= 0.86) {
         alcohol = 5;
         }
       else if (random_alcohol_temp <= 0.90) {
         alcohol = 6;
         }
       else if (random_alcohol_temp <= 1.00) {
         alcohol = 7;
       }
     else {
          cout << "error in alcohol selection" << "\n";
     }
     //alcohol = 1;

    //Smoking group selected
      if (random_smoking_temp <= 0.44) {
        smoking = 1;
        }
    else if (random_smoking_temp <= 0.78){
        smoking = 2;
        }
      else if (random_smoking_temp <= 0.85){
        smoking = 3;
        }
      else if (random_smoking_temp <= 1.0){
        smoking = 4;
        }
      else {
           cout << "error in smoking selection" << "\n";
     }
      //smoking = 4;

    //Physical activity group selected
      if (random_exercise_temp <= 0.00) {
        exercise = 1;
        }
    else if (random_exercise_temp <= 0.19) {
        exercise = 2;
        }
    else if (random_exercise_temp <= 0.40) {
        exercise = 3;
        }
      else if (random_exercise_temp <= 0.60){
        exercise = 4;
        }
      else if (random_exercise_temp <= 0.77) {
        exercise = 5;
        }
      else if (random_exercise_temp <= 1.00) {
        exercise = 6;
        }
      else {
           cout << "error in exercise selection" << "\n";
     }
      //exercise = 6;

    //Sex assosiated risk
      if (sex == 0) {
        sex_risk = 0.58;
        }
    else if (sex == 1) {
        sex_risk = 1.0;
        }
    else {
        cout << "fail to allocate sex" << "\n";
        }

    //Apo E x Diet risk factors
      if (female_genotype == 33) {
```

```
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.9842;
        }
    else if (diet == 2) {
        diet_risk = 1.1042;
        }
    else if (diet == 3) {
        diet_risk = 1.3042;
        }
    else if (diet == 4) {
        diet_risk = 1.5842;
        }
    else {
        diet_risk = 1.9442;
        }
    }
else if (female_genotype == 34) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.98725;
        }
    else if (diet == 2) {
        diet_risk = 1.1164;
        }
    else if (diet == 3) {
        diet_risk = 1.33165;
        }
    else if (diet == 4) {
        diet_risk = 1.633;
        }
    else {
        diet_risk = 2.02046;
        }
    }
else if (female_genotype == 23) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.981499;
        }
    else if (diet == 2) {
        diet_risk = 1.09339;
        }
    else if (diet == 3) {
        diet_risk = 1.27989;
        }
    else if (diet == 4) {
        diet_risk = 1.54098;
        }
    else {
        diet_risk = 1.87666;
        }
    }
else if (female_genotype == 44) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.990388;
        }
    else if (diet == 2) {
        diet_risk = 1.12895;
        }
    else if (diet == 3) {
        diet_risk = 1.35989;
        }
    else if (diet == 4) {
        diet_risk = 1.6832;
        }
    else {
        diet_risk = 2.09889;
        }
    }
else if (female_genotype == 24) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.9842;
        }
    else if (diet == 2) {
        diet_risk = 1.1042;
        }
    else if (diet == 3) {
        diet_risk = 1.3042;
        }
    else if (diet == 4) {
        diet_risk = 1.5842;
        }
    else {
        diet_risk = 1.9442;
        }
    }
else if (female_genotype == 22) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.978866;
        }
    else if (diet == 2) {
        diet_risk = 1.08287;
        }
    else if (diet == 3) {
        diet_risk = 1.2562;
        }
    else if (diet == 4) {
        diet_risk = 1.49886;
        }
    else {
            diet_risk = 1.81086;
        }
    }
else {
    cout << "error in genotype-diet selection" << "\n";
    }
```

```
  //Apo E x Alcohol risk factors
  if (female_genotype == 33) {
    if (alcohol == 1) {
        alcohol_risk = 1.7119;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.3164;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.0791;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.00;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.0791;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.3164;
        }
    else {
        alcohol_risk = 1.7119;
        }
    }
else if (female_genotype == 34) {
    if (alcohol == 1) {
        alcohol_risk = 1.7631;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.33915;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.08479;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.00;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.08479;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.33915;
        }
    else {
        alcohol_risk = 1.7631;
        }
    }
else if (female_genotype == 23) {
    if (alcohol == 1) {
        alcohol_risk = 1.6631;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.29614;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.07403;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.0;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.07403;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.29614;
        }
    else {
        alcohol_risk = 1.66631;
        }
    }
else if (female_genotype == 44) {
    if (alcohol == 1) {
        alcohol_risk = 1.81546;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.36243;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.09061;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.0;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.09061;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.36243;
        }
    else {
        alcohol_risk = 1.81546;
        }
    }
else if (female_genotype == 24) {
    if (alcohol == 1) {
        alcohol_risk = 1.7119;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.3164;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.0791;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.00;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.0791;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.3164;
        }
    else {
        alcohol_risk = 1.7119;
        }
    }
else if (female_genotype == 22) {
    if (alcohol == 1) {
        alcohol_risk = 1.62166;
        }
```

```
    else if (alcohol == 2) {
        alcohol_risk = 1.27629;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.06907;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.00;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.06907;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.27629;
        }
    else {
        alcohol_risk = 1.62166;
        }
    }
else {
    cout << "error in genotype-alcohol selection" << "\n";
    }


  //Apo E x smoking risk factors
  if (female_genotype == 33) {
    if (smoking == 1) {
        smoking_risk = 0.8055;
        }
    else if (smoking == 2) {
        smoking_risk = 1.7202;
        }
    else if (smoking == 3) {
        smoking_risk = 3.2447;
        }
    else {
        smoking_risk = 5.379;
        }
    }
else if (female_genotype == 34) {
    if (smoking == 1) {
        smoking_risk = 0.816793;
        }
    else if (smoking == 2) {
        smoking_risk = 1.76537;
        }
    else if (smoking == 3) {
        smoking_risk = 3.34634;
        }
    else {
        smoking_risk = 5.55969;
        }
    }
else if (female_genotype == 23) {
    if (smoking == 1) {
        smoking_risk = 0.795492;
        }
    else if (smoking == 2) {
        smoking_risk = 1.68017;
        }
    else if (smoking == 3) {
        smoking_risk = 3.15463;
        }
    else {
        smoking_risk = 5.21888;
        }
    }
else if (female_genotype == 44) {
    if (smoking == 1) {
        smoking_risk = 0.828403;
        }
    else if (smoking == 2) {
        smoking_risk = 1.81181;
        }
    else if (smoking == 3) {
        smoking_risk = 3.45083;
        }
    else {
        smoking_risk = 5.74545;
        }
    }
else if (female_genotype == 24) {
    if (smoking == 1) {
        smoking_risk = 0.8055;
        }
    else if (smoking == 2) {
        smoking_risk = 1.7202;
        }
    else if (smoking == 3) {
        smoking_risk = 3.2447;
        }
    else {
        smoking_risk = 5.379;
        }
    }
else if (female_genotype == 22) {
    if (smoking == 1) {
        smoking_risk = 0.785735;
        }
    else if (smoking == 2) {
        smoking_risk = 1.64114;
        }
    else if (smoking == 3) {
        smoking_risk = 3.06682;
        }
    else {
        smoking_risk = 5.06277;
        }
    }
else {
    cout << "error in genotype-smoking selection" << "\n";
    }


  //Apo E x Exercise risk factors
  if (female_genotype == 33) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.0087;
        }
```

```
    else if (exercise == 2) {
        exercise_risk = 1.0582;
        }
    else if (exercise == 3) {
        exercise_risk = 1.1407;
        }
    else if (exercise == 4) {
        exercise_risk = 1.2562;
        }
    else {
        exercise_risk = 1.4047;
        }
    }
else if (female_genotype == 34) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.01114;
        }
    else if (exercise == 2) {
        exercise_risk = 1.06796;
        }
    else if (exercise == 3) {
        exercise_risk = 1.16265;
        }
    else if (exercise == 4) {
        exercise_risk = 1.29522;
        }
    else {
        exercise_risk = 1.46567;
        }
    }
else if (female_genotype == 23) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.00654;
        }
    else if (exercise == 2) {
        exercise_risk = 1.04958;
        }
    else if (exercise == 3) {
        exercise_risk = 1.1213;
        }
    else if (exercise == 4) {
        exercise_risk = 1.22172;
        }
    else {
        exercise_risk = 1.35082;
        }
    }
else if (female_genotype == 44) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.01365;
        }
    else if (exercise == 2) {
        exercise_risk = 1.07801;
        }
    else if (exercise == 3) {
        exercise_risk = 1.18527;
        }
    else if (exercise == 4) {
        exercise_risk = 1.33544;
        }
    else {
        exercise_risk = 1.52851;
        }
    }
else if (female_genotype == 24) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.0087;
        }
    else if (exercise == 2) {
        exercise_risk = 1.0582;
        }
    else if (exercise == 3) {
        exercise_risk = 1.1407;
        }
    else if (exercise == 4) {
        exercise_risk = 1.2562;
        }
    else {
        exercise_risk = 1.4047;
        }
    }
else if (female_genotype == 22) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.00445;
        }
    else if (exercise == 2) {
        exercise_risk = 1.0412;
        }
    else if (exercise == 3) {
        exercise_risk = 1.10244;
        }
    else if (exercise == 4) {
        exercise_risk = 1.18819;
        }
    else {
            exercise_risk = 1.29843;
        }
    }
else {
    cout << "error in genotype-exercise selection" << "\n";
    }


//Total risk
  total_risk = sex_risk * diet_risk * alcohol_risk *
               smoking_risk * exercise_risk;
```

```
      //Gumbell PDF curve parameters calculation
      while (age1 <= 122) {
        ++age1;
        temp1 = age1 * temp_sum(age1, total_risk);
        temp2 += temp1;
        }

      mean_of_distribution = temp2;

    while (age2 <= 122) {
        ++age2;
        temp3 = (age2 - mean_of_distribution)*
            (age2 - mean_of_distribution)*
            temp_sum(age2, total_risk);
        temp4 += temp3;
        }

      double sigma = sqrt(temp4);

      double betaGumble = sigma * sqrt(6.0) / 3.14159;
      double alphaGumble = mean_of_distribution + beta * 0.5772;


      double temp_ran_gumbelcustom = gsl_ran_gumbelcustom(r,
alphaGumble, betaGumble);

      double random_death_age = temp_ran_gumbelcustom;

      /*cout << "\nthe female genotype is " << female_genotype <<
"\n"
           << "total risk factor " << total_risk << "\n"
           << "the mean of the disribution " << mean_of_distribution
<< "\n"
           << "sigma is " << sigma << "\n"
           << "random death age is = " << random_death_age <<
"\n";*/

      return (random_death_age);
}



//Male risk factor
double male (double frq22,double frq23,double frq33,
             double frq24,double frq34,double frq44)
{
      int diet, alcohol, smoking, exercise;
      int sex = 0;

      //Random numbers definitions
      double random_genotype = gsl_rng_uniform(r);
      double random_diet = gsl_rng_uniform(r);
     double random_alcohol = gsl_rng_uniform(r);
     double random_smoking = gsl_rng_uniform(r);
     double random_exercise = gsl_rng_uniform(r);

       double random_genotype_temp = random_genotype;
       double random_diet_temp = random_diet;
     double random_alcohol_temp = random_alcohol;
   double random_smoking_temp = random_smoking;
   double random_exercise_temp = random_exercise;

     //Parameters of the normal curve
     double sex_risk, diet_risk, alcohol_risk, smoking_risk,
           exercise_risk, total_risk;

     int age1 = 0;
     int age2 = 0;
   double temp1 = 0;
     double temp2 = 0;
     double temp3 = 0;
     double temp4 = 0;
     double alpha = 3.51364 * pow(10.0, -5.0);
     double beta = 0.093832;
     double mean_of_distribution = 0;


     //Genotype declaration
     if (random_genotype_temp <= frq22) {
       male_genotype = 22;
       }
    else if (random_genotype_temp <= (frq22 + frq23)) {
       male_genotype = 23;
       }
    else if (random_genotype_temp <= (frq22 + frq23 + frq33)) {
           male_genotype = 33;
           }
    else if (random_genotype_temp <= (frq22 + frq23 + frq33 + frq24))
{
       male_genotype = 24;
           }
    else if (random_genotype_temp <=
      (frq22 + frq23 + frq33 + frq24 + frq34)) {
       male_genotype = 34;
       }
    else if (random_genotype_temp <=
      (frq22 + frq23 + frq33 + frq24 + frq34 + frq44)) {
       male_genotype = 44;
       }
     else {
           cout << "error in genotype declaration" << "\n"
                << "random_genotype_temp " << random_genotype_temp
<< "\n"
                << "with frq22 " << frq22
                << "\n(frq22 + frq23) " << frq22 + frq23
                << "\n(frq22 + frq23 + frq33) " << frq22 + frq23 +
frq33
                << "\n(frq22 + frq23 + frq33 + frq24) " << frq22 +
frq23 + frq33 + frq24
                << "\n(frq22 + frq23 + frq33 + frq24 + frq34) " <<
frq22 + frq23 + frq33 + frq24 + frq34
                << "\n(frq22 + frq23 + frq33 + frq24 + frq34 +
frq44) " << frq22 + frq23 + frq33 + frq24 + frq34 + frq44
                << "\n";
      }

           //Diet group selected
      if (random_diet_temp <= 0.00) {
           diet = 0;
           }
```

```
  else if (random_diet_temp <= 0.200) {
    diet = 1;
        }
else if (random_diet_temp <= 0.679) {
    diet = 2;
    }
else if (random_diet_temp <= 0.715) {
    diet = 3;
    }
  else if (random_diet_temp <= 0.918) {
    diet = 4;
    }
else if (random_diet_temp <= 1.00) {
    diet = 5;
    }
else {
        cout << "error in diet selection" << "\n";
  }
  //diet = 5;

//Alcohol consumption group selected
  if (random_alcohol_temp <= 0.33) {
    alcohol = 1;
    }
else if (random_alcohol_temp <= 0.52) {
    alcohol = 2;
    }
else if (random_alcohol_temp <= 0.66) {
    alcohol = 3;
    }
else if (random_alcohol_temp <= 0.76) {
    alcohol = 4;
    }
else if (random_alcohol_temp <= 0.86) {
    alcohol = 5;
    }
  else if (random_alcohol_temp <= 0.90) {
    alcohol = 6;
    }
  else if (random_alcohol_temp <= 1.00) {
    alcohol = 7;
    }
  else {
        cout << "error in alcohol selection" << "\n";
  }
  //alcohol = 1;

//Smoking group selected
  if (random_smoking_temp <= 0.44) {
    smoking = 1;
    }
else if (random_smoking_temp <= 0.78){
    smoking = 2;
    }
  else if (random_smoking_temp <= 0.85){
    smoking = 3;
    }
  else if (random_smoking_temp <= 1.0){
    smoking = 4;
    }
  else {
        cout << "error in smoking selection" << "\n";
  }
  //smoking = 4;

//Physical activity group selected
  if (random_exercise_temp <= 0.00) {
    exercise = 1;
    }
else if (random_exercise_temp <= 0.19) {
    exercise = 2;
    }
else if (random_exercise_temp <= 0.40) {
    exercise = 3;
    }
  else if (random_exercise_temp <= 0.60){
    exercise = 4;
    }
  else if (random_exercise_temp <= 0.77) {
    exercise = 5;
    }
  else if (random_exercise_temp <= 1.00) {
    exercise = 6;
    }
  else {
        cout << "error in exercise selection" << "\n";
  }
  //exercise = 6;

//Sex assosiated risk
  if (sex == 0) {
    sex_risk = 0.58;
    }
else if (sex == 1) {
    sex_risk = 1.0;
    }
else {
    cout << "fail to allocate sex" << "\n";
    }

//Apo E x Diet risk factors
  if (male_genotype == 33) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.9842;
        }
    else if (diet == 2) {
        diet_risk = 1.1042;
        }
    else if (diet == 3) {
        diet_risk = 1.3042;
        }
    else if (diet == 4) {
        diet_risk = 1.5842;
        }
    else {
        diet_risk = 1.9442;
        }
    }
else if (male_genotype == 34) {
```

```
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.98725;
        }
    else if (diet == 2) {
        diet_risk = 1.1164;
        }
    else if (diet == 3) {
        diet_risk = 1.33165;
        }
    else if (diet == 4) {
        diet_risk = 1.633;
        }
    else {
        diet_risk = 2.02046;
        }
    }
else if (male_genotype == 23) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.981499;
        }
    else if (diet == 2) {
        diet_risk = 1.09339;
        }
    else if (diet == 3) {
        diet_risk = 1.27989;
        }
    else if (diet == 4) {
        diet_risk = 1.54098;
        }
    else {
        diet_risk = 1.87666;
        }
    }
else if (male_genotype == 44) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.990388;
        }
    else if (diet == 2) {
        diet_risk = 1.12895;
        }
    else if (diet == 3) {
        diet_risk = 1.35989;
        }
    else if (diet == 4) {
        diet_risk = 1.6832;
        }
    else {
        diet_risk = 2.09889;
        }
    }
else if (male_genotype == 24) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.9842;
        }
    else if (diet == 2) {
        diet_risk = 1.1042;
        }
    else if (diet == 3) {
        diet_risk = 1.3042;
        }
    else if (diet == 4) {
        diet_risk = 1.5842;
        }
    else {
        diet_risk = 1.9442;
        }
    }
else if (male_genotype == 22) {
    if (diet == 0) {
        diet_risk = 0.9442;
        }
    else if (diet == 1) {
        diet_risk = 0.978866;
        }
    else if (diet == 2) {
        diet_risk = 1.08287;
        }
    else if (diet == 3) {
        diet_risk = 1.2562;
        }
    else if (diet == 4) {
        diet_risk = 1.49886;
        }
    else {
            diet_risk = 1.81086;
        }
    }
else {
    cout << "error in genotype-diet selection" << "\n";
    }


  //Apo E x Alcohol risk factors
  if (male_genotype == 33) {
    if (alcohol == 1) {
        alcohol_risk = 1.7119;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.3164;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.0791;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.00;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.0791;
        }
        else if (alcohol == 6) {
```

```
        alcohol_risk = 1.3164;
        }
    else {
        alcohol_risk = 1.7119;
        }
    }
else if (male_genotype == 34) {
    if (alcohol == 1) {
        alcohol_risk = 1.7631;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.33915;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.08479;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.00;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.08479;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.33915;
        }
    else {
        alcohol_risk = 1.7631;
        }
    }
else if (male_genotype == 23) {
    if (alcohol == 1) {
        alcohol_risk = 1.6631;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.29614;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.07403;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.0;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.07403;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.29614;
        }
    else {
        alcohol_risk = 1.66631;
        }
    }
else if (male_genotype == 44) {
    if (alcohol == 1) {
        alcohol_risk = 1.81546;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.36243;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.09061;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.0;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.09061;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.36243;
        }
    else {
        alcohol_risk = 1.81546;
        }
    }
else if (male_genotype == 24) {
    if (alcohol == 1) {
        alcohol_risk = 1.7119;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.3164;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.0791;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.00;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.0791;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.3164;
        }
    else {
        alcohol_risk = 1.7119;
        }
    }
else if (male_genotype == 22) {
    if (alcohol == 1) {
        alcohol_risk = 1.62166;
        }
    else if (alcohol == 2) {
        alcohol_risk = 1.27629;
        }
    else if (alcohol == 3) {
        alcohol_risk = 1.06907;
        }
    else if (alcohol == 4) {
        alcohol_risk = 1.00;
        }
        else if (alcohol == 5) {
        alcohol_risk = 1.06907;
        }
        else if (alcohol == 6) {
        alcohol_risk = 1.27629;
        }
    else {
        alcohol_risk = 1.62166;
        }
    }
else {
```

```
    cout << "error in genotype-alcohol selection" << "\n";
    }


  //Apo E x smoking risk factors
  if (male_genotype == 33) {
    if (smoking == 1) {
        smoking_risk = 0.8055;
        }
    else if (smoking == 2) {
        smoking_risk = 1.7202;
        }
    else if (smoking == 3) {
        smoking_risk = 3.2447;
        }
    else {
        smoking_risk = 5.379;
        }
    }
else if (male_genotype == 34) {
    if (smoking == 1) {
        smoking_risk = 0.816793;
        }
    else if (smoking == 2) {
        smoking_risk = 1.76537;
        }
    else if (smoking == 3) {
        smoking_risk = 3.34634;
        }
    else {
        smoking_risk = 5.55969;
        }
    }
else if (male_genotype == 23) {
    if (smoking == 1) {
        smoking_risk = 0.795492;
        }
    else if (smoking == 2) {
        smoking_risk = 1.68017;
        }
    else if (smoking == 3) {
        smoking_risk = 3.15463;
        }
    else {
        smoking_risk = 5.21888;
        }
    }
else if (male_genotype == 44) {
    if (smoking == 1) {
        smoking_risk = 0.828403;
        }
    else if (smoking == 2) {
        smoking_risk = 1.81181;
        }
    else if (smoking == 3) {
        smoking_risk = 3.45083;
        }
    else {
        smoking_risk = 5.74545;
        }
    }
else if (male_genotype == 24) {
    if (smoking == 1) {
        smoking_risk = 0.8055;
        }
    else if (smoking == 2) {
        smoking_risk = 1.7202;
        }
    else if (smoking == 3) {
        smoking_risk = 3.2447;
        }
    else {
        smoking_risk = 5.379;
        }
    }
else if (male_genotype == 22) {
    if (smoking == 1) {
        smoking_risk = 0.785735;
        }
    else if (smoking == 2) {
        smoking_risk = 1.64114;
        }
    else if (smoking == 3) {
        smoking_risk = 3.06682;
        }
    else {
        smoking_risk = 5.06277;
        }
    }
else {
    cout << "error in genotype-smoking selection" << "\n";
    }


  //Apo E x Exercise risk factors
  if (male_genotype == 33) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.0087;
        }
    else if (exercise == 2) {
        exercise_risk = 1.0582;
        }
    else if (exercise == 3) {
        exercise_risk = 1.1407;
        }
    else if (exercise == 4) {
        exercise_risk = 1.2562;
        }
    else {
        exercise_risk = 1.4047;
        }
    }
else if (male_genotype == 34) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.01114;
        }
```

```
    else if (exercise == 2) {
        exercise_risk = 1.06796;
        }
    else if (exercise == 3) {
        exercise_risk = 1.16265;
        }
    else if (exercise == 4) {
        exercise_risk = 1.29522;
        }
    else {
        exercise_risk = 1.46567;
        }
    }
else if (male_genotype == 23) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.00654;
        }
    else if (exercise == 2) {
        exercise_risk = 1.04958;
        }
    else if (exercise == 3) {
        exercise_risk = 1.1213;
        }
    else if (exercise == 4) {
        exercise_risk = 1.22172;
        }
    else {
        exercise_risk = 1.35082;
        }
    }
else if (male_genotype == 44) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.01365;
        }
    else if (exercise == 2) {
        exercise_risk = 1.07801;
        }
    else if (exercise == 3) {
        exercise_risk = 1.18527;
        }
    else if (exercise == 4) {
        exercise_risk = 1.33544;
        }
    else {
        exercise_risk = 1.52851;
        }
    }
else if (male_genotype == 24) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.0087;
        }
    else if (exercise == 2) {
        exercise_risk = 1.0582;
        }
    else if (exercise == 3) {
        exercise_risk = 1.1407;
        }
    else if (exercise == 4) {
        exercise_risk = 1.2562;
        }
    else {
        exercise_risk = 1.4047;
        }
    }
else if (male_genotype == 22) {
    if (exercise == 0) {
        exercise_risk = 0.9922;
        }
    else if (exercise == 1) {
        exercise_risk = 1.00445;
        }
    else if (exercise == 2) {
        exercise_risk = 1.0412;
        }
    else if (exercise == 3) {
        exercise_risk = 1.10244;
        }
    else if (exercise == 4) {
        exercise_risk = 1.18819;
        }
    else {
            exercise_risk = 1.29843;
        }
    }
else {
    cout << "error in genotype-exercise selection" << "\n";
    }


//Total risk
  total_risk = sex_risk * diet_risk * alcohol_risk *
               smoking_risk * exercise_risk;

  //Normal curve parameters calculation
  while (age1 <= 122) {
    ++age1;
    temp1 = age1 * temp_sum(age1, total_risk);
    temp2 += temp1;
    }

  mean_of_distribution = temp2;

while (age2 <= 122) {
    ++age2;
    temp3 = (age2 - mean_of_distribution)*
        (age2 - mean_of_distribution)*
        temp_sum(age2, total_risk);
    temp4 += temp3;
    }

  double sigma = sqrt(temp4);

  double betaGumble = sigma * sqrt(6.0) / 3.14159;
```

```
      double alphaGumble = mean_of_distribution + beta * 0.5772;


      double temp_ran_gumbelcustom = gsl_ran_gumbelcustom(r,
alphaGumble, betaGumble);

      double random_death_age = temp_ran_gumbelcustom;

      /*cout << "\nthe male genotype is " << male_genotype << "\n"
           << "total risk factor " << total_risk << "\n"
           << "the mean of the disribution " << mean_of_distribution
<< "\n"
           << "sigma is " << sigma << "\n"
           << "random death age is = " << random_death_age <<
"\n";*/

      return (random_death_age);
}



//Function to add an array
int sum_int_array(int pt[],int n)
{
      int temp = 0;
      for (int i = 0; i < n; ++i)
            temp += pt[i];
      return temp;
}



//Generating the number of offspring per couple
int number_of_offspring (void) {

      double fecundity[51] = {0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0,
        0.007800, 0.007800, 0.007800, 0.007800, 0.007800, 0.051700,
        0.051700, 0.051700, 0.051700, 0.051700, 0.128300, 0.128300,
        0.128300, 0.128300, 0.128300, 0.116700, 0.116700, 0.116700,
        0.116700, 0.116700, 0.043000, 0.043000, 0.043000, 0.043000,
        0.043000, 0.006700, 0.006700, 0.006700, 0.006700, 0.006700,
        0.000200, 0.000200, 0.000200, 0.000200, 0.000200, 0};

      double *pt_fecundity = 0;
      pt_fecundity = fecundity;

      int offspring[51];
      int *pt_offspring;
      pt_offspring = offspring;

      int female_fin_age = (int) female(frq22, frq23, frq33, frq24,
frq34, frq44);
     int male_fin_age = (int) male(frq22, frq23, frq33, frq24, frq34,
frq44);
//    int temp = min(50, min(female_fin_age, male_fin_age));

      int temp1;

      int temp;

      //Alternative to min function
      /*if (female_fin_age <= male_fin_age)
      {
            temp1=female_fin_age;
      }
      else
      {
            temp1=male_fin_age;
      }

      if (50 <= temp1)
      {
            temp=50;
      }
      else
      {
            temp=temp1;
      }*/


      temp = MIN(50, MIN(female_fin_age, male_fin_age));



      for (int life = 0; life <= temp; life++)
            *pt_offspring++ = gsl_ran_bernoulli(r, *pt_fecundity++);
      int total_no = sum_int_array(offspring, temp);

      /*cout << "\nfemale death age " << female_fin_age << "\n"
           << "male death age " << male_fin_age << "\n"
           << "number of offspring " << total_no << "\n";*/

      return (total_no);
}



std::string convertToString(int x)
{
   std::ostringstream o;
   if (o << x)
     return o.str();
   // some sort of error handling goes here...
   return "conversion error";
}



int convertFromString(const std::string& s)
 {
   std::istringstream i(s);
   int x;
   if (i >> x)
     return x;
   // some sort of error handling goes here...
   return 0;
 }
```

```
//Finding the genotype of the offspring
 int offspring_genotype (void) {

      int temp_female_genotype = female_genotype;
      int temp_male_genotype = male_genotype;


      std::string female_genotype_string = convertToString
            (temp_female_genotype);
      std::string male_genotype_string = convertToString
            (temp_male_genotype);

      char offspring_char[3] = {' ', ' ', '\0'};
      char *pt_offspring_char = offspring_char;

      int random_offspring_temp = gsl_rng_uniform_int(r, 100);
    int random_offspring = random_offspring_temp;

      if (random_offspring < 25) {
            pt_offspring_char[0] = female_genotype_string[0];
            pt_offspring_char[1] = male_genotype_string[0];
      }
      else if (random_offspring < 50) {
            pt_offspring_char[0] = female_genotype_string[0];
            pt_offspring_char[1] = male_genotype_string[1];
      }
      else if (random_offspring < 75) {
            pt_offspring_char[0] = female_genotype_string[1];
            pt_offspring_char[1] = male_genotype_string[0];
      }
      else if (random_offspring < 100) {
            pt_offspring_char[0] = female_genotype_string[1];
            pt_offspring_char[1] = male_genotype_string[1];
      }
      else {
            cout << "error in offspring char" << "\n";
      }

      int temp_part1 = offspring_char[0];
      int temp_part2 = offspring_char[1];

      if (temp_part1 > temp_part2) {
            char temp_string[3];
            char *pt_temp_string = temp_string;
            pt_temp_string[0] = pt_offspring_char[0];
            pt_offspring_char[0] = pt_offspring_char[1];
            pt_offspring_char[1] = pt_temp_string[0];
      }
      //else {
            //cout << "a " << "\n";
      //}


      int fin_offspring_genotype = atoi(offspring_char);

      /*cout << "\n\nthe genotype of the offspring is " <<
offspring_char <<
            "\n";*/

      return(fin_offspring_genotype);
}



static int offspring_genotype_array[6];

int offspring_array (int offspring_gen){

      if (offspring_gen == 22) {
            offspring_genotype_array[0] += 1;
      }
      else if (offspring_gen == 23) {
            offspring_genotype_array[1] += 1;
      }
      else if (offspring_gen == 33) {
            offspring_genotype_array[2] += 1;
      }
      else if (offspring_gen == 24) {
            offspring_genotype_array[3] += 1;
      }
      else if (offspring_gen == 34) {
            offspring_genotype_array[4] += 1;
      }
      else if (offspring_gen == 44) {
            offspring_genotype_array[5] += 1;
      }
      else {
            cout << "error in offspring_genotype_array" << "\n";
      }

      /*for (int z = 0; z < 6; z++) {
    cout << offspring_genotype_array[z] << " ";
      }

      cout << "\n-----------------------------------------------
\n";*/

      return(0);
}


int temp_offspring_array;

void entire_population (double population) {

      int temp_number_of_offspring;

      for (int pop = 1; pop <= population; pop++) {
            temp_number_of_offspring = number_of_offspring();
        for (int f = 1; f <= temp_number_of_offspring; f++)
                offspring_array (offspring_genotype());}
}


int main()
{
```

```
    gsl_rng_env_setup();

    T = gsl_rng_default;
    r = gsl_rng_alloc(T);

    seed = time(NULL) * getpid();
    gsl_rng_set(r, seed);

    char *buffer;
    int decimal, sign;
    double file_random = gsl_rng_uniform(r);
    buffer = fcvt (file_random, 5, &decimal, &sign );
    char file_name[39];
    strcpy (file_name,"fotios");
    strcat (file_name, buffer);
    strcat (file_name,".nb");


    //Opening the output file
    ofstream my_file(file_name);

    if (!my_file) {
          cerr << "Failed to open my_file.\n";
          exit(EXIT_FAILURE);
    }

    //Initial Population and frequencies
    int population = 200000;
    frq22 = 0.017;
    frq23 = 0.116;
    frq33 = 0.558;
    frq24 = 0.019;
    frq34 = 0.251;
    frq44 = 0.039;
    freq2 = frq22 + (frq23/2) + (frq24/2);
    freq3 = (frq23/2) + frq33 + (frq34/2);
    freq4 = (frq24/2) + (frq34/2) + frq44;

    //~74sec for population of 1000 for 10 generations
    clock_t start_time = clock();

    //Loop and number of generations
    for (int gen = 1; gen <= 50; ++gen) {

          /*cout << "--------------- GENERATION "<< gen << " ------
---------------\n"
               << "frequency 2 is " << freq2 << "\n"
               << "frequency 3 is " << freq3 << "\n"
               << "frequency 4 is " << freq4 << "\n"
               << "frq22 is " << frq22 << "\n"
               << "frq23 is " << frq23 << "\n"
               << "frq33 is " << frq33 << "\n"
               << "frq24 is " << frq24 << "\n"
               << "frq34 is " << frq34 << "\n"
               << "frq44 is " << frq44 << "\n";*/

          entire_population(population/(2 * 0.88));
          //0.886

       my_file << "generation[" << gen << "] = {" <<
            freq2 << " ," << freq3 << " ," << freq4 <<
            "};" << "\n";

       my_file << "OffspringGenotypes[" << gen << "]={" <<
            offspring_genotype_array[0] <<
            ", " <<
            offspring_genotype_array[1] <<
            ", " <<
            offspring_genotype_array[2] <<
            ", " <<
            offspring_genotype_array[3] <<
            ", " <<
            offspring_genotype_array[4] <<
            ", " <<
            offspring_genotype_array[5] << "};\n";


          population = offspring_genotype_array[0] +
            offspring_genotype_array[1] +
offspring_genotype_array[2] +
            offspring_genotype_array[3] +
offspring_genotype_array[4] +
            offspring_genotype_array[5];

          if (population == 0) {
            cout << "population extinct" << "\n";
            exit(EXIT_FAILURE);
          }
          //else {
          //    cout << " b"<< "\n";
          //}
          double tempbottom = 2 * population;

          double tempfreq2top = 2 * offspring_genotype_array[0] +
          offspring_genotype_array[1] +
            offspring_genotype_array[3];
          freq2 = tempfreq2top / tempbottom;

          double tempfreq3top = 2 * offspring_genotype_array[2] +
          offspring_genotype_array[1] +
            offspring_genotype_array[4];
          freq3 = tempfreq3top / tempbottom;

          double tempfreq4top = 2 * offspring_genotype_array[5] +
          offspring_genotype_array[3] +
            offspring_genotype_array[4];
          freq4 = tempfreq4top / tempbottom;

          double temp_population = population;

          frq22 = offspring_genotype_array[0] / temp_population;
        frq23 = offspring_genotype_array[1] / temp_population;
        frq33 = offspring_genotype_array[2] / temp_population;
        frq24 = offspring_genotype_array[3] / temp_population;
        frq34 = offspring_genotype_array[4] / temp_population;
        frq44 = offspring_genotype_array[5] / temp_population;


          offspring_genotype_array[0] = 0;
          offspring_genotype_array[1] = 0;
```

```
            offspring_genotype_array[2] = 0;
            offspring_genotype_array[3] = 0;
            offspring_genotype_array[4] = 0;
            offspring_genotype_array[5] = 0;

    }

      my_file.close();

      clock_t stop_time = clock();

      cout << "Time taken = " <<
            static_cast<double>(stop_time - start_time) /
            CLOCKS_PER_SEC << " secs.\n"
            << "frequency 2 is " << freq2 << "\n"
            << "frequency 3 is " << freq3 << "\n"
            << "frequency 4 is " << freq4 << "\n"
            << "frq22 is " << frq22 << "\n"
            << "frq23 is " << frq23 << "\n"
            << "frq33 is " << frq33 << "\n"
            << "frq24 is " << frq24 << "\n"
            << "frq34 is " << frq34 << "\n"
            << "frq44 is " << frq44 << "\n";


      return (EXIT_SUCCESS);
}
```

Appendix 9. *Mathematica* notebook for the calculation of the random walk change per generation

Loading Packages

```
<< Statistics'DescriptiveStatistics'
<< Graphics'MultipleListPlot'
<< Statistics'ConfidenceIntervals'
<< Statistics'DataManipulation'
```

Setting Directory

```
Directory[]
```

C:\Program Files\Wolfram Research\Mathematica\4.0

```
SetDirectory["C:\backed up\Mathematica\ApoE results/1st_run"]
```

C:\backed up\Mathematica\ApoE results\1st_run

```
Directory[]
```

C:\backed up\Mathematica\ApoE results\1st_run

```
ListOfFiles = FileNames["fotios*.nb"];

NumberOfFiles = Length[ListOfFiles];
```

Number Of Generations

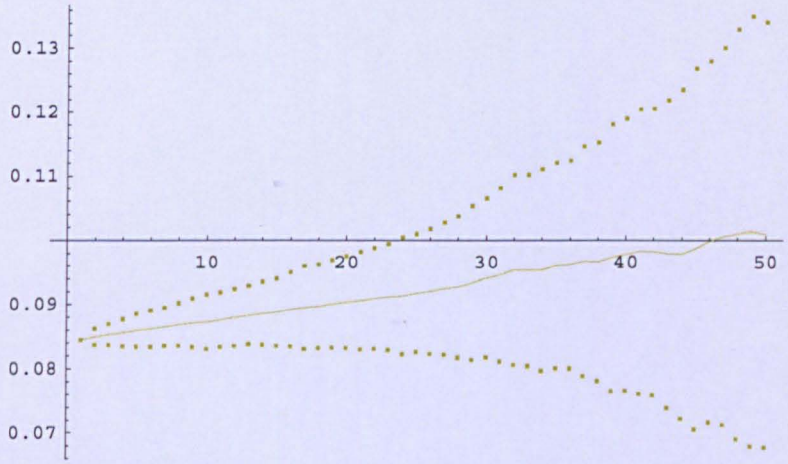
```
gen = 50;
```

Modules for Alleles Frequencies and Random Walk

```
TotalSampleRun[generationNumber_?NumberQ, alleleNumber_?NumberQ] :=
  Module[{ListOfFiles = FileNames["fotios*.nb"], NumberOfFiles = Length[ListOfFiles]},

    ListTest = Table[Get[ListOfFiles[[i]]];
        generation[generationNumber][[alleleNumber - 1]], {i, 1, NumberOfFiles}]
  ]

RandomWalk[alleleNumber_?NumberQ] :=
  Module[{ListOfFiles = FileNames["fotios*.nb"]},

    RW = Flatten[Table[{Table[Get[ListOfFiles[[SetOfValues]]]; generation[i + 1][[alleleNumber - 1]] -
            generation[i][[alleleNumber - 1]], {i, 1, 49}]}, {SetOfValues, 1, 60}]]
  ]
```

For Allele ϵ2

```
alleleϵ2Mean = Table[Mean[TotalSampleRun[i, 2]], {i, 1, gen}];

alleleϵ2SD = Table[StandardDeviation[TotalSampleRun[i, 2]], {i, 1, gen}];

alleleϵ2SDabove = Table[alleleϵ2Mean[[i]] + 2 × alleleϵ2SD[[i]], {i, 1, gen}];

alleleϵ2SDbelow = Table[alleleϵ2Mean[[i]] − 2 × alleleϵ2SD[[i]], {i, 1, gen}];

alleleϵ2Plot = MultipleListPlot[alleleϵ2Mean, alleleϵ2SDabove, alleleϵ2SDbelow,
    PlotJoined → {True, Automatic, Automatic}, SymbolShape → {None, PlotSymbol[Box, 1], PlotSymbol[Box, 1]},
    SymbolStyle → {RGBColor[0.8, 0.8, 0], RGBColor[0.8, 0.8, 0]}, PlotStyle → RGBColor[0.8, 0.8, 0]]
```




```
- Graphics -
```

Random Walk Analysis

Confidence Intervals

```
RWe2 = RandomWalk[2]

RWe2Mean = Mean[RWe2]
```

0.000335217

```
RWe2SE = StandardDeviation[RWe2]
```

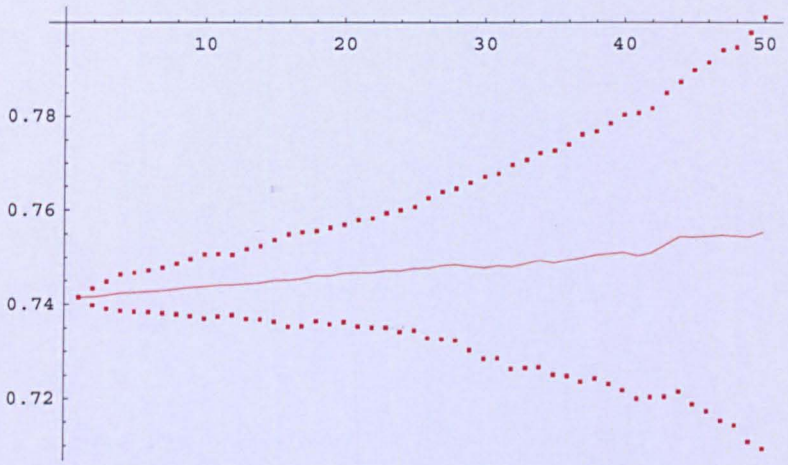
0.00193138

$$\text{RWe2CI} = \left\{\text{RWe2Mean} - \frac{\text{RWe2SE}}{\sqrt{\text{gen} \times \text{NumberOfFiles}}} \times 1.96,\ \text{RWe2Mean} + \frac{\text{RWe2SE}}{\sqrt{\text{gen} \times \text{NumberOfFiles}}} \times 1.96\right\}$$

{0.000266103, 0.00040433}

For Allele ϵ3

```
alleleϵ3Mean = Table[Mean[TotalSampleRun[i, 3]], {i, 1, gen}];

alleleϵ3SD = Table[StandardDeviation[TotalSampleRun[i, 3]], {i, 1, gen}];

alleleϵ3SDabove = Table[alleleϵ3Mean[[i]] + 2 × alleleϵ3SD[[i]], {i, 1, gen}];

alleleϵ3SDbelow = Table[alleleϵ3Mean[[i]] − 2 × alleleϵ3SD[[i]], {i, 1, gen}];

alleleϵ3Plot = MultipleListPlot[alleleϵ3Mean, alleleϵ3SDabove, alleleϵ3SDbelow,
    PlotJoined → {True, Automatic, Automatic}, SymbolShape → {None, PlotSymbol[Box, 1], PlotSymbol[Box, 1]},
    SymbolStyle → {RGBColor[0.9, 0, 0], RGBColor[0.9, 0, 0]}, PlotStyle → RGBColor[0.9, 0, 0]]
```




```
- Graphics -
```

Random Walk Analysis

Confidence Intervals

```
RWe3 = RandomWalk[3]

RWe3Mean = Mean[RWe3]
```

0.000278457

```
RWe3SE = StandardDeviation[RWe3]
```

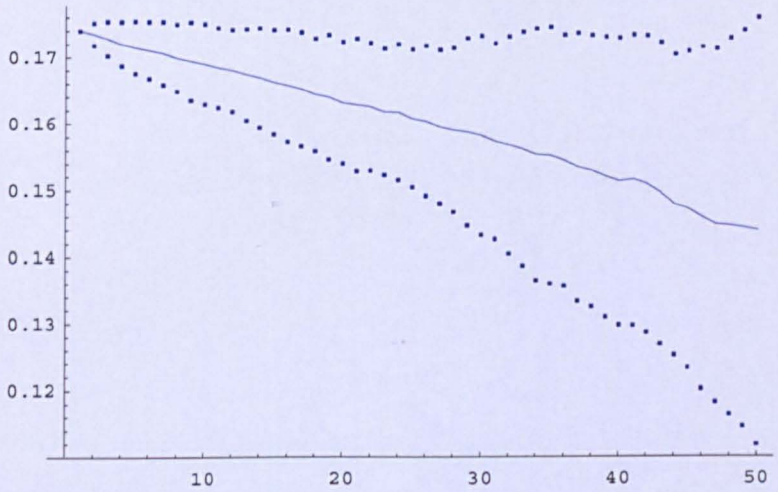
0.00272725

$$\text{RWe3CI} = \left\{\text{RWe3Mean} - \frac{\text{RWe3SE}}{\sqrt{\text{gen} \times \text{NumberOfFiles}}} \times 1.96,\ \text{RWe3Mean} + \frac{\text{RWe3SE}}{\sqrt{\text{gen} \times \text{NumberOfFiles}}} \times 1.96\right\}$$

{0.000180864, 0.000376051}

For Allele ϵ4

```
alleleϵ4Mean = Table[Mean[TotalSampleRun[i, 4]], {i, 1, gen}];

alleleϵ4SD = Table[StandardDeviation[TotalSampleRun[i, 4]], {i, 1, gen}];

alleleϵ4SDabove = Table[alleleϵ4Mean[[i]] + 2 × alleleϵ4SD[[i]], {i, 1, gen}];

alleleϵ4SDbelow = Table[alleleϵ4Mean[[i]] − 2 × alleleϵ4SD[[i]], {i, 1, gen}];

alleleϵ4Plot = MultipleListPlot[alleleϵ4Mean, alleleϵ4SDabove, alleleϵ4SDbelow,
    PlotJoined → {True, Automatic, Automatic}, SymbolShape → {None, PlotSymbol[Box, 1], PlotSymbol[Box, 1]},
    SymbolStyle → {RGBColor[0, 0, 0.9], RGBColor[0, 0, 0.9]}, PlotStyle → RGBColor[0, 0, 0.9]]
```




```
- Graphics -
```

Random Walk Analysis

Confidence Intervals

```
RWe4 = RandomWalk[4]

RWe4Mean = Mean[RWe4]
```

−0.000613672

```
RWe4SE = StandardDeviation[RWe4]
```

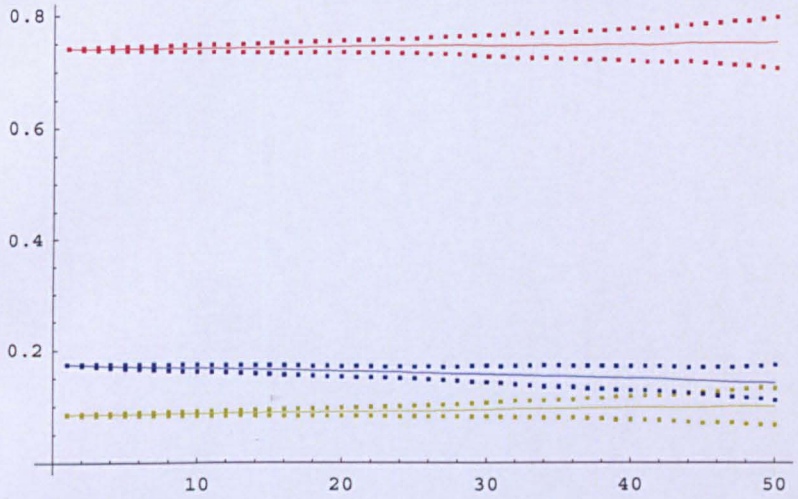
0.00226254

$$\text{RWe4CI} = \left\{\text{RWe4Mean} - \frac{\text{RWe4SE}}{\sqrt{\text{gen} \times \text{NumberOfFiles}}} \times 1.96,\ \text{RWe4Mean} + \frac{\text{RWe4SE}}{\sqrt{\text{gen} \times \text{NumberOfFiles}}} \times 1.96\right\}$$

{−0.000694636, −0.000532708}

Graph representing the total allele frequencies change

```
Show[alleleϵ2Plot, alleleϵ3Plot, alleleϵ4Plot]
Print["The change ϵ2 per generation is –> ", RWe2Mean, " with 95% CI –> ", RWe2CI ]
Print["The change ϵ3 per generation is –> ", RWe3Mean, " with 95% CI –> ", RWe3CI ]
Print["The change ϵ4 per generation is –> ", RWe4Mean, " with 95% CI –> ", RWe4CI ]
```




```
- Graphics -
```

```
The change ϵ2 per generation is -> RWe2Mean with 95% CI -> RWe2CI

The change ϵ3 per generation is -> RWe3Mean with 95% CI -> RWe3CI

The change ϵ4 per generation is -> RWe4Mean with 95% CI -> RWe4CI
```