Chapter 8

Evolutionary Biology of Aging: Future Directions

Daniel E. L. Promislow, Kenneth M. Fedorka, and Joep M. S. Burger

I. Introduction

Over the past two decades, we have seen extraordinary progress in evolutionary studies of senescence. Beginning with the early quantitative genetic tests of theories of senescence (Edney & Gill, 1968; Luckinbill et al., 1984; Rose & Charlesworth, 1980), we have progressed to the point where evolutionary biologists rely on state-of-the-art molecular tools, and the ties between evolutionary and molecular approaches in the study of senescence are often seamless. More than perhaps any other research area in evolutionary ecology, molecular biologists working on senescence appreciate how evolutionary biology contributes to our understanding of senescence. Many will tell you about George Williams’ antagonistic pleiotropy theory of senescence (Williams, 1957), or perhaps Tom Kirkwood’s disposable soma theory (Kirkwood, 1977), and the especially well-rounded researcher might mention that molecular biologists have identified individual genes that appear to exhibit antagonistic pleiotropic effects (Campisi, 2003).

Serious effort has been expended in testing these classic evolutionary theories of senescence (e.g., Hughes & Charlesworth, 1994; Promislow et al., 1996; Rose & Charlesworth, 1980). The field is now moving beyond tests of existing hypotheses, but what exactly is the future direction for the evolutionary studies of aging? We do not have a scientific crystal ball, and making definite predictions places us somewhere between hubris and folly. Nevertheless, with this caveat in mind, we hope that the ideas presented here may spur the next generation of biogerontologists to consider evolutionary studies of senescence.

What differentiates evolutionary studies of aging from studies that take a strictly molecular, physiological, or demographic approach? One distinction is that non-evolutionary biologists frequently ask proximate questions (e.g., how do specific biological processes change with age, and how do genes affect these changes?), whereas evolutionary biologists are
interested in ultimate questions (why has aging evolved, and why have particular genes evolved to influence longevity?). The molecular biologist is trying to identify the mechanisms that cause aging. The evolutionary biologist is trying to place those mechanisms in a broader perspective, asking why those mechanisms and not others are important to aging, and how those mechanisms are shaped by other forces—mutation, selection, genetic drift—acting over evolutionary time.

In the past few years, the lines between these two fields have begun to blur. Molecular geneticists working on aging have used evolutionary theory to motivate studies of single genes (Campisi, 2003; Walker et al., 2000), and evolutionary biologists have embraced techniques that are firmly rooted in modern molecular biology (e.g., Pletcher et al., 2002; Tatar et al., 2001).

Although evolutionary biologists now rely on 21st-century techniques, the greatest evolutionary contributions to aging studies date back half a century. Until the middle of the 20th century, the standard argument for the evolution of aging was that it was good for the species (Weismann, 1891). In the 1940s, we began to move beyond that argument, which we now recognize as fallacious. Following from the insights of Fisher (1930) and Haldane (1941), in 1946, Medawar argued that because the strength of selection declines with age, senescence will arise as an inevitable consequence of this decline (Medawar, 1946). Consider a novel, germ-line mutation that reduces survival at just one age. If it reduces survival before the age at maturity, then the probability of surviving to produce offspring at any age is uniformly reduced. Such a deleterious mutation would experience strong negative selection and eventually be lost from the population. In contrast, a mutation that reduces survival at some late age, after most individuals in the population had died, would experience only very weak selection and could spread through random genetic drift. Over evolutionary time, early-acting deleterious mutations will continually be removed by selection, whereas late-acting ones will accumulate. According to Medawar’s mutation accumulation theory, it is inevitable that we carry a relatively high load of late-acting deleterious mutations inherited from our ancestors. As we age, we experience the effects of these mutations, which cause a decrease in rates of survival and/or fertility.

A decade later, George Williams developed his antagonistic pleiotropy theory, in which he argued that senescence would arise if late-acting deleterious mutations were actually favored by selection due to their early-acting beneficial effects (Williams, 1957). In this case, senescence evolves due to tradeoffs between early-age benefits and late-age costs, an idea that was further developed by Tom Kirkwood in his disposable soma theory (Kirkwood, 1977). A half-century after Medawar and Williams, evolutionary biologists are still trying to determine which of these theories provides the best explanation for senescence (Charlesworth, 2001; Charlesworth & Hughes, 1996; Hughes et al., 2002; Partridge & Gems, 2002a; Snoke & Promislow, 2003). More recently, molecular gerontologists have begun to embrace these ideas, with a particular interest in finding genes with antagonistic pleiotropic effects (Campisi, 2003; Walker et al., 2000).

The current focus in the evolutionary biology of aging encompasses three main areas:

1. Ongoing studies are trying to characterize the genetic architecture of aging (see Chapter 7, Mackay et al.), asking not only which model can explain genetic variation for senescence, but also whether genes that have been found to extend life span in the lab (so called “aging genes”) show allelic variation that
correlates with longevity in wild-caught isolates, and whether the effect of aging genes depends on the sex or external environment in which they are expressed (Fry et al., 1998; Leips & Mackay, 2000; Nuzhdin et al., 1997).

2. A second body of work has focused specifically on the shape of mortality trajectories (see Chapter 1, Gavrilo and Gavrilova), asking why mortality curves increase exponentially with age (Abrams & Ludwig, 1995; Charlesworth, 2001) and why mortality rates appear to decelerate late in life in some cases (Mueller et al., 2003; Mueller & Rose, 1996; Service, 2000; Vaupel et al., 1998, but see Finch & Pike, 1996; Linnen et al., 2001).

3. Finally, many evolutionary biologists have become interested in the central role that the endocrine system may play in determining the evolution of the suite of traits that make up an individual’s “life history strategy,” including development time, age at maturity, growth rate, body size, fecundity, and, of course, life span (Tatar et al., 2003). The evolution of aging in general, and these three subjects in particular, have all been reviewed recently in other sources (e.g., Promislow & Bronikowski, in press; Tatar et al., 2003) and in this book (see Chapter 15, Tu et al.; Chapter 19, Miller and Austad; Chapter 20, Carter and Sonntag). Accordingly, rather than going over well-tilled ground, in the rest of this chapter, we will look at five areas that are less studied at present but which we think may provide fertile soil for the growth of future evolutionary studies of aging. These include (1) molecular evolution and gene networks; (2) the intersection of physiology and demography; (3) parasites and immunity; (4) sexual selection and sexual conflict; and (5) genetic variation in natural populations.

The rationale for focusing on these particular areas comes from our thinking about early models of aging. The first mathematical model for the evolution of senescence was developed by W. D. Hamilton (1966). Hamilton started with the well-known Euler-Lotka equation (Euler, 1760; Lotka, 1925):

\[ 1 = \sum e^{-rx}l(x|m(x)) \]  

in which \( r \) is the intrinsic rate of increase in a population (also called the Malthusian parameter, and used as a measure of Darwinian fitness), \( l(x) \) is the probability of surviving from birth to age \( x \), and \( m(x) \) is the number of daughters that a female produces at age \( x \). Hamilton’s model used this equation to provide exact descriptions of how the strength of selection acting on rates of mortality or fecundity would change with age. One can think of the strength of selection as a measure of how fast a new mutation will be fixed or lost in a population. In particular, the strength of selection acting on \( P(x) \), the survival rate from age \( x \) to \( x + 1 \), is given by

\[ \frac{\partial r}{\partial \ln p(x)} = \frac{\sum_{y=x+1}^{\infty} e^{-ry}l(y)m(y)}{\sum xe^{-rx}l(x|m(x))} \]  

The strength of selection acting on fecundity is given by

\[ \frac{\partial r}{\partial m(x)} = \frac{e^{-rx}l(x)}{\sum xe^{-rx}l(x|m(x))} \]  

The important point that these equations illustrate is that the demographic parameters themselves determine the rate at which selection declines with age. We show how the strength of selection declines for one particular set of values for age-specific survival and fecundity in Figure 8.1. These equations do not include information about such factors as social behavior (e.g., mate choice, parent-offspring conflict), host-parasite interactions, variation in the environment, physiology, or the underlying
genetic architecture of these demographic traits. All of these factors could alter the shape of the curves described by equations (2) and (3).

Researchers interested in the evolution of aging have begun to develop ways to extend Hamilton’s equations, enhancing both biological realism and predictive power. For example, Peter Abrams showed that if one adds density dependence to classic models for the evolution of aging, in some cases extrinsic mortality rates no longer influence the evolution of rates of senescence (Abrams, 1993). Ron Lee has developed models that incorporate interactions between parents and their offspring into models of aging; these models make impressively accurate predictions about the shape of human mortality curves (Lee, 2003). And most recently, Jim Vaupel and colleagues have created models to demonstrate that species with indeterminate growth (i.e., no asymptotic size as adults) may evolve “negative senescence,” where mortality rates actually decline with age (Vaupel et al., 2004). In the following sections, we explore a range of biological phenomena that may allow us to further extend classic models of aging. Clearly there is much work to be done, and we are optimistic that as we unite new molecular tools and new evolutionary ideas, the coming years will bring a more comprehensive understanding of the evolution of senescence.

II. Genetics of Senescence

For almost 20 years, evolutionary biologists and molecular biologists working on the biology of aging appeared to be working on utterly different problems, with little communication between the two groups (perhaps not unlike the gap between those working on cellular senescence and those working on aging in animal models; Campisi, 2001). This has changed in the past few years, with evolutionary biologists now embracing the newest molecular techniques, and molecular biologists starting to test evolutionary hypotheses. For example, a collaborative effort among evolutionary and molecular biologists led to the first large-scale microarray analysis of patterns of gene transcription associated with senescence in the entire fly genome (Pletcher et al., 2002). The sequencing of entire genomes has made it extremely easy to identify genes associated with aging that occur across taxonomically diverse organisms (and so, by inference, to identify aging processes that have deep evolutionary roots). George Martin wondered whether genes associated with aging were likely to be specific to each species.
private mechanisms) or constant across evolutionary time (public mechanisms) (Martin, 1997). At least two gene pathways associated with aging—sirtuins and insulin signaling—appear to be public mechanisms (Partridge & Gems, 2002b; Tissenbaum & Guarente, 2001).

Current work on the molecular genetics of aging is focused on characterizing the function of the genes and gene pathways that have already been identified, and on finding new genes (Partridge & Gems, 2002b; Tissenbaum & Guarente, 2001). Clearly, this is where the action is. So where does evolutionary biology fit within this decidedly molecular enterprise? In the introduction, we discussed how the strength of selection changes with age. One challenge is to determine how the strength of selection acts not only on demographic traits, but also on the individual genes that shape those demographic traits. If we can do this, we may be able to develop an evolutionary genetic model of aging that allows us to actually predict what kinds of genes should be associated with aging. A more refined analysis of selection at the level of individual genes and genomes will come about in two ways—first, by combining molecular evolutionary studies of gene sequences or genome structure with analyses of life span, and second, through the study of gene- and protein-interaction networks.

**A. Molecular Evolution and Aging**

Molecular evolutionists study the process of evolution by analyzing variation in DNA and protein sequences within and among species. Many workers in the field have focused on the historical and current patterns of selection that act on genes—an area of inquiry that began in earnest with the debate over whether most genetic variation was due to selection or had occurred in the absence of selection due to the effects of random drift of neutral or nearly neutral mutations (Kimura, 1968). In some ways, the history of the neutralist–selectionist debate in molecular evolution mirrors that of the mutation accumulation–antagonistic pleiotropy debate in senescence. After a long argument between neutralists and selectionists, most evolutionary biologists now accept that some genes have evolved under a neutral scenario whereas others have evolved due primarily to selective forces. Nevertheless, the debate started by Kimura fueled decades of exciting and productive research. Similarly, in the world of evolutionary gerontology, it will likely turn out that some genes have evolved due to mutation accumulation and others due to antagonistic pleiotropy, but the debate over which is the better explanatory hypothesis has sparked invaluable research progress.

However, we see the parallel between molecular evolution and aging research as having far more than just historical interest. Molecular evolution has led to fundamental advances in the way that we understand the biology of organisms, and some of these advances could inform future studies in the biology of aging.

Work on the genomes of a diversity of organisms has found that substantial portions of the genome are often made up of transposable elements (TEs). These “jumping genes” can function as parasites in the genomes of their host and can lead to substantial increases in background mutation rates (McDonald, 1993). To the extent that aging is affected by age-related increases in somatic mutation rates, TEs may turn out to play an important role in the aging process. Surprisingly, relatively few studies have examined the relationship between TEs and aging (Nikitin & Shmookler Reis, 1997; Woodruff & Nikitin, 1995). One study makes the interesting (and cautionary) point that when genes associated with longevity in *Drosophila* are identified by knocking the gene out with a P-element (a type of TE found in flies), the P-element insertion alone may influence life span, independent of the effects of the target
gene (Kaiser et al., 1997). It would be useful to quantify the extent to which age-related accumulation of somatic mutations accounts for the aging process and to examine the role that TEs play in this process.

Molecular studies have found that over evolutionary history, some genes have experienced periods of very strong selection, whereas others appear to have been strongly shaped by drift (Li, 1997). Might this variation in selection translate to differences not just among genes but also among tissues? Early studies examined age-related declines in specific tissues, such as the early 20th-century work of Krumbiegel (1929) on fat body in aging Drosophila. More recent studies have found that the age-related rate of decline varies among tissue types. For example, in nematodes, muscle cells appear to age at a much faster rate than nerve cells (Herndon et al., 2002). Furthermore, the effect of gene signaling on longevity is often tissue-specific (Hwangbo et al., 2004; Libina et al., 2003). One obvious challenge is to determine whether tissues differ in their rate of aging because of different patterns of selection acting on different tissues. For example, the fitness consequences of a cell failing to produce hair follicles are likely to be quite different than the consequences of a pancreatic cell failing to produce insulin. If tissue-specific variation in rates of aging (what we might call “senescent heterochrony”—the opposite of the “one hoss shay” effect1) turns out to have deep evolutionary roots, it would be of interest to look for common genetic or epigenetic causes of this pattern.

Finally, comparisons of whole genomes across species have been used to infer physiological process from genetic pattern. For example, Eisen and Hanawalt (1999) determined the presence or absence of various DNA repair genes across more than 20 species of microbes. On the basis of their “phylogenomic” analysis, they were able not only to determine the degree of ubiquity of different repair genes or gene pathways, but also to predict the repair phenotypes of different microbes based on their underlying genotypes. Interestingly, they found that although some repair processes, such as those associated with the recA gene, are strikingly constant across taxa, others have evolved relatively recently. The genetic basis of repair differs among species, and in some cases, specific repair mechanisms have evolved convergently in different taxa. These findings should serve as a warning to biogerontologists. Although some aging genes, such as those in the insulin signaling pathway, may have deep evolutionary roots, our current focus on these ubiquitous pathways may lead us to overlook many others that evolve rapidly and are highly variable among species. We are confident that “phylogenomic” approaches will lead to profound insights into the evolutionary history of genetic mechanisms of aging in the coming years.

**B. Gene Networks**

Classic studies of aging relied on forward or reverse genetic techniques to identify individual genes that extended longevity. But a one-gene/one-trait perspective is clearly an oversimplification (Lewontin & White, 1960; Wright, 1932). The past few years have given us a new appreciation for the complex interactions among genes and proteins that affect the

---

1The observed patterns of senescent heterochrony contrast with the classic example of the “One Hoss Shay,” made famous in Oliver Wendel Holmes’ “The Deacon’s Masterpiece”:

The poor old chaise in a heap or mound,
As if it had been to the mill and ground!
You see, of course, if you’re not a dunce,
How it went to pieces all at once, –
All at once, and nothing first, –
Just as bubbles do when they burst.
formation of the final phenotype (Gibson & Honeycutt, 2002; Wolf et al., 2000). Already, studies have shown that in worms (Shook & Johnson, 1999) and flies (Jackson et al., 2002; Leips & Mackay, 2000), the way in which a particular allele affects longevity can depend on the presence of specific alleles at other loci (Spencer et al., 2003).

However, the shift from thinking about single genes to epistatic interactions between pairs of loci is still an oversimplification. We need to begin thinking about age-related changes to whole networks of interacting elements. Recent studies have shown that when genes or proteins interact within a complex network, the network structure itself can make the network resilient to damage in a way that would not be possible if all the elements in the network were operating independently (Albert et al., 2000; Flatt, 2005; Maslov et al., 2004; Siegal & Bergman, 2002; Wagner, 2000).

Networks can describe a wide array of interactions, from the regulatory interactions among genes, to social interactions among individuals, to transfer of electricity from power stations to users. In general, networks consist of “nodes” or “vertices” connected to each other by “edges.” The number of edges that a particular node has is called its “degree” or “connectivity,” and the frequency distribution of connectivity across all nodes in a network is the network’s degree distribution (Albert & Barabási, 2002). Biological networks typically have a degree distribution that approximates a power law (Albert & Barabási, 2002), such that the majority of nodes have just one or two edges, but some may have tens or hundreds of edges. Studies on network robustness illustrate that the strength of selection acting on a single gene depends largely on the network context within which that gene functions. For example, we might hypothesize that more highly connected genes or proteins are under stronger selection. This idea is supported by findings that more highly connected proteins evolve more slowly (Fraser et al., 2002), that genes that produce these proteins are more likely to have a lethal phenotype when knocked out (Jeong et al., 2001), and that these genes are less likely to be lost over evolutionary time (Krylov et al., 2003).

In light of these studies, we might expect that the structure of gene- and protein-interaction networks may influence which genes are associated with senescent decline (Sozou & Kirkwood, 2001). In a study of the yeast protein–protein interaction network, Promislow (2004) found that proteins with relatively high connectivity were more likely to be associated with replicative aging than proteins with fewer interactions. Furthermore, aging genes tended to have a higher degree of functional pleiotropy than expected by chance. Although Promislow (2004) argues that these results are consistent with the antagonistic pleiotropy theory of senescence, just why we observe these patterns is still an open question.

Molecular studies have identified complex pathways that affect senescence, best exemplified by work on the insulin-like/insulin growth factor signaling pathway in C. elegans and D. melanogaster (see Chapter 13, Henderson et al.; Chapter 15, Tu et al.). We now face the exciting challenge of using classical genetics approaches (Van Swinderen and Greenspan, 2005) and microarray studies to describe the larger networks of interacting genes that might affect aging. At the same time, we need theoretical models to predict how the network of nodes with which a single gene interacts can be used to predict (1) whether this gene will be associated with longevity and (2) how likely the gene is to fail as the organism ages.
III. From Physiology to Demography

In his book on the evolution of aging, Michael Rose defines senescence as "a persistent decline in the age-specific fitness components of an organism due to internal physiological deterioration" (our italics, page 20, Rose, 1991). Most evolutionary biologists studying senescence have focused on the decline in age-specific fitness components (mortality and fecundity). Whereas physiologists have focused on age-related changes in a wide array of physiological systems (Masoro, 1995), much less attention has been devoted to testing the hypothesis that demographic senescence is due to internal physiological deterioration (see Williams, 1999), or to exploring the possibility that physiological homeostasis may even limit the senescent decline in survival or fecundity (Kowald & Kirkwood, 1996).

In standard evolutionary models, the currency that measures how likely a gene is to make it into subsequent generations is made up solely of age-specific survival and fecundity (Equation 1). Evolutionary biologists are interested in changes in gene frequencies over time, so it seems logical to measure senescence by observing age-specific declines in survival and fecundity. Until recently, the physiological factors that are presumably the proximate cause of age-related changes in survival or fecundity have been considered of secondary importance among evolutionary gerontologists. In fact, some models suggest demographic senescence may evolve even in the absence of physiological decline. Houston and McNamara (1999) showed that mortality rates can increase with age solely as a result of how individuals optimize reproductive effort, and without any age-related deterioration in physiological state. Where evolutionary biologists have looked at both demographic and physiological parameters, the results have sometimes been counterintuitive. Recently, Reznick et al. (2004) found that guppies from high-predation environments had higher rates of decline in neuromuscular function, but they lived longer and had lower rates of reproductive aging than guppies from low-predation environments. Thus, there may not be a simple one-to-one relationship between physiological senescence and demographic decline.

The task of mapping out the relationship between genotype, physiological senescence and demographic senescence is no small challenge, but drawing these connections is crucially important. In the coming decades, the social and medical costs associated with physiological decline in aging humans will increase rapidly. We may gain most if we turn at least some of our attention from demographic quantity of life to physiological quality of life in evolutionary studies of aging. We see three primary areas where a more integrated approach, one that unites demography and physiology, can lead to a more comprehensive understanding of the biology of aging. In particular, we need to (1) develop more physiologically based theoretical models of senescence; (2) use classical quantitative genetic approaches to determine whether the genes that determine rates of physiological decline are the same ones that determine rates of demographic decline; and (3) include both physiological and demographic measures in molecular studies that are searching for specific genes and gene pathways that can slow the aging process. In the following section, we use the term physiology in its broadest sense—that is, the overall functioning of an organism. This includes organ performance, system function such as immunity (see Section IV), cell morphology, organismal behavior, and so forth. We need a broad definition as these non-demographic traits may overlap and interact in complex ways to ultimately
determine the fitness parameters of age-specific fertility, fecundity, and survival.

A. Physiological Models of Senescence
Within the theoretical literature on aging, not all studies have ignored physiology. For example, Mangel (2001) incorporated an organism’s energy level and the accumulated level of metabolic damage as physiological states into life-history models. These models predict how caloric restriction and reproduction affect the shape of the mortality curve. A physiologically structured model by Mangel and Bonsall (2004) predicts that the actual shape of the mortality curve can depend on physiological processes, such as growth and the level of repair. As information becomes available about the genetic control of declining organ function (e.g., Wessells et al., 2004), we should be able to construct ever-more realistic models for the way in which physiological senescence relates to demographic senescence.

B. Quantitative Genetic Analyses of Physiology and Demography
We discussed earlier the need to develop a more complex and subtle model for the genetic architecture of aging. We would further suggest that a complete understanding of the genetics of demographic aging must include physiology. But the genetic relationship between physiological and demographic senescence may be complicated. Physiological senescence may be the functional intermediary between so-called “longevity genes” and demographic senescence (see Figure 8.2a). Alternatively, the same genes may regulate rates of senescence in both demography and certain physiological processes independently, such that the demographic decline may not be specifically caused by those physiological processes (see Figure 8.2b). Last, the two processes could be both functionally and genetically independent (see Figure 8.2c). In a study on age-related changes in heart failure rate in Drosophila, Wessells and colleagues (2004) showed that long-lived insulin signaling mutants had much lower rates of heart failure. Although this study demonstrates that a gene that influences mortality rates also affects a physiological parameter, we do not know if heart failure plays any causal role in aging and/or death in fruit flies (the case illustrated in Figure 8.2b).

Figure 8.2 The influence of physiology in the genetics of senescence. The figure presents a schematic for three possible scenarios: [a] Genes influence physiological processes, which then lead to downstream effects on demographic senescence in mortality or fecundity. [b] The same genes that influence senescence in physiological processes independently determine rates of senescence in demographic traits. [c] Different genes determine rates of senescence in physiological and demographic processes.
Quantitative genetic studies have demonstrated that there is genetic variation for the rate of decline in age-specific survival (Hughes & Charlesworth, 1994; Promislow et al., 1996) and fecundity (Tatar et al., 1996). There is ample evidence to suggest that there is genetic variation for physiological traits in diverse taxa, from dairy cattle to Drosophila (Kiddy, 1979; Zera & Harshman, 2001). And we also know that genetic variation in longevity is correlated with underlying physiological differences [e.g., Djawdan et al., 1998; Gibbs et al., 1997]. But information about genetic variation for the age-related rate of decline in physiological function is rare (Wessells et al., 2004), and we know little about the genetic correlation with longevity and mortality parameters.

To map the relation between genotype and physiological and demographic senescence, we need to address a series of specific questions: Do genotypes that show a fast decline in physiological traits also have a higher rate at which intrinsic mortality increases with age? If an intervention extends life span through a decrease in the rate of increase in age-specific mortality rate, does it also decrease the rate of physiological deterioration? Alternatively, if an intervention extends life span through a decrease in the initial mortality rate, does the rate of physiological deterioration remain the same (see Chapter 1, Gavrilov and Gavrilova)? Is there a shared regulatory system that mediates the rate of deterioration for all physiological traits as predicted by Williams (1957), or is the genetic basis for the rate at which one trait deteriorates independent of the genetic basis for the rate at which another trait deteriorates? And finally, if the link between physiology and demography is not as straightforward as generally assumed, how does the shape of the selection curve change when physiological processes are incorporated into Hamilton’s (1966) theoretical model?

C. Molecular Genetic Analyses of Physiology and Demography

As with evolutionary geneticists, molecular biologists working on aging have also tended to focus on demographic traits. This has been motivated in large part by the search for genes that will make organisms live longer, no matter what the physiological state of the animal. Of course, most biogerontologists hope to find ways to increase not only the quantity of life, but also the quality of life at late ages (Arantes-Oliveira et al., 2003). But to do this, we need a balanced research program that includes both demographic and physiological perspectives.

Fortunately, we are beginning to see just such a shift in focus. For example, Huang and colleagues (2004) found that the age-related decline in pharyngeal pumping and body movement in C. elegans was positively correlated with life span among a series of mutants. This suggests either that the decline in physiological processes causes a reduction in survival probability (see Figure 8.2a), or that the declines in physiological function and survival are regulated by a shared mechanism (see Figure 8.2b). Future work should focus on attempts to test these two hypotheses explicitly and should include demographic parameters other than longevity. Interestingly, Huang and colleagues (2004) found that the age-related decline in pharyngeal pumping and body movement were not correlated with self-fertile reproductive span. The authors speculate that other measures of reproductive aging may correlate with physiological decline. Alternatively, genes that affect aging may have pleiotropic effects on body movement and survival, but not on fecundity.
We are also likely to see substantial progress from studies of cellular physiology. Using a histological approach to study aging in *C. elegans*, Herndon and colleagues (2002) found that long-lived age-1 mutants showed a slower rate of deterioration in cell ultrastructure. Interestingly, these lower rates were only seen in certain cell types.

Taken together, these two studies in *C. elegans* show that physiology is correlated with some but not all demographic traits (Huang *et al.*, 2004), and that demography is correlated with some but not all physiological traits (Herndon *et al.*, 2002). The challenge now before us is for theoretical, quantitative, and molecular geneticists to develop an integrated research program that incorporates genetics, physiology and demography to create a more integrated research program in biogerontology.

IV. Parasites and Immune Function

In the previous section, we argue that a critical challenge for evolutionary gerontologists is to bridge the gap between physiology and demography in studies of aging, and we propose some explicit ways in which this might be accomplished. One candidate for a ubiquitous factor that might tie together physiological senescence and demographic senescence is parasites.

We all have first-hand experience of the deleterious effects of parasites, and not surprisingly, there is abundant evidence for their life-shortening effects in model organisms. Likewise, when individuals are deprived of their normal bacterial flora they often show increased life span (e.g., Croll *et al.*, 1977; Garigan *et al.*, 2002; Houthoofd *et al.*, 2002; Larsen & Clarke, 2002; Min & Benzer, 1997). Recently, researchers have begun to identify molecular pathways that appear to have overlapping effects on survival and immune function. For instance, the secretion of juvenile hormone (JH) is crucial for several important reproductive pathways in insects, including gametogenesis and spermatophore production (Wigglesworth, 1965). However, increased JH titers have been shown to decrease immune function in the mealworm beetle (Rantala *et al.*, 2003; Rolff & Siva-Jothy, 2002) and to decrease longevity in monarch butterflies (Herman & Tatar, 2001). This apparent physiological antagonism between reproduction, immune function, and survivorship may play an important role in how insects age.

Although the proximate effects of parasites on longevity are often clear and straightforward, there are more subtle and interesting relationships between hosts and their parasites that may develop over evolutionary time. Parasites may play an important role in the evolution of a striking variety of biological traits, including the existence of sexual reproduction (Hamilton, 1980); the dramatic, dimorphic coloration in birds (Hamilton & Zuk, 1982); and the ability of organisms to invade novel habitats (Torchin *et al.*, 2003; Wolfe, 2002). And recent work on parasites and life-history strategies (Rolff & Siva-Jothy, 2003; Williams & Day, 2001) suggests that parasites may influence the way that natural selection shapes patterns of senescence.

A. From Immunosenescence to Demographic Senescence

We discussed the possibility that physiological decline may give rise to decline in fitness traits. One potentially important physiological factor is the age-related decline in immunocompetence, or “immunosenescence” (Walford, 1969). We have a pretty good idea of the proximate causes of immunosenescence in humans. For example, the thymus, where T cells mature, exhibits degenerative changes...
throughout life. Consequently, thymic tissue loses the capacity to influence a variety of important immune functions, including the repopulation of T cells in the lymph nodes (Hirokawa & Makinodan, 1975). Similarly, the sources of B cells, which fine-tune an antigen match for invading pathogens, begin to disappear over time (Leslie, 2004). This decline, coupled with a change in T-cell population with age, may lead to a decreased antibody response to most antigens (Weksler & Schwab, 1992). These age-specific changes in the immune system mark a dramatic physiological decline late in life that may greatly contribute to patterns of demographic senescence.

The cellular details differ, but this general pattern of immunosenescence is seen across an impressively broad array of species, including fish, birds, and reptiles (Torroba & Zapata, 2003). Insect species, including bumble bees, crickets, and dragonflies, also exhibit a marked age-related change in a variety of immune components, often leading to increased rates of parasitism and mortality (Adamo et al., 2001; Doums et al., 2002; Rolff, 2001). Even in the nematode, C. elegans, a positive association between age and susceptibility to bacterial infection is found (Laws et al., 2004). Whether a causal relationship exists between immunosenescence and the age-specific decline in fitness, however, is currently unknown.

Immune senescence could contribute to a general physiological decline if, as the immune system ages, it requires increased resources to maintain the status quo. Consequently, there may be fewer resources available for other costly physiological systems. For example, in the collared flycatcher (Ficedula albicollis), older females tend to suffer from a decline both in humoral immune function and offspring size (Cichon et al., 2003). However, more work is needed to determine whether the decline in offspring size is due to the increased energy demands of an aging immune system. Alternatively, the immune changes seen late in life may merely be a superficial marker of other underlying causes of aging. If so, then immunosenescence may serve as a useful biomarker of physiological age for future studies.

B. Immunocompetence Tradeoffs

Hamilton showed that the age-related decline in the strength of selection can be predicted solely on the basis of age-specific survival and fecundity (see Equations 2 and 3). Previous models have argued that these two traits should be negatively correlated because the amount of resources that can be invested in both is finite (de Jong & van Noordwijk, 1992). This tradeoff underlies Kirkwood’s disposable soma theory for the evolution of senescence (Kirkwood, 1977). However, this framework may be incomplete. We suggest here that in addition to investment in survival and reproduction, it may be worth considering other intermediate physiological components as distinct model elements, such as neuron, circulatory, or immune function. For instance, if investment in immune function is directly correlated with both age-specific reproduction and age-specific survival (e.g., if the physiological costs of immunity affect fecundity and survival simultaneously), immune function may account for the correlation between fecundity and survival. This, in turn, will influence the age-related decline in the strength of selection.

To avoid parasites, potential hosts invest in a variety of defenses, some of which prevent infection in the first place and others of which fight off the parasite once an infection has taken hold. But these defenses can be costly. For example, increasing investment in immunity is often paid for with lower fecundity (Zuk & Stoehr, 2002). Conversely, among individuals who increase their investment in reproduction, we see a decrease not only in survivorship (Partridge & Harvey, 1985), but also in immunocompetence (e.g.,
Fedorka et al., 2004). Most examples of these immune costs of reproduction come from insects. However, a recent study in women from 18th- and 19th-century Finland (Helle et al., 2004) found that women who had born twins at some time in their reproductive life span were more likely to succumb to an infectious disease after menopause than women who had only given birth to singletons. This result held even after controlling for the total number of offspring born to each woman.

There is still much work to be done as we try to sort out how costs of parasitism and immunity are translated into the currency of demographic senescence. One important problem will be to determine how the physiological costs of mounting an immune response and the mortality costs of being parasitized change with age. Such changes can have complicated and nonintuitive consequences on mortality rates. In traditional models of either host–parasite coevolution or senescence, as background mortality rates increase, virulence and rates of aging also increase, respectively. In two separate theoretical studies, Williams and Day (2001; 2003) demonstrate that when extrinsic sources of mortality interact with the host’s physiological state in a nonadditive fashion, virulence and rates of aging may actually decrease as background mortality rates increase. We now need theory that combines Williams and Day’s models of parasitism (Williams & Day, 2001) and senescence (Williams & Day, 2003) and empirical tests of the existing theory. Such models should help us determine how selection should shape patterns of age-specific investment in immunity and age-specific survival, and how the two may interact.

C. Parasites and the Genetics of Aging

Studies of aging and immunity may also further our understanding of the genetic architecture of senescence. Recent studies have shown that many of the genes that affect longevity play an integral role in immune defense (Caruso et al., 2001; Garsin et al., 2003; Ivanova et al., 1998; Lagaay et al., 1991; Laws et al., 2004). In some cases, immunity may be the proximate mechanism that gives rise to genes with antagonistic pleiotropic effects. For example, human centenarian studies have uncovered a strong association between Major Histocompatibility Complex (MHC) haplotypes and life span (Caruso et al., 2001; Ivanova et al., 1998; Lagaay et al., 1991). At least one human leukocyte antigen haplotype, 8.1 AH, may provide a selective advantage in early infancy by providing protection from infectious disease (Caruso et al., 2000). However, this haplotype is also associated with susceptibility to several autoimmune disorders that occur during reproductive adulthood, such as sarcoidosis and systemic lupus erythematosus (Lio et al., 1997; Price et al., 1999). Interestingly, this haplotype also provides a late-life, sex-specific advantage for male carriers, whereas female carriers continue to suffer from early morbidity and mortality late in life (Caruso et al., 2000).

Such tradeoffs may turn out to be quite common for genes associated with immune function. Another example comes from the MHC mutation C28Y. This mutation increases the intestinal absorption of iron, which is crucial in many immune pathways (Salter-Cid et al., 2000). However, the mutation is also associated with haemochromatosis in homozygotes, a disease characterized by a reduced life expectancy due to an excess accumulation of iron in the organs and concomitant reduction in immunity (Waheed et al., 1997). Researchers have also suggested relationships between tuberculosis and Tay-Sachs disease (Spyropoulos, 1988), and cholera and cystic fibrosis (Gabriel et al., 1994). In the past few years, biologists have begun to focus on the evolutionary consequences of immunity-related tradeoffs (reviewed in DeVeale et al., 2004),
although a detailed connection to senescence has yet to be made. Just how common these tradeoffs are, and whether some tradeoffs show age-specificity consistent with the antagonistic pleiotropy theory of senescence, remains to be seen. Finding such genes will provide an interesting challenge for those interested in developing an evolutionary perspective on immunosenescence.

Parasites may turn out to influence the genetics of aging in an even more general fashion. A recent study found that the ability of some genes to increase longevity depended on the presence of the bacterial flora (Brummel et al., 2004). For example, flies carrying the DJ817 mutant, which are normally long-lived, lost their life span advantage when treated with antibiotics. The same study showed that D. melanogaster males deprived of their normal bacterial complement in early adulthood suffered a significant reduction in life expectancy. In light of this interesting set of results, we should now set out to determine how many genes associated with longevity depend on the presence of bacteria for their phenotype, and of the many bacteria found in Drosophila, which ones are early-age symbionts and why.

D. Parasites and Sex Differences in Longevity

Finally, studies of parasites, immunity, and senescence may help us to understand why males and females often have quite different patterns of aging. Studies of immunosenescence have found that many of the MHC haplotypes associated with longevity are sex-dependent (Lagaay et al., 1991; Lio et al., 2002). Ivanova and colleagues (1998) found that of the three MHC alleles linked to human longevity in their study, two had a sex-dependent effect. Similarly, Lio and colleagues (2002) found a male-biased pattern among centenarians for polymorphisms in the promoter region of the IL-10 gene. IL-10 inhibits the pro-inflammatory immune response, which is believed to contribute substantially to mortality in late life. These patterns are not surprising considering that sex differences in immune function are widespread. Male mammals often have a lower level of immunocompetence (e.g., Klein & Nelson, 1997) and a higher rate of parasitic infection when compared to females (Moore & Wilson, 2002). Invertebrate males show a similar pattern (Adamo et al., 2001; Kurtz & Sauer, 2001; Kurtz et al., 2000; Radhika et al., 1998; Rolff, 2001), although due to the complex nature of life-history trade-offs, sex differences in immune defense are often difficult to predict (Doums et al., 2002; Moret & Schmid-Hempel, 2000; Zuk et al., 2004; Zuk & Stoehr, 2002). These sex-specific patterns suggest that the selection pressures that influence immunity and longevity are different for males and females.

We have suggested here that sex-specific differences in immune function may account, in part, for sex differences in longevity. In the following section, we move away from immunity and explore causes of sex-related differences in longevity in greater detail.

V. Sex, Sexual Selection, and Sexual Conflict

Sex differences in longevity are common. Female mammals generally live longer than their male counterparts (Promislow, 1992), with a few interesting exceptions, such as among anthropoid primates with extended paternal care, where males live longer than females (Allman et al., 1998). Conversely, in birds (Promislow et al., 1992) and nematodes (McCulloch & Gems, 2003), males tend to outlive females. Many specific mechanisms have been proposed to account for these patterns, from differences in hormon profiles (e.g., testosterone levels in males) to...
differences in predation risk (e.g., vulnerability of nesting females).

It is likely that at least part of these differences is due to a basic disparity in reproductive strategies. A female’s optimal reproductive strategy may be very different than that of her male partner (Rice, 1996). These differences may lead to very different age-specific selection pressures acting on the two sexes. We are only beginning to understand the full implications of these differences. In the following section, we describe three specific areas that could further our understanding of sex-specific patterns of senescence, including costs of reproduction, the proximate and evolutionary consequences of female choice, and intersexual conflict.

A. Cost of Reproduction

Reproduction can increase the mortality rate in many organisms (e.g., Fedorka et al., 2004; Sgro & Partridge, 1999), but these costs can differ dramatically between the sexes in a range of organisms (Lyons & Dunne, 2003; Michener & Locklear, 1990; Rocheleau & Houle, 2001). Males and females may differ in the resources they allocate to gamete production, parental care, and mating effort. Dissimilar reproductive investment can, in turn, lead to differences in the risk of predation or sexually transmitted disease, or in the resources available for somatic maintenance. The evolution of sexually dimorphic traits through female choice or male competition exemplifies these sex-specific costs. To attract females or outcompete other males for access to mates, males often develop exaggerated secondary sexual characteristics, such as bright plumage or coloration, conspicuous calling songs, or large antlers. However, these traits may also attract predators or encumber the male when trying to escape from harm. Furthermore, these traits come at a large physiological cost and may reduce the resources that are available for immune defense or other traits, leading to reduced survival (Andersson, 1994). These changes in survival rates can then lead to sex-specific differences in the declining force of selection with age. At this point, we need sex-specific models to help us determine how sex differences in the risk of mortality might lead to sex differences in rates of aging.

B. Female Mate Choice

There is extensive evidence from different species suggesting that females will often choose to mate with males based on their genetic quality (Andersson, 1994). In so doing, choosy females may affect the evolution of senescence. For instance, according to Hamilton and Zuk (1982), female birds may use a male’s bright coloration to assess his underlying ability to resist parasitic infection. If female choice thereby increases immunocompetence in the population, mean survival rates may increase, and selection on senescence will change accordingly.

Female choice may have a more direct effect on senescence in a population if females prefer to mate with males of a particular age. Some theoretical studies have suggested that females might choose to mate with older males because they have proven their ability to live long (Beck & Powell, 2000; but see Hansen & Price, 1995; Kokko, 1998). In their computer simulations, Beck and colleagues (2002) found that when females were allowed to choose males based on their age, they typically evolved a preference for older males. In a genetically heterogeneous population, older males will have a higher proportion of alleles associated with increased life span than younger males. Thus, female preference for older males leads not only to lower mortality rates in the choosy female’s offspring, but will actually increase mean life span in the entire cohort over evolutionary time.
The theoretical findings that choosiness can lead to increased longevity are consistent with experimental studies in *Drosophila*. Promislow and colleagues (1998) compared mortality rates in strains that were maintained under enforced monogamy for multiple generations with strains where female choice and male–male competition were allowed. They found that the strains with higher levels of sexual selection evolved lower mortality rates. Similarly, among birds, female survival rates are highest in those species that appear to have been under stronger sexual selection (Promislow et al., 1992). Further empirical work is needed to determine whether female preference for older males, in particular, can alter patterns of senescence in natural populations.

C. Sexual Conflict

Sexual conflict has also been suggested as playing a significant role in the evolution of senescence (Promislow, 2003; Svensson & Sheldon, 1998). Sexual conflict arises when the optimal reproductive strategy differs between the sexes. For example, a female may enhance her fitness by mating multiple times with different males, but this may reduce the fitness of each of her mates by diluting his sperm with those of his rivals. These conflicts can give rise to a situation where males, in an effort to maximize their fitness, evolve the capacity to dramatically reduce female fitness.

We have long known that male behavior can be detrimental to females: Parker and Thompson (1980) observed that, in the quest for mating opportunities, competing male dung flies (*Scatophaga stercoraria*) would often drown potential mates, leading to an obvious conflict of interest between the dying female and the overzealous male. Similarly, male bed bugs (*Cimex lectularius*) forcibly inseminate females by piercing their abdominal wall with needle-like genitalia, leading to high rates of female mortality (Stutt & Siva-Jothy, 2001). Conflict can even take place at a molecular level. In *Drosophila* spp., males pass accessory gland proteins (Acps) during copulation that may decrease a female’s sexual receptivity, incapacitate previously donated sperm, prevent future sperm displacement (see Chapman et al., 1995, and references therein), and increase the rate of oviposition (Heifetz et al., 2000). Moreover, Acps tend to decrease a female's overall life expectancy in a dose-dependent manner (Chapman et al., 1995; Lung et al., 2002). And tying Acps more directly to longevity, Moshitzky and colleagues (1996) have shown that one of the sex peptides, a protein found in the male ejaculate, can induce the biosynthesis of juvenile hormone, which is associated with longevity (see Chapter 15, Tu et al.). Further experiments are needed to determine whether sexual conflict can affect rates of senescence and differences in patterns of senescence between the sexes. Are rates of aging higher in species with greater sexual conflict? Do species with higher levels of conflict show greater sex-differences in rates of aging than species with lower levels of conflict? And are genes associated with conflict (such as Acps in *Drosophila*) also associated with variation in longevity in either sex?

VI. Genetic Variation in Natural Populations

In a separate chapter in this volume, Brunet-Rossinni and Austad (see Chapter 9) examine evidence for senescence in natural populations. Both lab and natural populations generally show similar Gompertz-like increases in mortality. However, several studies have shown that when we bring organisms into a lab environment, we often inadvertently expose them to novel selection
pressures that lead to shorter life span [Sgrò & Partridge, 2000; Linnen et al., 2001]. Thus, what we learn about the genetics of aging in the lab may not always transfer to the wild.

For many years, senescence was assumed to be rare in the wild (Comfort, 1979) because organisms were most likely to die by accident before they would have a chance to senesce. However, later studies demonstrated that senescence is not an artifact created by ideal laboratory conditions but is actually common in natural populations of mammals [Promislow, 1991] and birds [Ricklefs, 1998; 2000]. We even see evidence for senescence in natural populations of insects. Bonduriansky and Brassil (2002) demonstrated aging in wild populations of very short-lived male antler flies, and Carey (2002) was even able to find evidence of aging in the famously short-lived mayfly. These new studies on aging in nature in short-lived animals suggest the possibility for developing free-living model systems to study the genetics of aging.

A few studies have demonstrated a genetic basis to variation in rates of aging in natural populations (e.g., Bronikowski et al., 2002; Reznick et al., 2004; Tatar et al., 1997). Not surprisingly, most of what we know about the genetics of aging comes from lab studies. But we need to be especially cautious in assuming that what we learn in the lab translates directly to the field. We usually assume that an allele that increases longevity in a lab incubator will also increase longevity in the wild. Unfortunately, when we introduce natural populations into the lab, the change in environment can select for quite dramatic changes in demographic characteristics. Lab-adapted organisms typically mature earlier and have increased early-age fecundity and greatly shortened life span [Clark, 1987; Houle & Rowe, 2003; Miller et al., 2002; Promislow & Tatar, 1998; Sgrò & Partridge, 2000]. One might argue that the life-extending effects of novel mutants or artificial selection simply restore short-lived lab strains to the longevity of their wild relatives [Linnen et al., 2001].

Recent studies from worms and flies point to two specific factors that we need to consider before assuming that lab results necessarily apply to wild populations. First, the effect of a mutation can depend on the environment in which an organism lives, due to gene by environment (G × E) interaction. The importance of G × E interactions in the genetics of longevity has been explored in great detail by Mackay and her co-workers (Chapter 7, Mackay et al.). Second, some studies appear to find life extension at no cost. For instance, the daf-2 mutation in C. elegans can greatly extend longevity in the laboratory without apparent loss of fertility or activity [Kenyon et al., 1993]. However, when the long-lived daf-2 mutant is combined with the wildtype strain in the same culture, the long-lived mutant is always outcompeted [Jenkins et al., 2004]. In general, it may turn out that when placed in a natural environment, genes that extend life span in the lab may have unanticipated negative consequences for fitness.

Several studies have now tried to determine whether alleles associated with longevity in the lab have similar functions in natural populations. Schmidt and colleagues (2000) showed that there is geographic variation in the frequency of methuselah, a Drosophila gene that can substantially increase longevity [Lin et al., 1998]. However, clinal variation in longevity was not correlated with clinal variation in single nucleotide polymorphisms at the methuselah gene. Other studies have had greater luck in applying lab-based results to field populations. Geiger-Thornsberry and Mackay (2004) studied genetic variation for genes associated with aging in the lab in wild-derived inbred strains of Drosophila.
Using quantitative complementation tests (see Chapter 7, Mackay et al.), they found that there was standing genetic variation at these loci in natural populations, and that this genetic variation was correlated with variation in longevity among strains. We now need to determine whether Geiger-Thornsberry and Mackay’s findings with the InR and Adh genes are the exception or the rule, and how selection has shaped these loci over evolutionary time.

It is clearly important to determine whether genes identified in lab-reared populations of flies are effective at extending life span in natural populations. But an even greater challenge beckons. Can we translate lab-based findings to natural populations of different taxa? Can the genetic results for flies or worms in the lab be applied to natural populations of vertebrates? Researchers have suggested that various fish species, including killifish (Herrera & Jagadeeswaran, 2004), zebrafish (Gerhard, 2003; Gerhard & Cheng, 2002), and guppies (Reznick, 1997), may be ideal model systems to help bridge the gap between invertebrate and mammalian taxa in studies of aging. Reznick recently used a natural population of guppies to test the hypothesis that rates of senescence should be highest in populations with high extrinsic mortality (Reznick et al., 2004). Interestingly, the data did not support this classic hypothesis, which reiterates the point we made earlier in this chapter that we need more biologically realistic models for the evolution of senescence.

Researchers have also argued that birds (Holmes et al., 2001), bats (Wilkinson & South, 2002), and other “slow-aging” organisms (Austad, 2001) may be valuable model systems for the study of senescence. By extending the range of taxa that we use to study aging, we should be able to determine the extent to which mechanisms of aging are shared among taxa. Although there are distinct advantages to studying short-lived species, studying long-lived organisms in the wild (or the lab, for that matter) could add a new dimension to aging research (Holmes & Ottinger, 2003).

VII. Conclusions

The most important conceptual advances in our understanding of the evolution of senescence came from the initial models (Medawar, 1946; 1952; Williams 1957). Since the time of Medawar’s and Williams’s pioneering efforts, a half-century of evolutionary research has led to extraordinary advances in our understanding of evolutionary patterns and processes, from adaptation to altruism, from speciation to sex ratios. However, with few exceptions, these enormous conceptual advances have occurred independently of evolutionary studies of aging.

In this chapter, we have explored five areas that have excellent potential to further our understanding of the evolution of senescence. These include the genetic architecture of senescence, the relationship between physiological and demographic decline, the importance of parasites and the immune system, sex differences in behavior and aging, and studies of aging in the wild. Some of these areas include those in which we have already seen important advances by evolutionary biologists working outside of the field of aging. All are linked by the general problem of which forces will shape the age-related decline in the force of selection.

We hope that the questions that we have posed here will encourage more evolutionary biologists to take on the problem of the evolution of senescence. In the meantime, most of the action in biogerontological research appears to be found in the molecular labs, where new genes and gene pathways that affect aging are being uncovered at an extraordinary
rate. One might ask, then, whether there is still a need for an evolutionary perspective in this fast-paced world of sequenced genomes and microarrays. We believe that the most exciting challenge facing all of us is to develop an integrated research program, where molecular and evolutionary biologists work hand in hand—molecular biologists asking evolutionary questions, evolutionary biologists using state-of-the-art molecular tools. With such an integration, as old problems are solved, new and interesting questions will arise at an even faster rate, ensuring that the study of aging continues to be exciting and enlightening.

Acknowledgements

We thank Thomas Flatt and members of the Promislow lab for thoughtful discussions. Richard Miller, Steve Austad, Edward Masoro, and one anonymous reviewer provided valuable comments on a previous draft of this manuscript. This work was supported by an NSF Minority post-doctoral fellowship from the National Science Foundation (KF) and a grant from the Ellison Medical Foundation (DP).

References


Mechanisms of Ageing and Development, 125, 179–189.


Tissenbaum, H. A., & Guarente, L. (2001). Increased dosage of a sir-2 gene extends...


I. Introduction

Senescence or, synonymously in this chapter, aging can be defined as the progressive deterioration in physiological function that accompanies increasing adult age. Like the proverbial elephant described by blind men, it has many visible forms, depending on your perspective. For instance, to a geriatrician, senescence is evident as an age-related increase in frailty or vulnerability and a mounting incidence and severity of degenerative diseases. To a demographer, it is most easily measured as an age-related increase in the probability of death, and to an evolutionary biologist, senescence might describe the progressive decline in age-specific Darwinian fitness components. These are all valid embodiments of senescence; however, some will be more easily observed and measured in natural populations and less easily confounded by other phenomena than others. Also, various signs of senescence may appear at different ages.

It has long been conjectured, particularly in the biomedical community, that animals in nature do not experience senescence (Comfort, 1979, Hayflick, 2000; Medawar, 1952). Among ecologists, such a general assumption has never been advanced; however, some animal groups, particularly birds and fishes, have been identified as likely to die at a constant rate in nature, irrespective of age (Deevey, 1947). Even this more restricted statement has met with subsequent skepticism from field biologists. Botkin and Miller (1974) pointed out that if annual mortality rate were indeed independent of age, then one or more of the royal albatrosses seen by Captain Cook during his first visit to New Zealand in 1769 should still be alive today! As remarkable as avian longevity might be, that is a hard tale to swallow. The Guinness Book of Records lists the oldest royal albatross identified to date as 53 years old.

The notion that animals do not senesce in nature arises from an implicit belief that life in the wild is so nasty and brutish, so beset by predation, pestilence,
foul weather, famine, and drought that animals must inevitably die before signs of senescence appear. According to this intuition, only under the protected conditions of captivity (or civilization) does aging become manifest. No doubt such a scenario describes life for some species. For instance, multiple field studies of house mice (*Mus musculus*) indicate that median longevity in nature is only about 3 to 4 months, with 90 percent of deaths occurring by about 6 months of age (Phelan & Austad, 1989)—ages at which senescence is indeed not readily observable even in the laboratory. This is in contrast to the 2- to 3-year mean longevity under protected conditions (Turturro et al. 1999).

On the other hand, the general claim of negligible senescence in the wild, despite its persistence in the biogerontological literature, is quite clearly incorrect, as we will show. We are not suggesting that animals in nature reach the advanced state of decrepitude found in the city or the laboratory, but only that an observable age-related decline in function is not rare.

In this article, we review evidence for senescence in birds and mammals in the wild. We have limited our survey to birds and mammals not because they are uniquely senescence-prone, but to keep the size of our review manageable and also because of the availability of numerous long-term field studies in which individuals have been monitored for a number of years, sometimes throughout life, for these two animal groups. Such studies are not required to detect senescence, but they are the most likely to do so. Such studies also avoid many of the pitfalls of less intensive studies. We discuss the nature of these pitfalls later on.

It is important to note that detecting senescence is the primary goal of few if any field studies. Data relevant to senescence are usually adventitiously acquired while investigating other issues. Therefore, the failure to detect senescence in a study may say more about the length, design, and intensity of the study than the probability that the population under study did, or did not, exhibit signs of senescence.

II. Evidence of Senescence in Wild Populations

A. Demographic Senescence

The majority of articles assessing senescence in natural populations have focused on detecting an age-related increase in mortality rate (actuarial or survival senescence) and/or decrease in reproductive rate (reproductive senescence). Together we call these measures of *demographic* senescence. Increasing age-specific mortality in adulthood is widely used as the gold standard of senescence, both in natural and in laboratory populations. This choice stems from the assumption that age-specific death rate is a good indicator of intrinsic physiological hardiness. As hardiness declines with age, the probability of death should rise. Such a rise has often been found in natural populations. For instance, Promislow’s (1991) analysis of 56 published life tables from mammal field studies found statistically significant evidence for survival senescence in 44 percent (26) of the studies and a nonsignificant trend in the same direction in a further 34 percent (20) of the studies. Thus, there was at least some reason to suspect the existence of senescence in nearly four of five published mammal field studies. Later, Gaillard and colleagues (1994) published a slightly more conservative reanalysis of the same data but still found statistically compelling evidence for senescence in 42 percent (25 of 59) of data sets.

Sometimes, as in the previously mentioned Royal Albatross, survival senescence can be inferred by the lack of survivors above a certain age. For instance, Ian Nisbet has been studying the common tern, a small seabird, at an island off the Massachusetts coast for
more than 30 years and has individually marked thousands of birds. The annual observed adult survival rate is 0.9. If survival rate were independent of age, then about 5 percent of adults (assuming adulthood begins at age 2) should survive to 30 years of age and 1 percent should survive 45 years. However, direct observation finds that the 5 percent survival age is 18 years and the longest-lived bird seen so far is only 26 years old (Nisbet, 2001).

Measurement of survival senescence has some practical difficulties when applied to wild populations, however. One such complication is seasonality. For instance, a number of small mammals develop in the spring, mature in the summer and autumn, and tend to die in the winter. Among species such as these, increasing age-related mortality rate could indicate senescence but could just as easily indicate increasing environmental harshness from summer to autumn to winter. That is, even if no physiological deterioration has occurred in the study populations, mortality rate could still increase with age. Consequently, increasing mortality rate with age does not by itself unequivocally demonstrate senescence in the wild.

Another problem is environmental change with unknown effects on demography. For instance, the common tern colony studied by Nisbet and colleagues grew from about 500 breeders in 1970 to about 4,500 breeders in 1992. An early report from this colony suggested that several markers of reproductive function (clutch size, egg volume) decreased with age (Nisbet et al., 1984). However a much later report found no evidence of reproductive senescence (Nisbet et al., 2002). Whether this difference was due to the increased population density, some other environmental factor, or simply more and better information is not clear.

In addition to complications involving seasonality or environmental change, several other factors warrant caution in interpreting the failure to find increasing mortality as indicating an absence of senescence. First, for any given cohort, nonseasonal, stochastic environmental factors may override or mask even the very real effects of age on physical state. As one example, a classic study of house sparrows found that a New England winter storm preferentially killed birds that were particularly large or small relative to the mean size in the population (Bumpus, 1898). Body size is not related to age in birds, therefore the storm likely killed birds with little if any relation to their age. However the mortality from that single event might have overwhelmed any underlying age-related pattern. Stochastic climatic events can also depress food availability, which is likely to lead to more risk-taking by foragers and thus a generally higher mortality rate. Environmental events of this type can confound detection of survival senescence or its absence.

Significantly, dead animals are seldom recovered in field studies. Death is typically assumed when marked animals disappear from a study area, but they could also have emigrated. As a result, mortality and emigration can easily be conflated. In addition, if individual identification is by tags or bands, these markers can fall off. If so, a death will be mistakenly recorded. Third, even if death can be unambiguously detected, mortality rates can easily be affected by behavior as well as by intrinsic deterioration. For instance, animals engaging in high-risk behaviors will obviously die at higher rates than their more cautious peers. If animals in certain age classes are particularly prone to risky behavior, underlying physical senescence patterns again may be masked (see Figure 9.1). Are there particular ages at which animals in nature are prone to engage in high-risk behaviors? Indeed, as in humans, adolescent and newly maturing animals frequently take exceptional risks. Typically, birds and mammals disperse from their birth site around the time
of sexual maturity to seek and compete for new territories and/or mates. Venturing through unknown areas and competing for territories or mates can be very dangerous. In one of the few studies in which this "dispersal risk" could be quantified, female water-voles (Arvicola terrestris) died at at least an 86-fold higher rate during dispersal than when remaining in their natal home range (Leuze, 1980). Any such behavior-mediated elevation of mortality rate in early life will require that statistical analyses be sensitive to the possibility that temporary, early life mortality spikes might make a simple Gompertzian analysis of the entire adult life somewhat misleading (see Figure 9.1).

Survival is only part of demography. An age-related decline in reproductive performance can also be a sensitive indicator of physical senescence. For instance, among human females, fertility declines are detectable by about age 30, an age of minimal senescence by other measures (Dunson et al., 2004; Shock, 1983). One main difficulty with using reproductive function as a measure of senescence is that, like age-specific mortality, environmental events can mask underlying patterns. Hard times—food shortages, temperature extremes, exceptional predator abundance—rather than aging may depress reproduction in a study population. Reproductive performance also has the disadvantage that it is generally much easier to monitor in one sex (females) relative to the other. So, too, there are subtleties of reproductive senescence that will be difficult to detect in either field or laboratory. For instance, reproductive senescence may manifest itself as a decline in the phenotypic quality of offspring from older females rather than a straightforward reduction in number of eggs or newborns (Saino et al., 2002). As another example, slow growth of unweaned pups rather than reduced litter size marked reproductive senescence in Virginia opossums (Austad, 1993).

Both reproductive and survival senescence can also be difficult to detect in the presence of infectious disease. The relationship between senescence and disease is complex (see Masoro, Chapter 2). Decreased reproductive performance near the very end of life might as easily reflect infectious status as senescence per se. For instance, in a long-term study of an oceanic bird species, the black-legged kittiwake, Coulson and Fairweather (2001) observed depressed reproductive performance in the final breeding period prior to death in birds of all ages. Because this was observed in young as well as old birds, the authors interpreted this pattern as a sign of terminal illness, not senescence, although these are clearly not mutually exclusive.

Such difficulties aside, an extensive but not exhaustive survey of the literature

![Figure 9.1 Misleading inference of no senescence caused by elevated risk-taking during dispersal or territory acquisition. A. Regression from points sampled. B. Underlying mortality pattern. Ages sampled. Dashed line = actual adult mortality trajectory, solid line = inferred mortality trajectory from 5 sampled ages.](image-url)
turned up evidence for demographic senescence [either survival or reproductive] in natural populations of 42 species of mammals and 35 species of birds (summarized in Tables 9.1 and 9.2). Included in this list are the 26 species for which Promislow (1991) found statistically significant (P < 0.1) evidence of senescence. We excluded from this list data on captive or semi-wild populations and claims of senescence based only on secondary sources. However, we included any article in which the authors claimed evidence of senescence in natural populations or in which we could easily detect such evidence from the published data even if the author(s) failed to note it.

In addition to the studies we cite, of course, a number of studies have failed to find demographic senescence. However, because these studies varied in their intensity and evidentiary methodology, it

<table>
<thead>
<tr>
<th>Species</th>
<th>Survival senescence</th>
<th>Reproductive senescence</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Didelphis virginiana</td>
<td>Austad, 1993</td>
<td>† infertility, ‡ pouch young growth rate (♀)</td>
<td>Austad, 1993</td>
</tr>
<tr>
<td>Virginia opossum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pipistrellus pipistrellus</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pipistrelle bat</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macaca mulatta</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhesus macaque</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macaca fuscata</td>
<td></td>
<td>‡ birth rate (♀)</td>
<td></td>
</tr>
<tr>
<td>Japanese macaque</td>
<td>Bronikowski et al. 2002</td>
<td>‡ maternity rate, ‡ fertility (♀)</td>
<td></td>
</tr>
<tr>
<td>Papio hamadryas</td>
<td>Packer et al. 1998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baboon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lutra canadensis</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>River otter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martes zibellina</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sable marten</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panthera leo</td>
<td>Packer et al. 1998</td>
<td>‡ maternity rate (♀)</td>
<td></td>
</tr>
<tr>
<td>African Lion</td>
<td>Promislow 1991</td>
<td>‡ surviving cubs/male (♂)</td>
<td></td>
</tr>
<tr>
<td>Ursus arctos</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown bear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ursus maritimus</td>
<td></td>
<td>‡ litter size and mass (♀)</td>
<td>‡ body mass (♀)</td>
</tr>
<tr>
<td>Polar bear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alces alces</td>
<td>Ericsson et al. 2001</td>
<td>‡ litter size, ‡ offspring mortality (♀)</td>
<td></td>
</tr>
<tr>
<td>Moose</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continues)
<table>
<thead>
<tr>
<th>Species</th>
<th>Survival senescence</th>
<th>Reproductive senescence</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fallow deer</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Aepyceros melampus</em></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Impala</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hemitragus jemlahicus</em></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Himalayan tahr</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ovis dalli</em></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dall’s sheep</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Kobus kob</em></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Kob</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Rupicapra rupicapra</em></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chamois</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cervus elaphus</em></td>
<td>Clutton-Brock et al. 1988</td>
<td>↓ fecundity,↑ calf mortality [♀]</td>
<td>↓ body mass Mysterud et al. 2001</td>
</tr>
<tr>
<td><strong>Red deer</strong></td>
<td></td>
<td>↓ reproduction probability [♂]</td>
<td>↓ control of harems [♂]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ Clutton-Brock 1984</td>
<td>Clutton-Brock et al. 1979</td>
</tr>
<tr>
<td><strong>Tragelaphus strepsiceros</strong></td>
<td>Owen-Smith 1993</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Greater kudu</strong></td>
<td>Sinclair, 1977</td>
<td>↓ pregnancy rate [♀]</td>
<td></td>
</tr>
<tr>
<td><em>Syncerus caffer</em></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>African buffalo</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ovis musimon</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mouflon</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ parental care [♀]</td>
<td>Réale &amp; Boussès 1995</td>
</tr>
<tr>
<td><strong>Hippopotamus amphibious</strong></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hippopotamus</strong></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Phacochoerus aethiopicus</em></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Desert warthog</strong></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lutra canadensis</em></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>River otter</strong></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sable marten</strong></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Panthera leo</em></td>
<td>Packer et al. 1998</td>
<td>↓ maternity rate [♀]</td>
<td></td>
</tr>
<tr>
<td><strong>African Lion</strong></td>
<td>Promislow 1991</td>
<td>↓ surviving cubs/male [♂]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Packer et al. 1998</td>
<td></td>
</tr>
<tr>
<td><strong>Ursus arctos</strong></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Brown bear</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ursus maritimus</em></td>
<td>Derocher &amp; Stirling 1994</td>
<td>↓ litter size and mass [♀]</td>
<td>↓ body mass [♀]</td>
</tr>
<tr>
<td><strong>Polar bear</strong></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>European hare</strong></td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continues)
<table>
<thead>
<tr>
<th>Species</th>
<th>Survival senescence</th>
<th>Reproductive senescence</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oryctolagus cuniculus</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European rabbit</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sylvilagus floridanus</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern cottontail</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equus burchelli</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burchell's zebra</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Callorhinus ursinus</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern fur seal</td>
<td>Lunn et al. 1994</td>
<td>↓ reproductive rates (♀)</td>
<td></td>
</tr>
<tr>
<td>Arctocephalus gazelle</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antarctic fur seal</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phoca hispida</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ringed seal</td>
<td>Promislow 1991</td>
<td>↓ birth rate (♀)</td>
<td>Sugiyama 1994</td>
</tr>
<tr>
<td>Pan troglodytes</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loxodonta Africana</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African elephant</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microtus agrestis</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field vole</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apodemus flavicollis</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yellow-necked mouse</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peromyscus maniculatus</td>
<td>Millar 1994</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deer mouse</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peromyscus leucopus</td>
<td>Promislow 1991</td>
<td>↓ litter size of old, large females</td>
<td>Morris 1996</td>
</tr>
<tr>
<td>White-footed mouse</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tamiasciurus hudsonicus</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red squirrel</td>
<td>Promislow 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spermophilus colombianus</td>
<td>Promislow 1991</td>
<td>↑ unsuccessful litters (♀)</td>
<td>Broussard et al. 2003</td>
</tr>
</tbody>
</table>

**Table 9.2**

<table>
<thead>
<tr>
<th>Species</th>
<th>Survival senescence</th>
<th>Reproductive senescence</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphelocoma coerulescens</td>
<td>McDonald et al. 1996</td>
<td>↓ number of fledglings, ↓ offspring</td>
<td></td>
</tr>
<tr>
<td>Florida scrub jay</td>
<td></td>
<td>survival</td>
<td>Fitzpatrick &amp; Woolfenden 1988</td>
</tr>
<tr>
<td>Pica pica</td>
<td></td>
<td>↓ clutch size</td>
<td>Birkhead &amp; Goodburn 1989</td>
</tr>
<tr>
<td>Black-billed Magpie</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continues)
Table 9.2  (Cont’d)

<table>
<thead>
<tr>
<th>Species</th>
<th>Survival senescence</th>
<th>Reproductive senescence</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bucephala clangula</td>
<td></td>
<td>↓ brood survival</td>
<td>Milonoff et al. 2002</td>
</tr>
<tr>
<td>Common goldeneye</td>
<td></td>
<td>↓ clutch and brood size</td>
<td>Dow &amp; Fredga 1984</td>
</tr>
<tr>
<td>Larus glaucescens</td>
<td></td>
<td>↓ egg volume, ↓ hatching success</td>
<td>Reid 1988</td>
</tr>
<tr>
<td>Glaucous-winged gull</td>
<td></td>
<td>↓ fledging success</td>
<td>Anderson &amp; Apanius, 2003</td>
</tr>
<tr>
<td>Sula granti</td>
<td></td>
<td>↓ hatching success</td>
<td></td>
</tr>
<tr>
<td>Nazca booby</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larus delawarensis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ring-billed gull</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larus californicus</td>
<td>Pugesek 1987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>California gull</td>
<td>Pugesek et al. 1995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larus canus</td>
<td>Rattiste &amp; Lilleleht 1987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common gull</td>
<td></td>
<td>↓ breeding success</td>
<td></td>
</tr>
<tr>
<td>Fulmarus glacialis</td>
<td>Dunnet &amp; Ollason 1978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern Fulmar</td>
<td></td>
<td></td>
<td>Ollason &amp; Dunnet 1978</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↓ fecundity, ↓ breeding success</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ollason &amp; Dunnet 1988</td>
</tr>
<tr>
<td>Rissa tridactyla</td>
<td>Aebischer &amp; Coulson 1990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black-legged Kittiwake</td>
<td>Coulson &amp; Wooller 1976</td>
<td></td>
<td>Thomas 1983</td>
</tr>
<tr>
<td>Somateria mollissima</td>
<td>Coulson 1984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Eider</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puffinus tenuirostris</td>
<td>Bradley et al. 1989</td>
<td></td>
<td>Bailey &amp; Milne 1982</td>
</tr>
<tr>
<td>Short-tailed shearwater</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phalacrocorax aristotelis</td>
<td>Aebischer 1986</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shag</td>
<td>Harris et al. 1994</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catharacta skua</td>
<td></td>
<td>↓ clutch volume</td>
<td>Hamer &amp; Furness 1991</td>
</tr>
<tr>
<td>Great skua</td>
<td></td>
<td>↓ clutch size</td>
<td></td>
</tr>
<tr>
<td>Diomedea exulans</td>
<td>Weimerskirch 1992</td>
<td>↓ egg size, ↓ breeding success and frequency</td>
<td>Weimerskirch 1992</td>
</tr>
<tr>
<td>Wandering albatross</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sterna hirundo</td>
<td></td>
<td>↓ clutch size, ↓ egg volume</td>
<td>Nisbet et al. 1984</td>
</tr>
<tr>
<td>Common tern</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sterna paradisaea</td>
<td></td>
<td>↓ clutch size and volume</td>
<td>Coulson &amp; Horobin 1976</td>
</tr>
<tr>
<td>Arctic tern</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen caerulescens</td>
<td></td>
<td>↓ hatchability, ↑ brood loss</td>
<td>Rockwell et al. 1993</td>
</tr>
<tr>
<td>Snow goose</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[continues]
<table>
<thead>
<tr>
<th>Species</th>
<th>Survival senescence</th>
<th>Reproductive senescence</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirundo rustica</td>
<td>Møller &amp; de Lope 1999</td>
<td>↓ offspring quality</td>
<td>↓ ectoparasite load,</td>
</tr>
<tr>
<td>Barn swallow</td>
<td>Saino et al. 2002</td>
<td>↓ number of fledglings,</td>
<td>↓ secondary sexual characters,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ reproductive value</td>
<td>↑ fluctuating asymmetry,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↓ body mass,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>delayed arrival from migration</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Møller and de Lope 1999;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↓ humoral immunity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Saino et al. 2003</td>
</tr>
<tr>
<td>Tachycineta bicolor</td>
<td></td>
<td>↓ breeding performance index</td>
<td></td>
</tr>
<tr>
<td>Tree swallow</td>
<td>Robertson &amp; Rendell 2001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calidris temminckii</td>
<td>Hilden 1978</td>
<td>↓ nesting success,</td>
<td></td>
</tr>
<tr>
<td>Temminck's stint</td>
<td></td>
<td>↓ brood size,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ juvenile survival</td>
<td></td>
</tr>
<tr>
<td>Parus major</td>
<td>Dhondt 1989</td>
<td>↓ hatching rate,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ fledgling survival</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Perrins &amp; Moss 1974</td>
</tr>
<tr>
<td>Great tit</td>
<td></td>
<td>↓ nesting success,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ clutch and brood size,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ juvenile survival</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dhondt 1989</td>
</tr>
<tr>
<td>Parus caeruleus</td>
<td></td>
<td>↓ nesting success,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ clutch and brood size,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ juvenile survival</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dhondt 1989</td>
</tr>
<tr>
<td>Blue tit</td>
<td></td>
<td>↓ hatching success,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ number of fledglings,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ clutch size</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parus montanus</td>
<td>Orell &amp; Belda 2002</td>
<td>↓ clutch size,</td>
<td></td>
</tr>
<tr>
<td>Willow tit</td>
<td></td>
<td>↓ fledging success</td>
<td></td>
</tr>
<tr>
<td>Pyrrhocorax</td>
<td>Reid et al. 2003b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyrrhocorax</td>
<td></td>
<td>↓ successful broods,</td>
<td></td>
</tr>
<tr>
<td>Red-billed chough</td>
<td></td>
<td>↓ number of fledglings,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ clutch size</td>
<td></td>
</tr>
<tr>
<td>Ficedula albicollis</td>
<td>Gustafsson &amp; Pärt 1990</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ hatching success,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ number of fledglings,</td>
<td></td>
</tr>
<tr>
<td>Collared flycatcher</td>
<td></td>
<td>↓ clutch size</td>
<td></td>
</tr>
<tr>
<td>Ficedula hypoleuca</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pied flycatcher</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parus atricapillus</td>
<td>Loery et al. 1987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black-capped chickadee</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acrocephalus</td>
<td>Komdeur 1996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sechellensis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seycelles warbler</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melospiza melody</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Song sparrow</td>
<td>Nol &amp; Smith 1987</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↑ Cell-mediated immune response</td>
</tr>
</tbody>
</table>

(continues)
is not possible to identify general patterns concerning species that exhibit senescence in nature relative to those that do not. For one thing, failure to observe a phenomenon of interest likely leads to reporting bias. That is, negative results often go unpublished.

This is not to say that some very thorough studies have failed to find demographic senescence. For instance, a 17-year study of Southern elephant seals tracked 1,650 individually marked pups throughout their lives. Although only 5 percent of the marked animals survived even to age 10, no statistically significant increase in age-specific mortality rate could be detected even as late as between ages 10 and 17 (Pistorius & Bester, 2002). A 30-year study of the common tern (Sterna hirundo) in which about 60,000 chicks have been banded has not found evidence of reproductive senescence even among the oldest 5 percent of birds in the population (Nisbet et al., 2002).

Besides practical difficulties in measuring demographic senescence in nature, there are also theoretical difficulties. For instance, if animals increase reproduction with age, and there is a tradeoff between reproduction and survival, as predicted by Williams (1957) and empirically verified by many others (e.g., Koivula et al., 2003; Orell & Belda, 2002), then declining survival with age might reveal nothing more than increasing reproduction rather than an independently deteriorating internal state. Although evolutionary theories of senescence, such as antagonistic pleiotropy and mutation accumulation (Medawar, 1952; Williams, 1957), do not specifically predict a decline in fertility or increase in death risk alone with age, they do predict declining reproductive expectations with age. Reproductive expectations incorporate both reproduction and survival. This is why Partridge and Barton (1996) suggest Fisher’s “reproductive value,” which utilizes both reproduction and survival to quantify expected current and future reproduction as the best metric for assessing senescence. An interesting case illustrating the problem and how such an analytical approach may be useful is Pugesek’s long-term study of California gulls (Pugesek, 1987; Pugesek & Diem, 1990; Pugesek et al., 1995). Older birds in this study

<table>
<thead>
<tr>
<th>Species</th>
<th>Survival senescence</th>
<th>Reproductive senescence</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthus spinoletta</td>
<td></td>
<td>↓ second clutch size</td>
<td>Askenmo &amp; Unger 1986</td>
</tr>
<tr>
<td>Rock pipit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geospiza conirostris</td>
<td>Grant &amp; Grant 1989</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large cactus finch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetrao tetrix</td>
<td></td>
<td>↓ copulation success (♂)</td>
<td>Kruijt &amp; de Vos 1988</td>
</tr>
<tr>
<td>Black grouse</td>
<td></td>
<td>↓ reproductive output</td>
<td></td>
</tr>
<tr>
<td>Accipiter nisus</td>
<td>Newton &amp; Rothery 1997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European Sparrowhawk</td>
<td></td>
<td>↓ clutch size, ↓ number of young/nest, ↓ egg size</td>
<td>Newton 1988</td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ annual production of young, ↓ reproductive value</td>
<td>Newton and Rothery 1997</td>
</tr>
<tr>
<td>Aegolius funereus</td>
<td></td>
<td>↓ clutch size (♂)</td>
<td>Laaksonen et al. 2002</td>
</tr>
<tr>
<td>Tengmalm’s owl</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9.2 (Cont’d)
died at higher rates than younger birds, hence by the standard criterion of survival senescence, they aged. However, older birds also raised more chicks to fledging than did younger birds. Feeding and protecting young in the nest is energetically taxing and physically risky. That is, there is a cost to reproduction. Controlling for reproduction by comparing young (3- to 10-year-old) birds with older (11- to 17-year-old) ones that fledged the same number of young, no difference in yearly survival could be detected (Pugesek, 1987).

A potentially useful way of envisioning demographic senescence is to assess the maintenance of adaptive tradeoffs between reproduction and somatic survival as they occur in young adults. For instance, Broussard and colleagues (2003) found that the oldest Colombian ground squirrels (*Spermophilus colombianus*) in their study population were more likely than younger age classes to experience reproductive failure. Moreover, young females failing to reproduce regained healthy body mass lost during their reproductive attempt, whereas old females did not. Thus, the adaptive tradeoff between reproductive and somatic investment appears to deteriorate in old age in the Colombian ground squirrel. It is not unexpected that tradeoffs like other adaptive traits deteriorate with age and decaying physiology. This type of deterioration may provide a better indication of the effects of physiological degeneration than purely demographic parameters. Moreover, such deteriorating tradeoffs could potentially produce physiological/demographic patterns that we do not normally recognize as senescence (Blarer et al., 1995).

**B. Nondemographic Measures of Senescence**

The failure to find demographic evidence of senescence does not necessarily mean that animals are not experiencing it. As previously mentioned, environmental or behavioral factors may overwhelm subtle demographic indicators; therefore, physiological markers of functional decline might serve as alternative or complementary sources of information. Note that the indicators used here only need document that later life decline in function has occurred. They do not need to be useful as markers of the rate of functional loss. For instance, studies of three bird species have presented evidence of a decline in immune function with advancing age (Cichón et al., 2003; Reid et al., 2003a; Saino et al., 2003). Although the long-held view that aging is accompanied by a monolithic decline in immune function is gradually being supplanted by a more nuanced view that aging alters immune response in complex, potentially adaptive ways (Effros, 2001), it is probably still a safe generalization that deteriorating resistance to infectious agents is a hallmark of physiological senescence. Indeed, autopsy data on humans implicated infections in the deaths of a majority of people older than 80 years in a Japanese population (Horiuchi & Wilmoth, 1997).

Declining immunity with age seems widespread among species. It is well documented in humans and laboratory rodents of course, but it has also been reported in song sparrows (Reid et al., 2003a) and rhesus macaques (Coe & Ershler, 2001). Particularly interesting field studies combine assessment of immune competence with demographic parameters. Intriguingly, in barn swallows, humoral immunity as measured by antibody response to injection of Newcastle Disease Virus declined in females 3 years old or older relative to 1-and 2-year olds (Saino et al, 2003), but no such decline was observed in males. Survival senescence could only be detected in birds age 5 and older, whereas reproductive senescence as measured by number of fledglings produced began to decline by age 4. Thus, the decline in female immune function preceded indicators of
It might therefore be assumed that immune function decline is a more sensitive indicator of senescence than demographic parameters. However, the situation is not so simple. In the collared flycatchers, a much different pattern is seen. Humoral immunity (response to injected sheep red blood cells) did not decline significantly until 5 to 6 years of age, whereas reproductive and survival performance declined after age 3 (Cichón et al., 2003; Møller & De Lope, 1999). In this case, demographic senescence preceded immunosenescence. Thus, one can’t generalize about the sensitivity of immune markers of physiological senescence in wild animals. It is most useful to have both physiological and demographic measures of senescence whenever possible.

A few studies have adduced age-related declines in body mass as evidence of senescence (Berubé et al., 1999; Derocher & Stirling, 1994; Mysterud et al., 2001). The general sensitivity of body mass decline as an indicator of aging is not clear. There is substantial variability among laboratory mouse sexes and genotypes in the timing of body mass decrease relative to patterns of demographic senescence (Turturro et al., 1999). However, in the laboratory, with ad lib feeding and minimal energetic demands, body mass may decline at a more advanced age compared to the wild. Senescence in nature can affect body mass by decreasing the ability to obtain, process, and store food (Ericsson et al., 2001). For instance, tooth wear may be a significant contributor to age-related mortality in some mammals (Ericsson et al. 2001; Skogland, 1988). Tooth wear increases markedly in roe deer after the age of 7 and coincides with a decrease in survival rate (Gaillard et al., 1993). Indeed, the strongest evidence for changes in body mass relating to senescence are when simultaneous changes occur in demographic parameters. Thus, both litter size and body mass increase for female polar bears (Ursus maritimus) until ages 14 to 16, and thereafter both decline (Derocher & Stirling, 1994). However, this is not always the case, as shown by bighorn ewe body mass, which began declining at about age 11 in one study, about three years before a reduction in fertility was observed but four years after survival senescence became evident (Jorgenson et al., 1997). In sum, the meaning and reliability of age-related body mass decline as a marker of physical senescence is far from clear.

Senescence could potentially be seen as well in a reduced ability to provide nourishment to offspring. Mammary glands themselves may senesce due to a decreasing replicative capacity of mammary epithelial cells with age (Daniel, 1977). In support of this phenomenon, the lambs of old mouflon (Ovis musimon) ewes suckle less frequently, decrease total suckling time, and spend more time grazing relative to lambs of younger ewes (Réale & Boussès, 1995).

Other nondemographic indicators of senescence exist in specific instances. For instance, the ability to defend a territory or other more mobile resource such as a harem may decline with age. On the Scottish island of Rhum, the reproductive success of red deer stags declines after age 11 due to a decreased ability to fight and control harems (Clutton-Brock et al., 1979). However, this is not always true even in harem-defending species, as shown by the absence of a similar phenomenon in the American pronghorn (Byers, 1998). From the female side, calf mortality increases for red deer hinds over 12 years of age. Survival of calves depends partially on its mother’s dominance rank and access to a good-quality home range, both of which decline as females age (Clutton-Brock, 1984). Therefore, change in dominance rank may serve as a marker of senescence in some species.
The ability to defend a territory declines with age in a number of species. In the small passerine bird, the great tit (Parus major), male territory size increases until 4 years of age and then decreases from 5 years throughout the rest of life. Some old great tits forego defending a territory altogether, although they still manage some breeding by mating with females late in the year (Dhondt, 1971). Male black grouse (Tetrao tetrix) defend territories that females attend only to breed. Those 6 years and older seldom occupy the highest-quality territories on display grounds, likely because they cannot defend them (Kruijt & de Vos, 1988). By contrast, young male greater white-lined bats (Saccopteryx bilineata) lurk on the periphery of older males’ breeding territories, copulating with females opportunistically. No reports exist of males so old that they can no longer defend their territory. This may be a case where significant senescence is not seen in the wild (Heckel & Von Helversen, 2002).

Hormones may serve as markers of senescence as well. Plasma testosterone level in Misaki feral horses correlates with age and harem size (Khalil et al., 1998). If senescent males no longer can mount and maintain appropriate testosterone levels during the breeding season, their ability to recruit a harem may become compromised. Thus, hormonal and reproductive aging decline synchronously.

Sometimes demographic senescence can be observed even when other well-characterized indicators of senescence cannot be detected. For instance, common terns exhibit survival senescence, yet no decline can be detected over their lifetime in immunological, endocrinological, or reproductive aging (Apanius & Nisbet, 2003).

In birds, where renewal of feathers must occur annually after molting, one might expect that the ability to produce long, healthy feathers might decrease with age. Indeed, Møller and De Lope (1999) found that both tail and wing length decreased in barn swallows 5 years old or older. In addition, these aged birds had less perfect symmetry between left and right wing and tail feathers compared with younger birds.

Finally, as biochemical markers of aging become available, studies that involve recapturing individuals can focus on directly measuring deteriorating physiology from blood or urine samples, though assays will need to be validated for the particular study organism. Regardless of the measures selected to test for senescence in natural populations, it is critical to have detailed understanding of the dynamics of the study population, the physiology of the study organism, and the life history trade-offs individuals face over their life span.

### III. Patterns of Senescence

Can we glean anything about general patterns of senescence in wild populations from the information currently available? The first and most important point is that one should exercise caution when accepting and comparing published data on senescence in the wild (Nisbet, 2001). Apparent senescence can arise from short-term climatic events such as El Niño or increasing population density over the course of a study. The literature must be interpreted with sensitivity to variation among studies in the methods used to collect and analyze demographic data. Second, senescence, measured as changes in life-history traits and physical function associated with old age, is a complex phenomenon. It can defy our expectations about which species should and should not show signs of senescence and at what point during the life course indicators of senescence might become apparent. It is also critical to remember that patterns of senescence are not necessarily species-specific but may vary
among populations within a species (Austad, 1993).

Generally, senescence is expected to be more apparent among species of low mortality rates and long life span because senescence reduces average life span more in such species (Ricklefs, 1998) and because high mortality rates significantly reduce the chances of any individual reaching a senescent age or remaining alive in a senescent state for long enough to be detected. However, some short-lived animals with high mortality rates have been demonstrated to experience senescence. For example, small rodents typically experience high mortality rates and small mice rarely live more than one year in the wild. Despite this, Promislow (1991) and Millar (1994) both found evidence of actuarial senescence in populations of deer mice (Peromyscus maniculatus), and Morris (1996) found evidence of reproductive senescence in white-footed mice (Peromyscus leucopus; see Table 9.1).

Is there a general age or developmental stage that is a threshold for the onset of senescence generally? Evolutionary senescence theory (Williams, 1957) suggests that aging should begin at about the time of sexual maturity. Although observations consistent with the theory have been reported for a German population of pied flycatchers (Sternberg, 1989) as well as Florida scrub jays (McDonald et al., 1996), it was not found in any of five populations of three species of ungulates (roe deer, bighorn sheep, Pyrenean chamois) (Loison et al., 1999). In a particularly thorough long-term (12 to 22 years) study, age at first reproduction in all five populations was 2 years, yet annual female survival rate high remained high (>0.9) with no statistical decline at least until age 7. Another potential stage at which senescence could conceivably begin is “social maturity,” the age at which animals are capable of actively competing for reproductive opportunities, often somewhat later than the time of sexual maturation. McElligott and colleagues (2002) evaluated such a situation in male fallow deer (Dama dama). They found that social maturation began at 4 to 5 years, yet annual survival rate did not begin to decrease until age 9. An even more complex situation is found in bighorn sheep (Ovis canadensis), in which the onset of survival senescence did not coincide with age of first reproduction in females, but did in males [i.e., rutting decreased male survival rate] (Jorgenson et al., 1997). Thus, no clear pattern about the timing of senescence’s onset emerges from field data.

One point that does emerge though is that patterns of mortality dynamics in nature can be dramatically different than in captive populations. For instance, animals in captivity often exhibit Gompertzian mortality dynamics (Finch, 1990)—that is, a monotonically log-linear increase age-specific mortality rate—but many long-lived mammals and birds in nature have a more phase-specific relationship between survival, reproduction, and age. That is, survival and reproduction are relatively low and variable in young adults, both increase and remain steady in prime-aged animals, and then both decrease again in oldest individuals (Berubé et al., 1999; Caughley, 1966; Festa-Bianchet et al., 2003). Some of the difference between captive and wild populations may be explained by the greater phenotypic variability within populations in nature due to some combination of differential genetic endowment, consequent differential access to important resources, and discrete social roles. A case in point is the previously mentioned study of fallow bucks (McElligott et al., 2002). Overall, mortality dynamics fit well with a Gompertz model. However, when the population was subdivided into nonproducers (animals that had ceased to breed later in life) and producers (animals still
actively reproducing) irrespective of age, nonreproducers’ survival declined after age 9 but reproducers showed no change in survival even to much later ages.

A related question concerns whether there are general patterns in the timing of the beginning of reproductive senescence vis à vis survival senescence? To take one well-known example, in humans, age-specific mortality increases from the age of 10 to 11 years in societies with access to modern medicine [Finch, 1990], whereas subtle evidence of reproductive senescence only becomes apparent by about age 30 [vom Saal et al., 1994]. Is this appearance of survival senescence prior to reproductive senescence a general phenomenon? Apparently not. Although such a pattern is seen in female bighorn sheep [Berubé et al., 1999, Jorgenson et al., 1997], European sparrowhawks [Newton & Rothery, 1997], European red deer [Clutton-Brock et al., 1988], baboons and African lions [Packer et al., 1998], the reverse—reproductive senescence preceding survival senescence—is seen in barn swallows [Møller & De Lope, 1999; Saino et al., 2003] and wandering albatrosses [Weimerskirch, 1992]. The difference between onset of reproductive and survival senescence can be substantial. In bighorn ewes, survival senescence appears by age 7 to 8 but reproductive senescence only after 13 years [Berubé et al., 1999; Jorgenson et al., 1997]. Similarly, both male and female red deer (Cervus elaphus) experience survival senescence after 8 years of age, but reproductive senescence does not appear until age 12, as evidenced by declining female fecundity and increasing calf mortality [Clutton-Brock, 1984] and the loss of males’ ability to control harems [Clutton-Brock et al., 1979]. In contrast, wandering albatrosses (Diomedea exulans) exhibit decreased egg size, breeding success, and breeding frequency after age 20 but survival decreases significantly only after age 27 [Weimerskirch, 1992].

Although the previous discussion and our summary tables have primarily focused on species patterns, it is worth recalling that not all populations of a species will necessarily exhibit the same pattern. For example, Sanz and Moreno (2000) found no evidence of declining age-specific survival rate of female pied flycatchers (Ficedula hypoleuca) breeding in central Spain, the southern part of this species’ breeding range, whereas Sternberg (1989), studying the same species in Germany, did observe a steady decline in annual survival after age 1. The former authors suggest that the lack of observed senescence in their population relative to Sternberg’s may be due to its shorter migration route to wintering grounds in West Africa. Loison and colleagues (1999) also found marked differences in senescence patterns among different populations of roe deer and bighorn sheep.

Gaps in our knowledge of senescence patterns in wild populations stem from the limited number of field studies systematically addressing the issue. However, there are also methodological difficulties associated with assessing senescence in natural populations that may continue to make the accumulation of new knowledge slow. Reliable studies need to be based on long-term observations and/or experiments, enhanced by detailed knowledge of the mechanisms affecting age-specific mortality and fecundity, of the life-history tradeoffs at play in the population, and of the causes and dynamics of these tradeoffs [Blarer et al., 1995].

IV. Methodological Difficulties in Evaluating Senescence in Wild Populations

The scarcity of studies on senescence in natural populations testifies to the logistical difficulty of observing and investigating it. Some authors urge readers to be
careful when accepting published data on senescence in wild populations (Gaillard et al., 1994; Nisbet, 2001) and cite a variety of methodological and interpretation problems that must be overcome before we can develop a true understanding of how, when, and why senescence occurs in the wild. Potential methodological problems include the lack of visible, anatomical markers of aging, the use of cross-sectional rather than longitudinal life tables, small sample sizes of individuals at older ages, and the assumption that an undetected individual is a dead individual. Interpretation problems include demographic heterogeneity, assumptions of temporal constancy of environmental and demographic conditions, and the choice of variables to measure senescence. We will discuss some of these problems briefly.

The first question to ask before evaluating any type of senescence in wild populations is whether you can be confident in assigning calendar age to study animals. The gold standard should continue to be unique marking of individuals as close to birth as possible. Even gold standards are not perfect, however, as individual marks such as bands or tags can be lost over time. Frequently, surrogate markers of calendar age, such as growth rings or plumage changes, are used as age estimators, sometimes without extensive validation. In fact, there are very few reliable markers of age in the wild, and even the ones that exist are often species-, population-, and even sex-specific. Thus, their uncritical use can be deceptive. Tooth wear categories, for example, are often used to estimate age in populations of ungulates. However, tooth wear changes have been shown to decelerate with age and differ between male and female Norwegian red deer (Loe et al., 2003). On the other hand, the precision required of an age estimator depends on the precision necessary to answer the research question. Sometimes relative rather than absolute age estimation is sufficient. Even when absolute age estimation is required, rough precision may suffice. For instance, if bowhead whales indeed frequently live 150 to 200 years, as one report suggests, then an age estimator such as amino acid racemization of the eye lens may be sufficiently precise for most demographic analyses, even if its margin of error is as much as a decade (George et al., 1999).

Even well-established age estimators can be flawed. For example, the correlation between growth layers in dentine and age used to establish calendar age of southern elephant seals was constructed with animals only up to 8 years of age. It is known from individually marked individuals that this species can live in excess of 23 years (Hindell & Little, 1988). Whether extrapolation from known to older unknown ages is warranted will depend on the biology of the species and marker in question. Some traditional markers, such as scale rings of fishes, have failed validation tests in at least some species (Nedreaas, 1990), whereas newer techniques, such as growth rings of fish otoliths, have been extensively validated (Cailliet et al., 2001).

Cross-sectional or “snapshot” studies are the easiest and quickest way to investigate senescence in the wild. That is, demographic parameters are estimated from the age distribution of a current population. Indeed, life tables used in human demography are virtually always cross-sectional. Cohort tables, in which a population of individuals born at the same time is tracked throughout life, will generally be preferred for studies of wild populations. Problems with cross-sectional life tables have to do with the assumptions required for using such life tables as surrogates for cohort tables. These include assumptions such as a stationary population age distribution and consistency of mortality patterns over time that are unlikely to be even
If individuals in a population are not identified, even serial snapshots of a single cohort can mislead as a consequence of, say, population heterogeneity. Consider, for instance, a hypothetical cohort of five females that were born the same year (see Figure 9.2). A researcher averages the number of offspring of the females at 1 year of age and then again at 6 years in an attempt to detect reproductive senescence. Only three of the original five females make it to age 6. The results from these two snapshots appear to indicate a decline in reproduction with age—that is, reproductive senescence. However, the reproduction of individual females never changed over time. A more accurate assessment of the population would be that for some reason—a tradeoff between reproduction and survival or genetic heterogeneity or chance—females with low reproductive rates live longer. Without data on survival and number of offspring for individual females, a researcher would have an inaccurate picture concerning reproductive senescence.

Even longitudinal studies are not foolproof, of course. For instance, the length of sampling intervals must be meaningful within the context of the life of the study organism. Sampling wild mice annually would clearly be useless given that their average longevity is only 3 to 4 months (Phelan & Austad, 1989). In addition, there is the question of how long a longitudinal study should last in order to obtain accurate measures of demographic parameters and actually detect individuals that reach a senescent stage. Climatic variability and changes in population density are only two possible confounders of field demography. Consider a population of fulmars in England that has been tagged and tracked since 1950 (Dunnet & Ollason, 1978). Based on data from 1950 to 1962, researchers calculated the average adult longevity to be 15.6 ± 1.9 years (Dunnet et al., 1963). Based on data from 1950 to 1970, adult longevity increased to 25.0 ± 4.3 years for males and 22.3 ± 4.2 years for females (Cormack, 1973). Finally, based on even later data (1958 to 1974), which now included only birds wearing a style of leg bands that were less likely to fall off, adult longevity was estimated at a substantially shorter 19.9 ± 1.8 years (Dunnet & Ollason, 1978). Although the analytical methods differed somewhat in each of these studies, the increasing study length combined with improved survival data from better leg bands account for at least part of the progressive changes in longevity estimates.

Although the length of a longitudinal study is important for evaluating trends in senescence, so is the number of individuals marked in the population. Needless to say, this is not a problem specific to field studies. Sample sizes limit demographic analyses in both field
and lab. For instance, if only 50 individuals are included in a study, a real 1 percent difference in mortality rate is impossible to detect (Promislow et al., 1999). Also, because fewer and fewer individuals are alive at later ages, the statistical power to determine late-life demographic trends inevitably decreases. For instance, in a typical study, only 4 percent of the willow tits (Parus montanus) reached an age at which demographic senescence could be detected (Orell & Belda, 2002).

Population heterogeneity can also obscure underlying demographic patterns (Carnes & Olshansky, 2001; Festa-Bianchet et al., 2003; McDonald et al., 1996; Service, 2000; Vaupel & Yashin, 1985), as the previous example with fallow deer indicated. This interpretation problem is not unique to field studies but may be exaggerated in nature relative to the laboratory. Such heterogeneity can arise from genotypic or phenotypic differences among individuals. Consider that it is likely that only the strongest, healthiest individuals in a population will make it to old age. As individuals of low quality are eliminated from the population, we see a resulting increase in survival rate. This increase could be interpreted as negative senescence. Yet it does not represent individuals becoming more robust as they get older. It is the result of a selection process that has changed the nature of the sample being measured and it thus may obscure a different underlying pattern if the same individuals were followed longitudinally. For instance, McDonald and colleagues (1996) found evidence of survival senescence in Florida scrub jays. The evidence was most clear when social role was controlled for and a homogenous subpopulation of high-quality individuals with the highest annual fledging rates and greatest longevity were identified. Another example in which phenotypic heterogeneity may be misleading with respect to demographic senescence may been seen in snowshoe hares (Lepus americanus), which experience dramatic and regular population cycles. That is, periods of very high hare density are followed by population crashes to low densities, which in turn are followed by rebounds to high density again. During a 16-year breeding program, Sinclair and colleagues (2003) found that genetic lineages established from females of the high-density phase had significantly lower reproductive output and longevity than lineages established from females of the low-density phase. Furthermore, high-phase females showed declining reproductive output with age, whereas reproductive output remained constant throughout life in low-phase females.

Despite the manifold complications associated with the study of senescence in nature, an increasing number of studies are addressing the issue, with progressively more sophisticated techniques.

V. Conclusions

Studies of senescence in natural populations of mammals and birds have improved significantly in recent years with the introduction of new methodological and analytical techniques and the continuation of long-term research projects. Clearly, senescence occurs in the wild, although just as clearly it is not ubiquitous. Only by increasing the breadth of study species and focusing specifically on the question of senescence will we eventually be able to comprehensively assess interspecies patterns of senescence. Further research is also necessary to enhance our understanding of variation in senescence patterns among populations of the same species. The future of this line of research will be based on longitudinal data from large samples of marked individuals in populations that are well characterized and organisms for which we have a sound
understanding of physiology, behavior, and life-history options, constraints, and tradeoffs.

Acknowledgments

We thank Ed Masoro, Daniel Promislow, Victor Apanius, and an anonymous reviewer for helpful comments on an earlier version of this manuscript. Preparation of the manuscript was supported in part by an Ellison Senior Scholar Award and NIH grant AG022873.

References


Chapter 10

Biodemography of Aging and Age-Specific Mortality in *Drosophila melanogaster*

James W. Curtsinger, Natalia S. Gavrilova, and Leonid A. Gavrilov

I. Introduction

For the last 15 years there has been a high level of interest in combining the methods of biology and demography to investigate aging in experimental populations. The hybrid field of biodemography addresses a wide range of questions about aging organisms and aging populations, and also attempts to provide insights into human aging [Wachter & Finch, 1997]. A handful of issues have preoccupied the nascent field: To what extent are the genetic phenomena that influence life histories age-specific in their effects? How malleable are the patterns of survival and death among the oldest organisms? Why do populations often exhibit mortality plateaus? How have observed survival patterns evolved under the influence of mutation and natural selection? To what extent do survival patterns in populations reflect underlying changes in individual organisms? All of these questions are challenging, and none fully answered yet. Addressing them requires a set of analytical techniques that are commonplace to demographers but foreign to most biologists. Here we review some basic analytic methods from demography and lay out essential biological methods and questions, hoping to introduce both biologists and demographers to the hybrid field.

The integration of genetic and demographic methods requires an experimental system that is genetically defined and amenable to large-scale population studies. The fruit fly *Drosophila melanogaster* is an obvious candidate, being one of the premiere experimental systems for basic research in genetics. The genome is completely sequenced, and the flies can be reared in large numbers (tens of thousands of organisms). The nematode *Caenorhabditis elegans* and some yeast species also have those desirable characteristics, but other standard experimental systems do not. The genetics of house mice (*Mus musculus*) is an important and growing area of research, but large-scale population studies with rodents are impractical. Demography of medflies
(Cerititus capitata) and several parasitic wasp species has been investigated in large experimental populations (Carey, 2003), but those systems are genetically undefined. An interesting feature of Drosophila as an experimental model is the similarity of its mortality kinetics to that of humans, first noted by Raymond Pearl (1922). Both species have a relatively short period of high initial mortality, followed by a relatively long period of mortality increase, and then deceleration at advanced ages (although the period of mortality deceleration and mortality plateau in Drosophila is longer than in humans).

Here we concentrate on D. melanogaster, a holometabolous insect. Larvae hatch from eggs about 24 hours after laying, feed voraciously for a week, and then pupate. Adults emerge from the pupal case after a few days of metamorphosis and are sexually mature within 24 hours. In the wild, D. melanogaster adults probably live one to two weeks. In laboratory culture, flies are normally maintained on a two-week generation schedule but can live much longer as adults. In a typical outbred population, adults survive 30 to 50 days on average, depending on temperature and other environmental conditions. Inbreeding and increased temperature reduce mean adult life spans, while artificial selection for increased life span is capable of doubling it. Maximum adult life spans observed in large experiments typically exceed 100 days. There is no precise definition of young, middle-aged, or old adult flies. At two weeks after emergence, metabolic rate and gene expression reach low levels characteristic of remaining adult life (Tahoe et al., 2004; Van Voorhies et al., 2003, 2004). For females, old age in flies is probably best understood as the age after egg laying has ceased, usually 40 to 60 days after emergence, depending on genotype and environmental conditions.

A. Collection of Survival Data
Survival experiments with laboratory populations of Drosophila are typically longitudinal, large scale, and complete. That is, age-synchronized cohorts consisting of thousands or tens of thousands of experimental animals are established with newly emerged adults and are observed over time. As the cohorts age, dead animals are removed, counted, and recorded on a daily basis. Observations continue until the last fly dies, typically around 100 days after emergence (depending on genotype and sample size; see below). Experimental populations are maintained under controlled environmental conditions, including temperature, light cycle, and humidity. Initial population density is also controlled, at least approximately; in smaller experiments, exact numbers of flies are counted, whereas in larger experiments, density is approximated by volume or weight of anesthetized flies (one large female weighs ~ 1mg., whereas males are typically ~30 percent smaller). Experimental populations are often housed in cages of one to several liters in volume, each holding up to a thousand individuals, but half-pint milk bottles and finger-sized glass vials are sometimes used. There is always fresh fly food in the containers, which serves as both an oviposition medium and a source of nutrition for adults and larvae. Frequent replacement of the medium and changing cages prevents unwanted recruitment of new adults into experimental populations.

Populations used for survival studies typically consist of males and females in approximately equal proportions when experiments are initially set up, but because of differential survival, the sex ratio changes over time. In mixed-sex population cages, females actively reproduce and generally exhibit shorter average life
spans than males (Curtsinger & Khazaeli, 2002; Curtsinger et al., 1998; Fukui et al., 1993, 1995, 1996; Khazaeli & Curtsinger, 2000; Khazaeli et al., 1997; Pletcher, 1996; Resler et al., 1998). Because flies reach sexual maturity soon after emergence, mating behavior begins almost immediately in mixed-sex populations. It is possible to study the survival characteristics of unmated flies in single-sex populations by anesthetizing newly emerged adults and then sorting the sexes under a dissecting microscope when cohorts are initially established (Miyo & Charlesworth, 2004; Semchenko et al., 2004).

There is significant uncontrolled environmental variation that affects death rates in experimental populations of Drosophila. The magnitude of the variation is perhaps underappreciated. For instance, it is not unusual to see four- or five-fold variation in individual life spans among flies of the same genotype sharing the same food and population cage. This is not a peculiarity of fly life spans; biologists have long recognized that quantitative traits vary between organisms, even if they are genetically identical and reared under carefully controlled conditions (for a review, see Finch & Kirkwood, 2000; Gavrilov & Gavrilova, 1991). Because of this irreducible variation, which is not well understood, survival experiments should be highly replicated, in some cases involving hundreds of populations. Ideally, data from genotypes or treatments that are to be contrasted are collected simultaneously in order to avoid confounding uncontrolled environmental variations with treatment or genotype effects.

B. Data Analysis: Mean Life Span and Survivorship

The central problem in survival analysis is to summarize and interpret large amounts of information hidden in the survival data. Raw data consist of estimated ages at death. Mean life span, the arithmetic average survival time, has intuitive appeal as a descriptor of survival ability, but the information contained in that summary statistic is limited. The most critical limitation in the present context is that the mean gives little information about the age-specificity of survival patterns. Two cohorts could have very similar means but experience vastly different life histories. For instance, if one population suffers mortality only at middle age, whereas a second experiences mortality equally and exclusively at early and late ages, mean life spans in the two populations will be similar. Maximum observed life span is also frequently reported but is similarly uninformative about age-specific events.

The central conceptual tool for organizing and analyzing age-specific aspects of survival data in experimental populations of Drosophila and other species (indeed, other objects) is the cohort life table. It is interesting that Drosophila was the second species, after humans, for which such demographic life tables were constructed (Pearl & Parker, 1921). The essential features of the life table are that age classes are defined by sampling intervals, and for each age class (life table row) specific variables (life table columns) are estimated. The first variable is the fraction of the total population dying while in age class $x$, denoted $d_x$. The distribution of $d_x$, a typical example of which is shown in Figure 10.1a, is approximately bell-shaped but not symmetrical, in contrast to the normal curve. The long right-skewed tail represents the oldest survivors of the cohort and is observed even in genetically homogeneous populations. The second variable, survivorship, is represented as $l_x$ and is defined as the probability of survival from the beginning of the experiment until the beginning age interval $x$. That probability is estimated by the proportion of the initial cohort that remains alive at age $x$. Survival curves, which show plots of $l_x$ versus $x$, start at
100 percent and decline to zero at the age when the last animal in the cohort dies (see Figure 10.1b). Survival curves have built-in smoothing because they are non-increasing (the proportion of the initial cohort remaining alive at age $x$ can only be the same or lower at age $x + \Delta x$). For this reason, even relatively small cohorts produce smoothly declining survivorship curves. Life-table values of $l_x$ and $d_x$ are related as follows: $l^{x+\Delta x} = l_x - d_x$, where $\Delta x$ is the length of the sampling (age) interval, typically equal to one day for fly experiments. It is important to emphasize that both $l_x$ and $d_x$ are cumulative indicators that depend on preceding death rates. Events early in the life history, such as a temporary epizootic, can affect survivorship and the fraction dying in later age classes, even in old age. In this sense, $l_x$ and $d_x$ reflect the survival history of the cohort up to and including age $x$.

C. Data Analysis: Probability of Death and Mortality Rate

Unlike survivorship and fraction dying, which have “memory,” some other life-table variables are noncumulative and better suited to detecting age-specific effects. Age-specific probability of death ($q_x$) is defined as the conditional probability of dying in the interval $\Delta x$ for individuals that survive to the beginning of interval $x$. It is estimated as the number of deaths that occur in age class $x$, divided by the number of individuals entering class $x$. An example of age-specific probability of death is shown in Figure 10.1c. Note that in this particular example, the age-specific probability of death grows monotonically with age up to an advanced age and then levels off, a phenomenon discussed in detail later.

Although probability of death is useful and intuitive, it has limitations. The main problem is that the value of $q_x$ depends on the length of the age interval $\Delta x$ for which it is calculated, which hampers both analyses and interpretation. For example, one-day probabilities of death may follow the Gompertz law of mortality, but probabilities of death calculated for other age intervals with the same data may not (Gavrilov & Gavrilova, 1991; le Bras, 1976). A meaningful descriptor of the dynamics of survival should not depend on the arbitrary choice of age intervals. Another problem is that, by definition, $q_x$ is bounded by unity, which makes it difficult to scale the variable for studies of mortality at advanced ages.

A more useful indicator of mortality is the instantaneous mortality rate, or hazard rate, $\mu_x$, which is defined as follows:

$$\mu_x = \frac{dN_x}{N_x d\Delta x}$$

where $N_x$ is the number alive at age $x$. The hazard rate does not depend on the length of the age interval; it reflects instantaneous risk of death. It has no upper bound and has the dimension of a rate (time$^{-1}$). One of the first empirical estimates of hazard rate $\mu_x$ was proposed by Sacher (1956):

$$\mu_x = \frac{1}{\Delta x} \left( \ln l_{x+\frac{\Delta x}{2}} - \ln l_{x-\frac{\Delta x}{2}} \right)$$

$$= \frac{1}{2\Delta x} \ln \frac{l_{x+\Delta x}}{l_{x-\Delta x}}$$

This estimate is unbiased for slow changes in hazard rate (Sacher, 1966). A simplified version of the Sacher estimate (for small age intervals equal to unity) is often used in biological studies of mortality: $\mu_x = -\ln(1-q_x)$ (see Carey, 2003) and assumes constant hazard rate in the age interval.

The Cutler-Ederer (1958) estimate (also called the actuarial hazard rate) is based on the assumption that deaths are
uniformly distributed in the age interval and that all cases of withdrawal (censoring) occur in the middle of the age interval:

$$\mu_x = \frac{d_x}{\Delta^x \left[ l_x - \frac{c_x}{2} - \frac{d_x}{2} \right]}$$

Here, $c_x$ is number of censored individuals during the age interval (for example, number of flies accidentally escaping the cage during food replacement). The hazard rate is measured at the midpoint of the age interval. Gehan and Siddiqui (1973) used Monte Carlo simulation to show that for samples less than 1,000,
the Sacher method may produce biased results compared to the Cutler-Ederer method, whereas for larger samples, the Sacher estimate is more accurate. The advantage of the Cutler-Ederer estimate is its availability in standard statistical packages [such as SAS and Stata], which compute actuarial life tables. Despite the apparent differences between Cutler-Ederer and Sacher estimates, the methods produce very similar results for real data [see Figure 10.1d]. Note that the mortality curve, depicting $\mu_x$ as a function of $x$, describes survival events in true age-specific fashion. It clearly illustrates the rate of actuarial senescence, usually defined as the slope of the mortality curve, and is particularly useful for examining details of death rates among the oldest survivors of a cohort. In contrast, the details of shape in a survivorship curve as it approaches the x-axis are generally indistinct [but see Pearl & Parker’s method, described below].

The differences between survivorship and mortality are fundamental. The former depends on all previous cohort history, whereas the latter reflects risk specific solely to the age group under study. This distinction has often been misunderstood or overlooked by biologists. Rose’s [1991] influential text on evolutionary biology of aging contains dozens of figures, extensive discussion of age-specific life-history phenomena, and not a single depiction of a mortality curve, either experimental or theoretical. Similarly, Kirkwood’s [1999] general text on causes of aging gives considerable notice to age-specific phenomena but employs survivorship rather than mortality throughout. In an otherwise excellent paper on chromosomal mapping of genes that influence mean life spans in Drosophila, Nuzhdin and colleagues [1997] test an evolutionary model of senescence by examining age-specific variance in $l_x$, when the issue is clearly variance in $\mu_x$.

Perhaps the most common misunderstanding among biologists about survivorship and mortality is the widespread assumption that rates of senescence can be easily seen in the slopes of survivorship curves. The apparent or actuarial rate of senescence, defined as the rate at which risk of death increases with age, is precisely reflected in the slope of the mortality curve: a steep slope indicates rapid actuarial senescence, a shallow slope indicates negligible senescence, and a zero slope indicates no senescence. Of course, the slope of the survivorship curve bears a mathematical relationship to the slope of the corresponding mortality curve, but not one that is easily grasped by visual inspection. The problem is that even populations that experience no apparent senescence (constant probability of death at all ages) will exhibit exponentially declining survivorship with increasing age. Thus, information about the rate of senescence is present in a survivorship curve only as a deviation from the exponential, a quantitative measure that is not well suited to casual inspection. Pearl and Parker [1924] addressed this problem by examining survivorship in semi-logarithmic plots. This approach may be useful in defining periods of mortality leveling-off [mortality plateaus]: survivorship curves in semi-logarithmic scale should be linear if mortality is constant. Economos [1979, 1980] used this method for demonstrating non-Gompertzian mortality kinetics at advanced ages, but the technique has not been widely used in recent years.

There are probably several reasons that biologists in some fields have not, until recently, adequately appreciated the information that can be gained by estimating mortality rates. Survivorship curves have intrinsic smoothing, as mentioned above, whereas mortality curves tend to be jumpy. For a single data set plotted both ways, the mortality estimation makes the data look noisy, whereas
the survivorship curve gives an appearance of orderly behavior. Accurate estimation of age-specific mortality rates requires larger sample sizes than those needed for estimating means or survivorship but provides extra sensitivity in studies of short-term response to phenomena such as heat shock (Khazaie et al., 1997) and dietary restriction (Mair et al., 2003; Pletcher, 2002). The sample-size requirement is especially critical for the oldest ages; large initial cohort sizes are required in order to have adequate numbers of animals alive for estimation of death rates at the older ages.

In the 1920s, Raymond Pearl, an early advocate of biostatistics and experimental investigation of populations, published a series of papers on *Drosophila* life spans that employed relatively large sample sizes. For instance, Pearl and Parker (1924) collected survival data on about 4,000 flies from two strains. Since the 1950s, radiobiologists have routinely employed large sample sizes to estimate mortality rates in survival studies with experimental organisms. However, in spite of those pioneering efforts, up until around 1990, it was standard practice among experimental gerontologists, evolutionary biologists, and geneticists to employ small populations in studies of *Drosophila* survival, typically on the order of 50 to 100 animals per experimental treatment or genotype. Such sample sizes sufficed to give reasonably accurate estimates of mean life spans and aesthetically pleasing survivorship curves but provided virtually no information about death rates in old age.

Sample size requirements will depend on the specific question being asked. For accurate estimates of hazard rates, it is necessary to have some events (deaths) in each age interval. At younger ages, when mortality rates are low, it would be desirable to have at least one death in each observation interval. In small samples there might be no deaths during some intervals, in which case intervals will have to be combined and the accuracy of hazard rate estimation will decline. Thus, the minimum sample size of experimental populations for hazard rate studies may be estimated on the basis of expected risk of death during younger ages, when mortality is low.

For example, if the expected risk of death is 1 per 1,000 during a one-day period, then the sample size should be at least 1,000. If mortality at younger ages is higher, then smaller sample sizes will suffice. This rule of thumb does not apply to studies of mortality deceleration and leveling-off. This phenomenon happens later in life, after a significant part of population has died and the remaining number of animals is a small fraction of the initial cohort. The empirical rule here may be to have at least 50 animals alive at the age when mortality deceleration starts so that hazard rate estimations would not be distorted by small numbers of deaths. If one is interested in short-term effects of caloric restriction or other interventions on mortality kinetics at middle ages close to the modal life span, then much smaller sample sizes may be sufficient because numbers of organisms at risk and numbers of deaths will be substantial.

D. Smoothing and Model Fitting

Two approaches are commonly used to describe trends in the (often noisy) data on age-specific mortality. One approach is to apply a non-parametric smoothing procedure. For data organized in the form of a life table, smoothing can be accomplished by widening the age intervals. If times to death for each individual in the sample are known with reasonable accuracy, and/or small sample size does not allow construction of a conventional life table, then the method of hazard rate smoothing using kernel functions may be more appropriate (Ramalu-Hansen,
1983). The latter method is more computationally complex, although special routines are available now in SAS and Stata. Applying methods of non-parametric smoothing decreases statistical noise and facilitates visual inspection of mortality plots but does not allow quantitative analysis of life-span data.

The second major approach for summarizing and simplifying mortality estimates is parametric model fitting, which allows researchers to describe the observed mortality kinetics using a small number of parameters of a specified mortality model. Although there are many possible models in the literature, three are widely used by biologists. The venerable model of Gompertz (1825) specifies exponentially increasing hazard rate with increasing age:

$$
\mu_x = Ae^{Bx}
$$

where $A$ is initial mortality rate, $e$ is the base of the natural logarithms, and $B$, the slope parameter, controls the rate at which mortality increases with age. Estimates of $A$ in laboratory populations of *D. melanogaster* are typically in the range 0.005 to 0.010 per day, whereas $B$ often lies in the range 0.04 to 0.10 per day (Fukui *et al.*, 1993). The Gompertz model produces a straight line in semi-log plots of hazard rate versus age, with the y-intercept estimating the initial mortality rate and the slope estimating the rate of senescence. The aging rate is sometimes summarized by the mortality rate doubling time (MRDT), defined as $\ln(2)/B$. However, this measure has limited applicability to *Drosophila* because of non-Gompertzian mortality dynamics at advanced ages; in particular, as $B$ approaches zero in old age, the MRDT approaches infinity.

A second widely used model is the logistic, which is motivated by the possibility that individuals in the same population can have different frailties (age-dependent chances of death).

Differences in frailty might be innate and fixed throughout life, or modified over the life history. Strehler and Mildvan (1960) showed that when there is such heterogeneity, the observed population mortality pattern deviates from the underlying mortality for individuals. Following Beard (1963), the observed mortality in the population is

$$
\mu_x = \frac{Ae^{Bx}}{1 + \sigma^2 \Lambda(x)}
$$

where $A$ and $B$ are as defined in the Gompertz model, $\sigma^2$ is the variance for frailty in the population, and $\Lambda(x) = (A/B)(e^{Bx} - 1)$. Note that when $\sigma^2 = 0$, there is no heterogeneity in the population, and the logistic reduces to the Gompertz model. However, if $\sigma^2 > 0$, then the logistic curve increases exponentially at early ages and plateaus at more advanced ages (as $x$ becomes large, $\mu_x$ approaches $B/\sigma^2$). Yashin and colleagues (1994) showed that this model applies under two biologically different circumstances: when individuals possess a fixed frailty from birth that differs from that of other individuals, and when all individuals start life with identical frailties but then randomly acquire differences in frailty during adulthood.

A third model used by biologists is also motivated by the observation that mortality data often exhibit plateaus at older ages. This approach involves fitting two curves to the mortality data. Curtsinger and colleagues (1992) proposed a two-stage Gompertz model, in which a Gompertz curve is fit to the data at young ages up to some breakpoint age, and then a second Gompertz curve with shallower slope is fit to the older ages. This model includes five parameters: two intercept and two slope parameters for two Gompertz curves, and a fifth parameter for the breakpoint. Zelterman and Curtsinger (1994, 1995) applied the method to fly data, and Vaupel and colleagues (1994) used it for nematodes.
Drapeau and colleagues (2000) employed a similar method, except older ages were fit to a linear rather than exponential curve. It should be noted that mortality trajectories following the Weibull (power) law of mortality may resemble a two-stage Gompertz model in semi-log coordinates (see Chapter 1 in this volume).

The two major methods of parameter estimation for nonlinear models are maximum likelihood and nonlinear least squares. The maximum likelihood approach is based on maximizing the likelihood function, or the probability of obtaining a particular set of data given the chosen probability model. Maximum likelihood methods provide unbiased and efficient parameter estimates for large data sets (though the estimates may be heavily biased for small samples). Another advantage is that maximum likelihood generates theoretically more accurate confidence bounds for parameter estimates. An important property of maximum likelihood for survival data is that censored observations can be readily introduced (see Filliben, 2004). The limitation of this method is the need for specifying the maximum likelihood equations for each particular function not implemented in the standard statistical software packages, which often is not trivial. Standard statistical packages provide maximum likelihood estimates for a limited number of models. For example, the Stata package has a procedure for maximum likelihood estimation of Gompertz and logistic models. Maximum likelihood estimation of Gompertz, Gompertz-Makeham, logistic, and logistic-Makeham models is implemented in WinModest, a program written and distributed by S. Pletcher (Baylor College of Medicine, Houston) specifically for calculating basic statistics, fitting mortality models to survival data, and partitioning mean longevity differences between populations (Pletcher & Curtsinger, 2000a).

The nonlinear least squares method provides an alternative to maximum likelihood. This method is implemented in most statistical software packages and allows researchers to fit a large variety of nonlinear models. The limitation of this method is its theoretically less desirable optimality properties compared to the maximum likelihood, and less applicability to censored data. Both methods are sensitive to the choice of initial parameter estimates and outliers.

There is a tradeoff between flexibility and convenience of the nonlinear least squares method and the accuracy of the maximum likelihood approach. In practice, the theoretical considerations mentioned above are apparently not crucial, and the two approaches generate similar results. For example, Gehan and Siddiqui (1973) conducted a simulation study of fitting Gompertz and some other hazard models to survival data. The authors concluded that the least squares estimates are nearly as efficient as maximum likelihood when sample size is 50 or more. They also found that the weighted least squares approach, which accounts for systematic decrease of the sample size with age, generated more efficient but less accurate parameter estimates compared to the nonweighted method. Thus, maximum likelihood is a preferred method in those cases where the statistical software is readily available or the optimization procedure can be easily implemented. Otherwise, the nonlinear least squares may be a reasonable choice.

It is important to recognize the limitations and pitfalls of model fitting. The main problem is uneven statistical power. At young ages, there are relatively few deaths; at the oldest ages, death rates are high, but there are relatively few organisms. At middle ages, there are large numbers of both organisms at risk and deaths, and so statistical power for estimation of mortality rates is
concentrated in those middle age classes. Consequently, model fitting to the entire life history can give very accurate descriptions of the dynamics of middle age and can be systematically biased at early and late ages.

II. Experimental Evidence for Age-Specific Effects

If new mutations and genetic variants segregating in populations modify chances of survival by a constant factor at all ages [a situation known among demographers as “proportional hazards”], then there is no true age specificity; all is known from observations at a single age. However, if genes alter survival characteristics specifically at certain prescribed ages or stages of the life cycle, with no effect or very different effect at other ages, then the situation is more complex, and much more interesting. The evolutionary theory for the evolution of senescence requires age-specificity of genetic effects [Charlesworth, 1980; Curtsinger, 2001; Hamilton, 1966; Medawar, 1952; Williams, 1957]. As we discuss below, evolutionary models currently under investigation are sensitive to the precise degree of age specificity. Proving the existence of such age-specific genetic variation is difficult, especially at the older ages, but mounting evidence suggests that there may be a substantial degree of age specificity of genetic effects in Drosophila. In the following sections, we describe several different types of experimental evidence that address that issue.

A. P-Element Tagging

P-elements are naturally occurring transposable genetic elements [transposons] specific to Drosophila. Their ability to insert into random chromosomal locations throughout the genome makes them useful tools for genetic research, because they potentially disrupt gene expression or function at the insertion site. Screening of P-element inserts led to the discovery of life-extending “methuselah” (mth) and “I’m not dead yet” (Indy) single-gene mutations [Lin et al., 1998; Rogina et al., 2000]. Clark & Guadalupe (1995) used P-element insertion lines to investigate the genetic basis of senescence and found that otherwise genetically identical lines differed in survivorship and mean life span under the influence of P-induced insertions. The authors claimed that some of the P-element insertions led to reduced post-reproductive survival without affecting early life history, and that P-element insertions altered the ages at which mortality curves leveled off, though few demographic details were given.

B. Mutation Accumulation Experiments

The term mutation accumulation refers to both a theory of the evolution of senescence [Medawar, 1952] and an experimental design pioneered in Drosophila [Mukai, 1964]. It is the latter sense of the term that concerns us for the moment, although the former will be relevant later. The goal of a mutation accumulation experiment is to measure the rate at which new genetic variation spontaneously arises in a population, and to measure the phenotypic effects of those new mutations. General features of mutation accumulation experiments using Drosophila are as follows: starting with a single highly inbred line of flies, multiple sub-lines are established and maintained separately in small populations for dozens or even hundreds of generations. Spontaneous germline mutations occur independently in the various sub-lines, causing them to diverge both genetically and phenotypically. The sub-lines are kept at small census numbers so that new mutations have a reasonable chance to increase to fixation within
each particular line by random genetic drift. The rate at which sub-lines diverge phenotypically provides an estimate of the rate of input of new genetic variation affecting the particular trait assayed.

The first mutation accumulation study of age-specific mortality was executed by Pletcher and colleagues (1998), who established 29 sub-lines of *D. melanogaster* from a single highly inbred progenitor pair. Sub-lines were maintained for 19 generations, and then survival data were collected on approximately 100,000 flies. Mutational effects were detected by comparing age-specific mortality rates in each sub-line with that of the progenitor stock, which was maintained in nonmutating condition by cryopreservation. Significant mutational variance for age-specific mortality was detected, but only for flies aged less than 30 days post-emergence. Most of the new mutations were highly age-specific, each affecting survival rates over a well-defined age window of one or two weeks. Mutations that affected mortality at all ages were also detected, but their contribution to overall mutational variance was small. The conclusion from this study is that most new mutations have age-specific effects, but the failure to detect mutational variance at very old ages is difficult to interpret. It is unclear whether the failure to detect late-acting mutations is due to smaller sample sizes and loss of statistical power, to inherently lower mutation rates for alleles that specifically affect old age survival, or a combination of those factors.

Pletcher and colleagues (1999) continued the mutation accumulation experiment, assaying mortality rates at 47 generations of divergence, and also jointly analyzing data at three time points (10, 19, and 47 generations). These assays involved approximately a quarter of a million flies. Further evidence for highly age-specific mutation was found, and once again there was evidence for higher levels of mutation affecting early survival than late survival. Surprisingly, there appeared to be no upward or downward bias of mutational effects on mortality rates (mutations increasing mortality are as frequent as mutations decreasing mortality), contradicting the usual assumption that almost all mutations are deleterious to carriers. One possible explanation of this paradox may be related to elimination of many deleterious mutations through selective deaths at early larval stages of *Drosophila* development.

Mack and colleagues (2000) and Yampolsky and colleagues (2001) used a different experimental design, the “middle class neighborhood” method, to accumulate mutations affecting mortality, fecundity, and male mating ability on a genetically heterogeneous background of recently collected flies. They found clear evidence of age-specific effects of new mutations after 20 generations of mutation accumulation, including many effects limited to middle and advanced ages. This result contrasts with that of Pletcher and colleagues (1998, 1999), who found mostly early age effects. In both studies, the degree of age specificity declined in later generations of the experiment.

Martorell and colleagues (1998) executed a large mutation accumulation experiment to study life history in *D. melanogaster*, maintaining 94 sub-lines for 80 generations. They found evidence for small mutational effects on mean life span, but because mortality rates were not assayed, the experiment provides no information about age specificity of genetic effects. If Pletcher and colleagues (1999) are correct about mutations decreasing mortality as often as they increase it, then Martorell and colleagues (1998) might have underestimated the rate of mutations that modify mean life spans. Similar remarks apply to studies of life span and related characters in flies exposed to mutagenic chemicals (Keightley & Ohnishi, 1998).
accumulation experiments on life-history traits have also been executed using the nematode *C. elegans* (Keightley et al., 2000).

C. Neurogenetics and Gene Expression

Adult *Drosophila* are entirely post-mitotic organisms; that is, all cell division is completed when the animal metamorphoses from larval to adult stage. This contrasts sharply with other organisms, such as humans, in which cell division continues throughout the adult life span. It has been suggested that the lack of cell division in adult flies precludes late-onset genetic effects in *Drosophila*. However, recent evidence from several areas of biology that are not normally part of the discourse of demography suggests otherwise.

Neurodegenerative diseases in human, including Alzheimer’s, Huntington, and Parkinson’s disease, are characterized by late onset of pathology. Because *Drosophila* and humans share many functionally and structurally related genes, it is possible to model some of the human neurodegenerative pathologies by creating lines of flies that carry foreign or artificially modified genes (Driscoll & Gerstbrein, 2003; Fortini & Bonini, 2000; Mutsuddi & Nambu, 1998). Feany and Bender (2000) constructed transgenic flies carrying normal or mutant forms of the human gene for α-synuclein, a candidate cause of Parkinson’s disease. All transgenics exhibited normal neural morphology and geotactic behavior as young adult flies, but beginning at 25 days after eclosion, mutant transgenics developed Parkinson-like neural morphology and a dramatic loss of locomotor ability, whereas nonmutant transgenics escaped the morphological and behavioral manifestations of disease. Of course, the primary importance of such research is its potential application to treating human disease, but the α-synuclein case and others like it also demonstrate that genetic variation can produce specific late-onset phenotypes in adult *Drosophila*. Evidently, lack of cell division in adults does not preclude age-specific effects in older flies.

There is also evidence for age-specific genetic effects in modern studies of gene expression. It used to be widely assumed that the regulation of gene expression, which is capable of transforming single cells into highly differentiated and spatially structured mature organisms, becomes chaotic in old age. This view is now rejected, in part because of evidence from *Drosophila* (Helfand & Rogina, 2000, 2003; Rogina & Helfand, 1995; Rogina et al., 1998). Regulation of gene expression throughout the adult life span, including old age, sets the stage for age-specific genetic effects. DNA microarrays are powerful tools for the study of genome-wide patterns of gene expression in *Drosophila* and other organisms. Microarrays have been used to detect genes that vary in expression levels over the lifetimes of flies, and to detect genome-wide transcriptional responses to experimental treatments that modify life spans (McCarroll et al., 2004; Pletcher et al., 2002). Results from microarray studies bolster the view that gene expression is regulated throughout the adult life span and is therefore likely to be subject to genetic modification. Tahoe and colleagues (2005) demonstrated that age-specific patterns of gene expression differ between lines of *Drosophila* with very different mean life spans, and in some cases, including the genes encoding anti-microbial peptides, the line differences are manifest only in old age. Such observations do not prove that there are genetic differences between lines that alter survival specifically at advanced ages, but the observation of late-onset transcriptional differences does render the existence of such effects more likely. As more longitudinal studies of genome-wide transcription levels are
published in the next few years, we can expect a more complete picture of genome function and its variability throughout the adult life span.

D. Mortality QTLs

Quantitative trait locus (QTL) mapping is a set of procedures for identifying approximate chromosomal locations of segregating genes that influence polygenic traits (Mackay, 2001, 2002; see Chapter 8, this volume). QTLs affecting mean life span in Drosophila have been identified in a number of studies (Curtsinger et al., 1998; De Luca et al., 2003; Forbes et al., 2004; Leips & Mackay, 2000, 2002; Luckinbill & Golenberg, 2002; Khazaeli et al., 2005; Nuzhdin et al., 2005; Nuzhdin et al., 1997; Pasyukova et al., 2000; Resler et al., 1998; Valenzuela et al., 2004; Vieira et al., 2000).

In principle, it is possible to apply the methods of QTL mapping to localize genes that affect age-specific mortality rates rather than just mean life spans. However, the requirements are stringent: not only is there the prerequisite for large sample sizes, as in any estimation of mortality rates, but it is also necessary that the populations be genetically highly defined and contain a high density of genetic markers for QTL localization. To date this has been accomplished in only two cases. Curtsinger and Khazaeli (2002) identified QTLs that affect age-specific mortality rates in recombinant inbred populations of D. melanogaster, finding evidence for several genetically variable chromosomal regions that influence survival in age-specific fashion. The authors also developed a graphical method for presenting age-specific QTL results, as follows. QTL mapping results are typically presented in two-dimensional graphs: the abscissa represents chromosomal position, measured in units of recombination from the left telomere, while the ordinate represents a statistical measure, likelihood or LOD score indicating the probability that a QTL is present at a particular chromosomal position. A typical QTL map has peaks and valleys; genes affecting the quantitative trait are most likely to be located in chromosomal regions under the peaks, provided that the peaks exceed some likelihood threshold. Curtsinger and Khazaeli (2002) extended the usual analysis by mapping QTLs that affect mortality in each week of adult life and then adding a third dimension to the QTL map, indicating age. An example of a three-dimensional QTL map is shown in Figure 10.2. There is a QTL that affects age-specific mortality near the left end of chromosome III; the QTL has significant effects on
mortality in the first few weeks of adult life but has no effect on survival at later ages.

One other study of age-specific mortality rates using QTL mapping methods is that of Nuzhdin and colleagues (2005). QTLs affecting weekly mortality rates in both sexes were mapped in 144 recombinant inbred lines. Twenty-five statistically significant QTLs were found; most had positively correlated effects on mortality at several different ages, but in two cases the correlations were negative. Overall, the results suggest that the standing genetic variation in survival consists of a mixture of transient deleterious mutations that tend to increase mortality at younger ages, and a few mutations with opposing age-specific effects that are maintained by balancing selection. The latter are potentially examples of antagonistic pleiotropy, although finer genetic resolution will be required to rule out the competing linkage hypothesis.

III. Leveling-Off of Mortality Rates

In many biological species, including Drosophila and humans, death rates increase exponentially with age for much of the life span. However, at extreme old ages, a “mortality deceleration” occurs—the pace of mortality growth decelerates from an expected exponential curve. Sometimes this mortality deceleration progresses to the extent that mortality “leveling-off” is observed, leading to a “mortality plateau.” Thus, at extreme old ages, a paradoxical situation is observed when one of the major manifestations of aging—increasing death rate—apparently fades away or even disappears.

The phenomenon of mortality deceleration has been known for a long time, although its mechanisms were not intensively studied prior to the 1990s. The first person who noticed that the Gompertz curve is not applicable to extreme old ages was Benjamin Gompertz himself (Gompertz, 1825, 1872; see review by Olshansky, 1998). In 1867, William Makeham noted that for humans “the rapidity of the increase in the death rate decelerated beyond age 75” (p. 346). In 1919, Brownlee wondered whether it is “possible that a kind of Indian summer occurs after the age of 85 years is passed, and that conditions improve as regards length of life” (p. 385). Perks (1932) observed that “the graduated curve [of mortality] starts to decline in the neighborhood of age 84” (p. 15). Greenwood and Irwin (1939) confirmed that “the increase of mortality rate with age advances at a slackening rate, that nearly all, perhaps all, methods of graduation of the type of Gompertz’s formula overstate senile mortality” (p. 14). They also suggested “the possibility that with advancing age the rate of mortality asymptotes to a finite value” (p. 14), and made the first estimates for the asymptotic value of human mortality plateau [expressed in one-year probability of death, $q_x$]. According to their estimates of human mortality plateaus, “the limiting values of $q_x$ are 0.439 for women and 0.544 for men” (Greenwood & Irwin, 1939, p. 21). In 1960, Science published an article on a “General theory of mortality and aging” that listed some “essential observations which must be taken into account in any general theory of mortality.” (Strehler & Mildvan, 1960, p. 14). The first of these essential observations was the Gompertz law of mortality, while the second essential observation stated that “the Gompertzian period is followed by a gradual reduction in their rate of increase of the mortality” (Strehler & Mildvan, 1960, p.14). This observation of mortality deceleration was confirmed for several species, including Drosophila and C. elegans (Economos, 1979). The author
concluded “that after a certain species-characteristic age, force of mortality and probability of death cease to increase exponentially with age . . . and remain constant at a high level on the average for the remainder of the life span.” [p. 74]. The author called these findings “a non-Gompertzian paradigm for mortality kinetics” [Economos, 1979, p. 74]. A year later, the same author analyzed data for thoroughbred horses (mares), Dall mountain sheep, houseflies, and some other species and came to a conclusion that “Gompertz’s law is only an approximation, not valid over a certain terminal part of the lifespan, during which force of mortality levels off.” [Economos, 1980, p. 317]. These findings failed, however, to receive attention, and the topic stagnated.

A. Recent Studies of Mortality Plateaus

Prior to 1990, the most popular explanation of mortality plateaus was based on the idea of initial population heterogeneity, suggested by British actuary Robert Eric Beard (1911–1983). Beard developed a mathematical model in which individuals were assumed to have exponential increase in their risk of death as they age, but their initial risks differed from individual to individual and followed a gamma distribution [Beard, 1959, 1963, 1971]. This model produces a logistic function for mortality kinetics that is very close to the exponential function at younger ages, but then mortality rates decelerate and reach a plateau in old age. This compositional interpretation of mortality plateaus explained them as an artifact of mixture, perhaps reducing their intrinsic interest to biologists.

The situation changed in 1991, when it was found that the general theory of systems failure [known as reliability theory] predicts an inevitable mortality leveling-off as a result of redundancy exhaustion, even for initially identical individuals [Gavrilov & Gavrilova, 1991]. Thus, a testable prediction from this theory was that mortality deceleration should be observed even for genetically identical individuals kept in strictly controlled laboratory conditions. Shortly thereafter, Carey and colleagues [1992] and Curtsinger and colleagues [1992] published back-to-back papers in Science demonstrating mortality plateaus in laboratory populations of medflies and Drosophila, respectively. The medfly study employed genetically heterogeneous populations, whereas the companion study in Drosophila used highly inbred lines that were essentially devoid of within-line genetic heterogeneity.

The medfly and Drosophila experimental papers generated a flurry of criticisms and responses [Carey et al., 1993; Curtsinger et al., 1994; Gavrilov & Gavrilova, 1993; Graves & Mueller, 1993, 1994; Kowald & Kirkwood, 1993; Nusbaum et al., 1993; Robine & Ritchie, 1993; Vaupel & Carey, 1993]. Within a few years, even the most ardent critics were convinced that mortality plateaus were real phenomena and not merely artifacts of contamination or declining density in population cages [Khazaeli et al., 1995a, 1996]. Mortality plateaus were subsequently documented on very large scales in a variety of experimental species, including yeast, nematodes, Drosophila, medflies, parasitic wasps, and humans [see Vaupel et al., 1998, for a review].

Typical characteristics of a mortality plateau in Drosophila are shown in Figure 10.3 [from Pletcher & Curtsinger, 1998]. In this sample of 122,000 males, age-specific mortality increases in approximately exponential fashion from emergence until 60 days. After 60 days, when 5 percent of the original cohort remains alive, mortality decelerates and remains fairly constant until 80 days of age. Thus,
for a period of 20 days, or about 20 percent of the maximum life span in this particular experiment, there is no trend toward increasing mortality with increasing age. After 80 days the mortality curve shoots up, as the last few survivors die. The latter behavior is of no particular significance, and is best understood as an artifact of finite sample size, occurring when fewer than 10 flies remain alive.

The turnaround in views about applicability of the Gompertz model, which had been revered for well over a century, raises an obvious question: Why was Gompertz widely accepted until recently, and even raised to the stature of “Gompertz’ law” despite various exceptions being pointed out? In addition to science’s predilection for simple laws of nature, the likely explanation is that most survival experiments prior to the 1990s had been too small to detect plateaus. Mortality plateaus are late-life phenomena. Small experiments fail to detect them because there are few survivors to the age at which mortality rates begin to level off. It is also possible that biologists’ habit of examining survivorship curves rather than mortality rates contributed to ignorance about plateaus; it is difficult to see a plateau in the tail of a survivorship curve, even if sample sizes are relatively large.

### B. Explaining Mortality Plateaus

Although the existence of mortality plateaus is now universally accepted, explaining why plateaus exist is controversial. It is convenient to define two general, non-exclusive classes of explanations: population heterogeneity and individual aging. Heterogeneity refers to the idea that individuals in a cohort differ in frailty, which is most conveniently parameterized as a multiplicative factor of the Gompertz hazard model. The hazard rate of an individual of age \( x \) and frailty \( Z \) is

\[
\mu_x Z = Z \alpha e^{\beta x},
\]

where \( Z \) is a gamma-distributed random variable with mean 1 and variance \( \sigma^2 \). Under those circumstances, the mean age-specific mortality in the population is given by the logistic equation. Individual differences in frailty can be genetic or environmental in origin and tend to produce mortality deceleration. This occurs because weaker organisms die first, leaving preferentially more robust members of the population alive for later survival measurements. The process of sorting weaker and stronger individuals by death within a generation is often referred to as “demographic selection,” the first part of the term being necessary to distinguish it from selection of the Darwinian sort.

Frailty may be fixed at birth, or acquired and modified through life experience, as mentioned above. For instance, for the fixed frailty situation, we might imagine that a population of flies contains different genotypes, each with its characteristic hazard rate. Or, in a genetically homogeneous population such as an inbred line or F1 cross between two inbred lines, differences...
in frailty between organisms could arise from micro-environmental effects, such as slight uncontrolled spatial variation in temperature or quality of food experienced at pupation sites. In either case, the essential feature of the fixed frailty models is that the organisms carry a certain frailty factor $Z$ with them throughout their lives. In contrast, flies could acquire different frailty factors during their adult lifetimes as a result of exposure to infectious organisms, or differential rates of reproduction. In either case, the logistic model predicts the expected population mortality dynamics (Yashin et al., 1994), and the magnitude of population variance for frailty has a strong influence on mortality dynamics.

Gavrilov and Gavrilova (1991, 2001; see Chapter 1, this volume) developed several classes of aging models based on reliability theory. Interestingly, all these models predict a mortality deceleration, no matter what assumptions are made regarding initial population heterogeneity or its complete initial homogeneity. Moreover, these reliability models of aging produce mortality plateaus as inevitable outcome for any values of considered parameters. The only constraint is that the elementary steps of the multi-stage destruction process of a system should occur by chance only, independent of age. The models also predict that an initially homogeneous population will become highly heterogeneous for risk of death over time [acquired heterogeneity].

Another class of explanations for mortality plateaus depends not on differences between individuals, but on changes within individuals as they age. If the hazard rates for individual organisms decelerate at older ages, then so, too, will the observed population mortality. One can imagine various biological reasons that individual hazard rates might decelerate. Older flies might incur less physiological and metabolic cost from mating behavior and reproduction, or lower activity levels in old age might entail less exposure to infectious agents and less generation of harmful oxygen radicals. For humans, a similar hypothesis was proposed by Greenwood and Irwin (1939), who suggested that lower-than-expected mortality of centenarians could be explained by their less risky behavior.

There is a growing body of evolutionary theory that addresses ultimate causes of mortality plateaus. The basic problem to be solved by theoreticians is that evolutionary models of age-specific mortality tend to generate very high mortality rates (“walls of death”) at post-reproductive ages (Charlesworth, 1980; Curtsinger, 1995a,b; Partridge & Barton, 1993; Pletcher & Curtsinger, 1998). Imagine a population in which there is initially no senescence—that is, the hazard rate is the same for all age classes. Over time, new mutations occur, some of which have age-specific effects on survival. Many of the new mutations are deleterious at all ages and are quickly eliminated from the population by natural selection. Some mutations, presumably very few, improve survival of carriers at early ages, are positively selected, and increase in frequency in the population; this causes an evolutionary lowering of the population mortality curve at juvenile and reproductive ages. Some mutations increase or decrease mortality specifically at post-reproductive ages, but because post-reproductive survival is irrelevant to Darwinian fitness, natural selection does not discriminate. The net result is that there is no evolutionary force “pushing down” on the late-life part of the mortality curve. If the majority of mutations that affect old-age survival cause a deterioration of vitality, then post-reproductive survival will erode under mutation pressure, with nothing to stop it from eventually producing a wall of death. This scenario presumes the existence of exclusively late-acting mutations, as originally postulated by Medawar (1952), and is known as the mutation accumulation model of the
evolution of senescence. The central problem for evolutionists trying to understand mortality trajectories is to discover some means of counteracting the tendency of recurrent mutation to drive post-reproductive hazard rate to infinity.

One possibility, not widely considered, is that mutations that affect only the old might improve survival as often as they erode it. This might seem at first glance to be nonbiological, violating the widely held view that the vast majority of mutations are deleterious to their carriers. However, reasonable scenarios can be imagined; for instance, a mutation that reduces mobility in old age might increase survival by causing carriers to generate fewer damaging oxygen radicals. There is some suggestion in the results of mutation accumulation experiments described above that mutations increase survival as often as they decrease it, but it must be admitted that the distribution of mutational effects for old-age-specific mutations is not known in detail.

Abrams and Ludwig (1995) addressed the mortality plateau problem in an evolutionary context by analyzing an optimality model in which organisms are presumed to allocate resources to either somatic repair or reproduction. The optimal allocation was presumed to be that which maximizes lifetime reproductive output. Abrams and Ludwig (1995) found that an optimal allocation involves declining investment in repair with increasing age, which, the authors suggest, could lead to late-life mortality plateaus. However, Charlesworth and Partridge (1997) re-examined the optimality model and found that the death rate tends to infinity with increasing age. We also note that the optimality approach does not specifically incorporate deleterious mutations with age-specific effects, an important omission.

Mueller and Rose (1996) used numerical simulations to study the evolution of mortality under antagonistic pleiotropy—that is, the assumption that mutations have negatively correlated effects on survival at young and old ages. They argued that such models easily explain mortality plateaus, but their results have been widely criticized. Mueller and Rose (1996) assumed that every mutation increases survival in one randomly chosen age class, and reduces it in another; there are no unconditionally deleterious mutations in the model. Charlesworth and Partridge (1997) noted that the Mueller-Rose model was not iterated to equilibrium, and suggested that late-life survival rates would approach zero in this model as more evolutionary time elapsed. In general, the evolutionary equilibrium state is difficult to define in numerical simulations of finite populations. Pletcher and Curtsinger (1998) argued that the Mueller-Rose model includes a strange feature that biases the results: there is an assumption that when the population mortality rate is low, new mutations tend to increase mortality, but when the mortality rate is high, new mutations tend to make it decrease. The net effect is that mortality rates are forced toward an intermediate value. Pletcher and Curtsinger (1998) showed that removing that assumption leads to a late-life wall of mortality. The most telling critique is by Wachter (1999), who obtained analytical results for a generalized class of Mueller-Rose–type models and concluded that mortality plateaus cannot be accounted for by their equilibrium behavior. Wachter (1999) states unequivocally that the Mueller-Rose model fails in this respect. Thus, it seems likely that the simulation of Mueller and Rose (1996) produced transient mortality plateaus that were erroneously interpreted as equilibrium evolutionary states.

Given strong criticisms of the Mueller-Rose simulation model and analytical invalidation of its results, it is surprising that Drapeau and colleagues (2000), Rose and Mueller (2000), and Rose and colleagues (2002) have continued to promote it. All three of those papers failed to
cite the analytical results of Wachter (1999). Mueller and colleagues (2003) address the various criticisms, including Wachter’s (1999) analytical results, but the responses are unconvincing (de Grey, 2003a, 2004; Service, 2004). Technical details aside, the broader point is that Rose, Mueller, and their associates endorse individual aging over population heterogeneity as a general explanation for mortality plateaus, a position that could ultimately prove to be correct. They refer to their argument as “the evolutionary theory” (Rose & Mueller, 2000, p. 1,660), implying that heterogeneity explanations are “un-evolutionary” or “anti-evolutionary.” The nomenclature is unfortunate. Phenotypic variability between organisms, including genetically identical ones, is an essential feature of quantitative genetic variability and micro-evolutionary change (Falconer & Mackay, 1996). Labeling the argument “evolutionary” is just a rhetorical device, with few constraints on its use: Graves and Mueller (1993, 1994; see also Curtsinger, 1995a,b) raised the “evolutionary” flag when they argued against the existence of mortality plateaus in Drosophila, a stance that was eventually abandoned.

Pletcher and Curtsinger (1998) presented simulation results for the evolution of mortality plateaus, focusing on positive pleiotropy, in which mutations exert positively correlated effects on mortality rates at different ages. In these simulations, positive pleiotropy seemed to produce mortality plateaus, but, as in any simulation of finite populations, the definition of stable evolutionary state is difficult, and the outcomes were probably transient. Charlesworth (2001) used analytical techniques to study a similar situation by assuming that all deleterious mutations have deleterious effects at reproductive ages. This assumption prevents mutation frequencies from exploding at older ages and, thus, preserves mortality plateaus. Service (2000a) simulated mortality dynamics under the assumption of population heterogeneity in individual age-specific risk of death. Heterogeneity was modeled by assigning each individual a unique Gompertz mortality function, with means and variances of Gompertz parameters based on the published literature for Drosophila. He found that the heterogeneity generated by variation in Gompertz parameters was sufficient to explain late-life mortality plateaus and could also account for late-life declines in genetic variance of mortality rates. Similar conclusions were reported by Pletcher and Curtsinger (2000b).

The reliability models of multistage destruction (Gavrilo & Gavrilo, 1991, 2001) were recently reformulated in mathematical terms of a stochastic Markov process (Steinsaltz & Evans, 2004). The authors define a Markov mortality model as a stochastic process, which is “killed” at random stopping times according to the behavior of a Markov process. A general feature of such multistage models is that they usually produce mortality plateaus, as it was demonstrated earlier with a more simple approach (Gavrilo & Gavrilo, 1991, 2001). As Steinsaltz and Evans (2004) put it, “the mortality rate stops increasing [with increasing age], not because we have selected out an exceptional subset of the population, but because the condition of the survivors is reflective of their being survivors, even though they started out the same as everyone else.” Thus, the Markov mortality models explain mortality plateaus by a type of heterogeneity in acquired frailty because the underlying assumptions are similar to the earlier reliability models.

In evaluating the various theories, it is important to remember that the fact that a particular mathematical model or simulation can fit or “predict” an experimental outcome is not proof that the assumptions of the model are correct. For
example, the venerable Hardy-Weinberg model of population genetics predicts certain genotypic frequencies, but observation of those frequencies in real populations does not validate the underlying assumptions of the model (random mating, absence of natural selection, etc.). Theory guides our thinking, but critical tests must come from well-designed experiments, efforts at which are described in the next section.

C. Testing the Theories

Designing critical experiments to address the causes of mortality plateaus has proven to be exceptionally difficult; in fact, all experimental tests in this area are flawed in one way or another. Thus, no final answers can be given at present, but it is instructive to review the relevant experiments and consider the pitfalls.

The first experiment specifically designed to test heterogeneity theory used lethal stress to manipulate the magnitude of population heterogeneity (Khazaeli et al., 1995b) and was inspired by demographic studies of human populations after a catastrophe (Vaupel et al., 1987). Using a single highly inbred line of flies, multiple age-synchronized cohorts were established. In control populations, flies were maintained under the usual conditions, whereas in experimental populations, flies were subjected to 24 hours of desiccation at a young age. About 20 percent of the flies died during and immediately after the desiccation stress. Post-stress mortality rates are informative about population heterogeneity; in particular, in the absence of population heterogeneity, post-stress mortality in experimental and control populations is expected to be identical. However, if there is significant population heterogeneity at the time of the stress, then post-stress mortality in the experimental populations is expected to drop below that of the control populations, at least temporarily, because the more frail individuals will have been eliminated. The latter pattern was observed, and was interpreted by Khazacli and colleagues (1995b) as evidence for significant levels of heterogeneity. However, the authors retracted that result when it was realized that there was a flaw in the interpretation (Curtsinger & Khazacli, 1997). The problem is that exposure to an external stress does more than kill the more frail flies; it also induces a stress response in the survivors. This phenomenon, known as hormesis, is well documented in a variety of species and involves a rapid genomic response to severe stress. The stress response is an interesting phenomenon, but it creates difficult problems in the interpretation of the stress experiment. In particular, the post-stress decline in mortality among experimentalists compared to controls could be due to reduced heterogeneity through elimination of weaker flies, hormesis induced among survivors, or both factors. The experimental design of Khazacli and colleagues (1995b) does not permit separation of the heterogeneity and hormetic effects, and so the result is inconclusive regarding heterogeneity. Recently, the stress experiment was redesigned to correct the confounding flaw, and data have been collected in the Curtsinger lab on 100,000 male flies of one inbred genotype. Five intensities of stress were applied, including one sufficient to induce a stress response but not severe enough to cause immediate deaths. The mild stress will allow estimation of the hormesis effect independent of the heterogeneity effect, unconfounding the variables. Data analysis by the authors of this chapter and Dr. A. Khazacli is underway.

A different and more benign experimental design was used by Khazacli and colleagues (1998), who attempted to manipulate population heterogeneity by fractionating genetically homogeneous populations. Working with two
highly inbred lines, experimental populations were subjected to the most stringent environmental controls possible, far beyond what is normally employed in fly husbandry. Eggs were collected over a seven-hour period, instead of the usual 24 hours. First instar larvae were collected from that sample for only three hours, and emerging adults were collected in three-hour windows. The result of all this careful timing of development is that within a cohort, adult flies experienced larval and pupal environmental conditions that are as similar as possible. The question then is whether the environmentally “homogeneous” populations exhibit mortality plateaus to a lesser extent than normal environmentally “heterogeneous” control populations. Khazaeli and colleagues (1998) found that 93 percent of experimental populations and 100 percent of control populations exhibited statistically significant mortality deceleration late in life. The authors concluded that reducing environmental heterogeneity during larval and pupal stages has negligible effect on adult mortality trajectories. Drapeau and colleagues (2000, p. 72) overstated this experimental result when they wrote that “Khazaeli et al. (1998) found no evidence to support the hypothesis that environmental heterogeneity among individual flies is a primary factor in determining late-life mortality rates.” The experiment actually gives information only about larval and pupal stages, and is in the strictest sense relevant only to the “fixed-heterogeneity” model. The results are not informative about heterogeneity acquired in adulthood, which may be substantial. Perhaps a broader lesson from this study is that there is a substantial and intrinsic environmental heterogeneity in experimental populations that cannot be removed experimentally, even by Herculean efforts.

The most widely discussed experimental test of heterogeneity theory is that of Drapeau and colleagues (2000), who argued that there is a close connection between frailty and sensitivity to environmental stresses in experimental populations of Drosophila. They further suggested that, according to heterogeneity theory, populations differing in tolerance to stress should have different late-life mortality characteristics, though the nature of the expected differences was not spelled out. They compared mortality trajectories in fly populations that had been selected for resistance to starvation with those of unselected controls. No statistically significant differences were found, which the authors interpreted as evidence against the heterogeneity theory. Service (2000b) questioned the assertion that the populations are expected to differ in late-life mortality, noting that for the logistic model the plateau occurs at $B/\sigma^2$. Consequently, populations could differ in the intercept parameter $A$ and have the same levels of late-life mortality. As noted by Mueller and colleagues (2000) in their response to Service (2000b), the force of this criticism is blunted by the generally accepted theoretical observation that large and biologically unrealistic amounts of variation in the intercept parameter would be required to produce mortality plateaus, if that were all that varied between individuals. Service (2000b) also noted that if $\sigma^2$ is lower in the selected population, then it is expected to have higher mortality rate than controls (when all other parameters are fixed), especially at early ages, as observed. Service concludes that the results of Drapeau and colleagues (2000) are entirely consistent with the predictions of the heterogeneity model. de Grey (2003b) criticized the use of maximum likelihood methods by Drapeau and colleagues (2000) and argued that heterogeneous Gompertz parameters could explain the experimental results. Steinsaltz (2005) reanalyzed the experimental results of Drapeau and colleagues (2000) and questioned the claim that
there is no difference in late-life mortality schedules between populations. The original claim was based on comparisons of means averaged over populations. Steinsaltz (2005) noted that the data are bimodal, and means are therefore misleading. He reanalyzed the data and found that populations were actually quite different, the mortality plateau being lower in the selected populations. He concluded that the experimental results lend mild support to the heterogeneity theory, although the expected differences in timing of the plateau were not observed. In sum, the critiques of Drapeau and colleagues (2000) are varied and instructive, and illustrate some of the difficulties of the experimental task and complexities of the analysis.

Rose and colleagues (2002) studied mortality trajectories in populations of Drosophila that had been artificially selected for long life and compared them to unselected control populations. Mortality trajectories had previously been studied in the same populations by Service and colleagues (1998), who invoked a heterogeneity explanation. Rose and colleagues (2002) showed that control populations consistently exhibited earlier onset of mortality plateaus than selected populations. This result was interpreted as being consistent with an “evolutionary” (i.e., individual aging) model. The result is suggestive, but not critical; it is not clear that the observations are inconsistent with predictions of any particular heterogeneity model. In general, we consider it very unlikely that critical tests of heterogeneity and individual aging models can be executed with outbred experimental populations. The problem is that the variance parameter plays a central role in the predictions of heterogeneity models but is generally unknown in either relative or absolute terms for outbred, genetically uncharacterized populations. It is widely assumed that selected populations are less heterogeneous than unselected controls because some genotypes have been eliminated by selection. However, several factors could cause selected populations to be more heterogeneous, both in genetic and environmental variance. If the selection response entails an increase in frequencies of initially rare alleles, genetic variance is expected to increase under selection, a prediction that has been verified experimentally (Curtsinger & Ming, 1997). This counterintuitive result occurs because the contribution to total genetic variance by any particular locus depends on $2pq$, where $p$ and $q$ are allelic frequencies (Falconer & Mackay, 1996); rare and common alleles contribute little to population genetic variance, but alleles at intermediate frequencies potentially contribute much. The same effect occurs if new mutations increase to appreciable frequencies during the selection process. Another factor that complicates matters is genetic homeostasis. It is well known that homozygous genotypes generally exhibit greater environmental variance than heterozygotes (see review by Phelan & Austad, 1994). If selection and/or inbreeding increase homozygosity in selected populations, then the environmental component of variance is expected to increase. In short, there are too many unknown variables in genetically uncharacterized outbred populations to allow critical tests of predictions of heterogeneity models. A better experimental design is that of Miyo and Charlesworth (2004), who studied mortality rates in hybrid progeny of crosses between inbred lines of Drosophila. In such populations, all individuals are genetically alike, except for recent mutations, and heterozygous at loci that differ between parental lines. Miyo and Charlesworth (2004) found that populations of both mated and unmated hybrid males exhibited mortality plateaus, and suggested that their results were consistent with underlying heterogeneity of mortality rates.
In the final analysis, evaluating the various heterogeneity models is a purely quantitative question. No reasonable person would deny that there is some heterogeneity for frailty within populations, even genetically homogeneous ones; the question is whether there is sufficient heterogeneity to produce late-life mortality plateaus. We are optimistic that large-scale, multilevel stress experiments and other designs using genetically defined populations will provide the relevant estimates. On the other hand, if the individual aging theory is correct, then there must be some important biological processes that differ between organisms at pre- and post-plateau ages and account for the change in mortality trajectory.

IV. Conclusions

The integration of biology and demography proceeded sporadically for most of the 20th century. Pearl, Sacher, Strehler, and others showed the way toward integration of the fields, but their efforts were not always widely appreciated. Now we are in a period of widespread dissemination of demographic techniques among experimental biologists. The new field of biodemography is flourishing and has rich conceptual bases to draw on in demography, evolutionary biology, reliability theory, and even theoretical physics (Pletcher & Neuhauser, 2000). Its first major conceptual challenge is to explain mortality plateaus. We are optimistic that consensus will emerge in this area as experimental designs and methods of data analysis become more sophisticated. Other important challenges include defining the nature of age-specific genetic variation and explaining the high degree of environmental variation in demographic parameters.

Acknowledgements

Research is supported by grants from the National Institute of Aging at the National Institutes of Health. We thank Dr. A. Khazaeli for comments.

References


Leips, J., & Mackay, T. F. C. [2000]. Quantitative trait loci for lifespan in
Genes affecting aging: mapping quantitative trait loci affecting longevity in Drosophila melanogaster using amplified fragment length polymorphisms (AFLPs). Genetica, 114, 147–156.


Pearl, R., & Parker, S. (1921). Experimental studies on the duration of life. I.


I. Introduction

The promise of genome-wide platforms for biological discovery has been received by biological scientists with great enthusiasm. Of the global discovery technologies, the increasing accessibility of microarrays for analysis of gene expression has perhaps stirred the greatest interest, certainly within the field of gerontology. As this chapter will discuss, recent literature includes applications of these arrays to studies of aging in yeast, invertebrates, rodents, and humans. However, the very nature of the technology—the measurement of thousands or tens of thousands of variables at once—presents new challenges for the analysis and interpretation of the data, requiring the development and application of new statistical and informatics tools. We have begun to see initial fruition of this work, especially in yeast and invertebrate models. Studies in mice and humans are well underway; however, the greater inter-individual heterogeneity and tissue and genomic complexity of these organisms remain as appreciable challenges. Because global analysis of gene expression offers great potential for the elucidation of common mechanisms, pathways, and biomarkers of aging, it is safe to predict continued enthusiasm for application of this technology toward these goals.

II. Technical Issues

A. Design of Aging Studies

1. Biological Aspects of Design

One of the most important (and often neglected) considerations in the successful design of microarray experiments is the unique properties of the biological question being explored. In the case of the biology of aging, with pleiotropic and often subtle phenotypes, careful examination of the precise biological features under scrutiny is paramount. Unfortunately, important considerations, such as how to define terms like “old” and “premature aging,” are often
neglected. This section attempts to deal with these issues in relationship to the experimental design of microarray studies. In particular, we will address how experimental design determines the types of gene expression biomarkers obtained. Several classes of aging-related microarray experiments are considered, including analysis of gene expression at different ages, analysis of gene expression in short- and long-lived models, and analysis of individual longevity.

a. Gene Expression Studies at Different Age. Perhaps the most common application of microarray technology to the study of aging is the search for gene expression changes that correlate with organismal age. Studies of this type typically employ a design in which RNA is obtained from “young” individuals and compared against RNA obtained from “old” individuals. Microarray analysis is then carried out and a comparison is made between “young” and “old,” with lists of genes presented that either increase or decrease in expression as a function of age. These types of studies have been carried out in all of the model systems commonly used to study aging, including mammals, flies, worms, and yeast, as described in subsequent sections.

In the vast majority of “young” versus “old” studies, only two timepoints have been used. This two-timepoint design is fraught with danger because no information is gained regarding the kinetics of gene expression change. In some cases, intermediate age timepoints have been collected in addition to “young” and “old.” This type of design has the advantage that it may be possible to identify genes that show trends in expression correlated across multiple age groups. In addition, it may be possible to identify and classify genes based on the kinetics with which expression changes occur. This is particularly important if any inferences regarding the causality of observed changes with respect to aging or age-associated phenotypes are to be made. Whenever possible, it is strongly recommended that multiple timepoints be used rather than only “young” and “old.”

One important consideration in the design of a study comparing young and old organisms is how to define the “young” and “old” populations. For young populations, the primary criteria should be organisms that are reproductively and developmentally mature. In order to define “old” populations, the most straightforward definition is derived from statistical parameters of the life-span distribution, such as population median and maximum life span. For example, one definition of “old” could be individuals that have achieved at least 75 percent of the population maximum life span. The potential for degenerative changes present in very old animals must be considered, however, as these secondary gene expression changes may complicate the observation of those more intimately tied to the biology of aging. An alternative definition of what age constitutes “old” might be based on the appearance of one or more phenotypes associated with old age. Such a definition, however, is complicated by the fact that a majority of aging phenotypes show incomplete penetrance, with large individual variation in age of onset and severity. On the other hand, an appropriate set of phenotypic markers might reflect biological age more accurately than does chronological age (see section II.A.1.e). In some cases, the age of the “old” population is determined by the availability of donor samples. Human studies in particular are often constrained by sample availability and must make use of tissues or cells obtained from donors of a variety of different ages. For microarray studies, as with other types of analyses, there is no clear right answer as to what the definition of “old” should be. What is clear, however, is that the parameters used to define the old population
should be carefully considered and explicitly presented during both the experimental design and data interpretation. In addition, the statistical analyses used should reflect the experimental design in this regard. As stated above, a multiple timepoint design in which samples are obtained at several different ages is preferred, as this provides additional information about the kinetics of gene expression change with age and a measure of flexibility in choosing appropriate donor ages.

b. Gene Expression Analysis of Long-Lived Models. In addition to studies comparing young organisms with aged organisms, microarrays can be used to compare individuals or populations of similar age but with different aging potentials. For example, many studies have examined the gene expression profile of young mice fed a control diet relative to young mice fed a calorie restricted (CR) diet (see section III.C.2). When age-matched animals are compared across the two dietary regimens, differences reflect gene expression changes associated with caloric intake rather than age. Because CR animals have a longer life span than control-fed animals, observed gene expression changes also have been correlated with increased longevity, at least within the experiment. This distinction is an important one. In “young” versus “old” experiments, the potential exists to discover gene expression changes that correlate with chronological age [biomarkers of age or aging]. In experiments carried out on young control versus long-lived individuals, it is possible to identify gene expression changes that correlated with longevity [biomarkers of longevity] and/or the rate of aging. Such studies may permit the discovery of genes with the potential to extend life span that may not be altered in normal aging.

c. Studies of Both Longevity and Age. In several studies, these two types of experimental design have been combined such that microarray analysis is carried out on control and long-lived animals at two or more ages. The primary advantage associated with this type of design is that a single experiment can be used to identify gene expression changes correlated with aging, the model of longevity and the interaction of aging and the model of longevity. Figure 11.1 shows the possible gene expression changes in a “four-way design” experiment where young and old animals are used with and without a modification that affects longevity [Mod]. Biomarkers of aging [light bars in the Old WT column of Figure 11.1, rows B, D, G, and H], which are unaffected in old members of the group modified for longevity (Figure 11.1, row B), can be distinguished from biomarkers of aging, which are attenuated in old members of the group modified for longevity (Figure 11.1, row D). The latter may include age-associated gene expression changes that are functionally important for longevity; these are predicted to be attenuated in long-lived animals relative

![Figure 11.1](image-url)
to controls. The classic example of this phenomenon is the observation that many gene expression changes associated with age in control mice are reduced or absent in old mice subjected to CR (Cao et al., 2001; Dhahbi et al., 2004). Of note, many of the studies of CR are missing the young treated (short-term CR) control group, and this may complicate the interpretation of results. In addition, biomarkers of longevity in the modified group can be examined in the young animals (Figure 11.1, row F), old animals (Figure 11.1, row E), or both (Figure 11.1, row C), and effects of the longevity modification occurring early and late in the aging process can be distinguished. Furthermore, complex changes associated with aging and altered by the longevity modification can be identified (Figure 11.1, rows G and H). In these cases, biomarkers of aging may also be affected in the young modified group. This might represent a more complex phenomenon, such as a stress induced by the longevity modification that leads to a stress response similar to that seen in aging. A related design, typically used for dietary modification studies such as caloric restriction, is a “three-way design” in which young animals from an age prior to the start of caloric restriction are compared to older animals that have been subjected either to caloric restriction or normal feeding [Lee et al., 1999]. This design is effective for the study of modifications that are started after development.

d. Microarray Studies on Short-Lived Mutants. Several mutations that result in shortened life span have been suggested as models of premature aging in mammals [Warner and Sierra, 2003]. Microarray analysis of tissues or cells derived from short-lived mutants offers the opportunity to identify gene expression changes correlated with short life span and, potentially, accelerated aging. In principle, such studies are identical in design to those examining gene expression changes in models of increased longevity. It should be noted, however, that none of the “premature aging” models proposed to date recapitulate all of the phenotypic changes associated with aging in normal animals, and most have additional pathologies that do not occur during the normal aging process. Differentiating gene expression changes associated with accelerated aging (if present at all) from those associated with non-aging related pathologies is a daunting task. The most effective approach would most likely be a comparative analysis of gene expression profiles across multiple long-and short-lived mutants. In this way, gene expression biomarkers that reflect the rate of aging could potentially be identified with higher confidence.

e. Biomarkers that Predict Individual Longevity. Another potential use of microarrays applied to aging research is the large-scale identification of gene expression biomarkers that predict individual longevity. Among humans, it is clear that different individuals age at different rates, due to both genetic and environmental factors. For a particular person, chronological age may not be an accurate predictor of remaining life expectancy. In fact, several phenotypes have already been suggested as potential biomarkers of biological age in humans, including body temperature [Roth et al., 2002], serum insulin levels, age-related rate of decline in serum dehydroepiandrosterone sulfate (DHEAS) [Roth et al., 2002], and telomere length in blood cells [Cawthon et al., 2003]. To date, however, it has not been demonstrated that these biomarkers can be accurately used to predict survival.

Microarrays offer the opportunity to detect a group of gene expression biomarkers that more accurately reflect biological age (and hence life expectancy) than currently possible. Hundreds of potential biomarkers can be assayed simultaneously in a single array experiment. Appropriate experimental design for the identification of individual biomarkers that are associated with aging in humans may be necessary.
biomarkers of longevity using microarrays, however, is nontrivial.

One approach might be to compare gene expression profiles from a specific tissue type using samples from centenarians versus samples from the general population. A problem with studies of this type, however, is the confounding effect of gene expression changes due to age-associated disease and degenerative changes that are likely to be present in the centenarian population simply due to the extreme chronological age of these individuals. One way to get around this complication would be to compare samples from age-matched siblings or offspring of centenarians versus age-matched individuals from the general population. Because the probability of achieving extreme longevity is quantifiably higher in close relatives of centenarians [Perls et al., 2000; Perls et al., 2002], it might be possible to extract the genetic component of gene expression changes associated with a predisposition for longevity in this manner.

Alternatively, a mammalian model system, such as mice, could be used to identify gene expression biomarkers predictive of individual longevity. Similar to the human study mentioned above [Roth et al., 2002], reduced body temperature and serum insulin are associated with longevity in mice [Weindruch and Walford, 1988]. Recent work has also suggested that body weight and levels of T-cell subsets and thyroxin can be used to predict individual longevity in animals as young as 8 months old [Harper et al., 2004]. One approach for using microarrays to identify these types of biomarkers would involve analysis of global gene expression in particular tissue, such as blood, from individual animals at multiple age points (e.g., every 6 months). Following the death of all animals in the cohort, it should be possible to identify gene expression patterns that correlate with individual longevity relative to the population. A computational algorithm could then be generated to predict life expectancy based on these gene expression markers.

From a clinical perspective, the use of microarrays has even greater promise than just identifying candidate genes for potential therapeutic intervention. Microarray analysis to identify biomarkers of aging and longevity could strongly influence preventative care and risk management–based decisions about screening for diseases of the elderly. It is possible that a specialized version of a microarray analysis for potential biomarkers of aging could be used in routine assessment of aging adults.

2. Sources of Variability in Microarray Experiments

When considering the design of a microarray experiment, the sources of biological and technical variability must be considered, as they affect the ability to detect true gene expression changes. Microarray experiments can most clearly detect gene expression changes with low biological variability. However, changes in genes with high biological variability may be of interest to the researcher for several reasons. Some of the sources of biological and technical variability as well as methods for addressing them are discussed below.

a. Cellular Heterogeneity. One important factor to consider in the analysis of variability in microarray experiments is the cellular heterogeneity of the samples being analyzed. In most experiments conducted on multicellular species, a single sample is derived from whole animals or organs. Therefore, changes in gene expression due to changes in underlying cell type distribution are indistinguishable from gene expression changes due to transcriptional events within a cell. This is a particularly important cautionary note when considering aging experiments where the time-dependent structural changes associated with atrophy or hypertrophy are well known and when tissues...
are rarely cleared of blood before RNA preparation. Gene expression differences arising from a population of cells may still provide insight into changes in the function of an organ or organism with age and are certainly interesting. However, when attempting to infer the cause of gene expression changes, it is important to consider that cell type differences may be at the root of the observed differences and that assumptions about alterations in intracellular signaling pathways may be unfounded when whole tissues are analyzed.

Techniques for measuring gene expression from single cells in an organism are being actively developed in order to address this concern. Amplification procedures allow for the labeling of RNA from samples as small as 50 nanograms (ng). Most of these amplification procedures are based on the methods developed by Van Gelder and colleagues (1990). Concerns about bias introduced by amplification appear to differ depending on the microarray platform. For Affymetrix arrays, a systematic bias introduced by amplification has been noted (Wilson et al., 2004). However, this bias does not affect the genes identified as differentially expressed when data is only compared directly between samples that have undergone the same amplification procedures. For cDNA arrays, Feldman and colleagues (2002) reported that the bias introduced by amplification is negligible. Additionally, there are several reports suggesting that amplification may reduce the noise inherent in rare transcripts for cDNA arrays and produce data that is more likely to be verified by reverse transcriptase polymerase chain reaction (RT-PCR) (Feldman et al., 2002; Gomes et al., 2003). In addition to potential bias due to amplification, analysis of RNA from single cells may introduce stochastic and micro-environmental variability that would normally be averaged out in a sample composed of thousands of cells. The magnitude of gene expression differences between neighboring cells in a tissue due to these effects has not been sufficiently characterized to date.

b. Temporal Heterogeneity. In addition to cellular heterogeneity, effects of temporal differences may also introduce variability into a data set. Gene expression differences can be associated with time of day or level of activity, and because a typical gene expression study is a temporal cross-section, it is difficult to determine whether the differences seen between individuals are due to a true stable heterogeneity or are the result of transient gene expression in certain animals. Care should be taken when preparing for a microarray experiment to minimize temporal factors that may affect apparent gene expression.

c. Age-Associated Biological Variability. There are several potential sources of biological variability inherent in the aging process that can affect the outcome of a microarray experiment. For example, age-associated dysregulation at the level of the genes (e.g., alterations in gene silencing), at the level of cellular signaling (e.g., alterations in functionality of signaling pathways), at the level of tissue (e.g., alterations in cell type composition), or at the level of the organism (e.g., alterations in circulating hormone levels) may cause an increase in the variability of gene expression with age. These changes are clearly associated with the aging process but may not affect the same genes in each individual to the same extent, thus appearing as an increase in the variance of samples from older individuals. Similarly, discarding as uninteresting genes that show significant expression changes in only a subset of individuals may overlook biologically important candidates. One positive aspect to the increase in gene expression variance with age is that it should be possible to obtain information regarding the pathways most affected by age-associated transcriptional dysregulation. A thorough study looking at the gene-specific variation as a function of age would be of interest.
d. Sources of Technical Variability.

There are multiple sources of variability in measurements of gene expression that can mask biologically relevant signals (Parmigiani et al., 2003). These sources of variability can be grouped into two main categories: systematic and stochastic (Huber, 2004). Systematic sources include the amount and quality of the labeled RNA in the sample, dye-specific effects, and proper calibration of the instrumentation used for array manufacture, hybridization, and scanning. Stochastic sources include variability in the quality of the arrays themselves, particularly the DNA on the arrays, nonspecific hybridization, stray signals, inherent variability in labeling and extraction of RNA, and dayspecific effects. For spotted arrays, spotting efficiency and spot size and shape contribute to stochastic variability, and for arrays built in situ, efficiency of incorporation of each base is included in stochastic variability. Systematic errors can be greatly reduced by careful methodology and appropriate instrument calibration. However, stochastic errors are inherent to the microarray system being used and are handled by an experimental design that is balanced across the sources of error and the use of an appropriate statistical error model that factors in the multiple sources of variation.

The most important and correctable source of systematic variability in any microarray experiment derives from the quality of the input RNA. Differences in RNA degradation between two samples can lead to false positives upon comparison of the gene expression profiles that are indistinguishable from true positives even after statistical analysis. Use of a fluidics system such as the Agilent Bioanalyzer for determination of RNA quality is recommended. This system uses nanogram amounts of the total RNA preparation and can quantify degradation in the extracted mRNA or in the 18s and 28s rRNA bands of total RNA (a 28s:18s ratio of 1.3–2:1 is typically considered intact). Although measuring bulk degradation in total RNA does not directly address mRNA quality, intact total RNA profiles are consistently associated with measurements of RNA quality on the arrays, such as the similarity of the signal intensity histogram across samples. Affymetrix arrays, for example, provide indicators of 3' to 5' ratio for two genes using probes that hybridize along the sequence. Additionally, contaminant-free preparation, proper storage at −70 °C, and minimization of freeze-thaw cycles all contribute to RNA quality. Microarray applications are highly sensitive to RNA quality and require extreme attention at this step.

Another important and correctable source of systematic variability is in the methods used for RNA labeling. Direct comparison of array results obtained through different labeling protocols should not be attempted, although differential expression (ratios) may be compared. Because even small deviations in a protocol can produce differences in the resulting signal intensities, simultaneous labeling of all samples compared in a study is preferred. When this is not possible, simultaneous labeling of samples that are balanced across the experimental conditions will minimize the bias introduced by labeling. For experiments where two or more samples are labeled with different dyes and compared directly on the same chip, systematic differences in the incorporation of dyes during RNA labeling can also lead to large artifacts. One method to control for this variability is to implement a dye flip control for each sample (see section II.A.3.c).

3. Managing Variability in Microarray Experiments

a. Replication and Sample Size Considerations. Like any other scientific experiment, the need for replication in microarray gene expression studies is well established (Lee et al., 2000b). Broadly speaking, there are two types of
replication: biological replicates and technical replicates. In addition, there are two main types of technical replicates: independent oligonucleotides or cDNA products representing the same gene present in multiple locations on the array, and the use of multiple RNA samples from the same source to hybridize onto multiple arrays. Only the biological replicates allow for making some inference about the population from which the individuals are drawn. Technical replicates, on the other hand, allow for the determination of measurement error or “noise.” When measuring technical variability in a platform, it is common to measure the same labeled RNA sample on multiple arrays. However, there is a stochastic component to the RNA extraction and labeling, even when systematic variability is minimized, and use of independently prepared RNA samples will incorporate this aspect of technical variability. Even so, estimates of the technical variability will always under-represent the true variability of the system, which includes large biological components.

Because the goal of a microarray experiment is to identify differentially expressed genes, it is important to have enough replicates to keep the probability of having false positives as low as possible. Having sufficient replicates to examine the probability of random variability accounting of “positive” results is an intrinsic aspect of algorithms for estimating false discovery rates (see below). There are several standard approaches for calculating how many replicates are needed when the variances of differential gene expression are known. For example, see Cui and Churchill (2004), Lee and Whitmore (2002), or Parmigiani and colleagues (2003).

**Figure 11.2** Power for detecting a two-fold change between two treatments at various combinations of number of mice per treatment (biological replicates) and number of arrays per mouse (technical replicates). Circles, triangles, squares, and diamonds represent 2, 4, 6, and 8 mice per treatment, respectively. Dotted lines represent the same number of array pairs (8 or 12) for each treatment. Significance level is 0.05 after Bonferroni correction. Biological and technical variance components are estimated from Project Normal data http://www.camda.duke.edu/camda02/datasets. The plotted data is derived from Cui and Churchill (2004).
Churchill (2004), shows the power for detecting a two-fold change for various combinations of biological and technical variations.

Sample size considerations are especially important in aging studies because the variability between individuals and between tissues can be many times larger than the changes due to aging; that is, individual variability may overwhelm changes due to aging alone. In general, biological variability is greater in higher eukaryotes, and greater in outbred, rather than inbred, organisms (although it is possible that F1 hybrids may have reduced variability compared to the parental strains; Phelan & Austad, 1994). It has been suggested, for example, that a minimum of six individual mice are required to reduce sampling errors to satisfactory levels in cDNA array experiments with mice (Cui & Churchill, 2004) and seven individual mice to detect a 1.5-fold difference in 95 percent of genes at the 0.01 level of significance with 90 percent power (Han et al., 2004). However, experiments examining expression differences in human tissues show that at least 36 individuals would be required to obtain similar results in human experiments due to the increased variance present in the human population (Han et al., 2004).

b. Microarrays on Individuals Versus Sample Pools. While increasing the number of arrays used in a microarray experiment increases the statistical power for detecting population differences, the cost of processing a microarray, along with the need to validate results by an independent technique, has driven researchers to look for methods to achieve similar data from a smaller number of arrays.

Pooling samples has been proposed as a strategy to identify genes displaying differences in mean expression between groups. Though pooling masks underlying biological variability, a more accurate estimate of the population mean can be determined when independent pools are used in an experiment with a fixed number of arrays. This is because more individuals from the population can be assayed than if an experiment used a single individual per array. However, this technique is highly sensitive to outliers because a single outlier may skew the perceived gene expression for an entire group. Furthermore, caution must be used when pooling across animals, organs, or even a heterogeneous single organ because small changes present in a region-specific manner will be undetectable. It has been demonstrated that when using a pooling strategy, apportioning individual samples into multiple smaller independent pools (a given sample is added to only one pool) can provide useful information on the biological variability, which cannot be obtained if all samples are pooled together and only technical replicates are performed (Kendziorski et al., 2004). Technical replicates also do not provide satisfactory estimates of interassay variance, which are used to determine statistical significance and false discovery rates; this challenge has not yet been addressed in pooling experiments.

Pooling is a standard strategy when samples are from very small organisms such as yeast and invertebrates (Drosophila and C. elegans), where acquisition of RNA from a single organism is difficult and leads to an extremely low yield. This is discussed further in section II.A.2.

c. Design of Two-Channel Arrays. Array platforms such as the Affymetrix GeneChip® use a single fluorescent signal, and all replicates must be hybridized to separate chips. Two-channel spotted cDNA microarrays, however, allow investigators to perform direct comparisons of two samples on the same array; these also pose more of a challenge for determining ideal experimental design.
For more than two samples, a widely used approach is the reference design, which compares all samples to a common reference. However, as pointed out by Kerr and Churchill in a series of papers, there are more sophisticated alternatives, such as the loop design, that are more efficient for certain types of experiments, such as those with fewer than 10 samples. Because there are differences in the incorporation of the two fluorophores, leading to systematic dye-specific bias, a dye flip design (replicates balanced across both fluorescent channels) is recommended. The importance of a good design cannot be overemphasized because it is a major factor in the estimation of precision as well as the power to detect differential gene expression. Reviews of two-channel design issues can be found in Churchill (2002), Kerr (2003), Kerr and Churchill (2001a), Kerr et al. (2000), Lee et al. (2000b), Parmigiani et al. (2003), and Yang and Speed (2002).

4. Available Technologies for Microarrays and Validation

a. Microarray Technologies. There are two main microarray technologies in current use: cDNA spotted glass arrays and oligonucleotide arrays (both in situ synthesized and spotted) (Holloway et al., 2002). Each type has its advantages and disadvantages. Spotted cDNA arrays usually offer the advantage of lower cost, whereas oligonucleotide arrays have much higher specificity (Hughes et al., 2000). There are three main factors underlying these differences: (1) cDNA products may be recombined or contaminated such that sequences from multiple genes may be present in a single spot; (2) oligonucleotide sequences can be chosen to distinguish from among related gene family members, which cDNA sequences frequently do not; and (3) cDNA products are often both variable and large in size (close to 1 kb), leading to inconsistent hybridization. Commercial oligonucleotide microarrays have seen dramatic price reductions over the last two years, which makes them more attractive for academic research. There are three main commercial oligonucleotide platforms: Affymetrix GeneChip® microarrays, Agilent Oligo microarrays, and Amersham Biosciences [GE Healthcare] CodeLink™ Bioarrays. Excellent reviews of the issues involved in both oligonucleotide and cDNA microarray platforms can be found in Li and colleagues (2003) and Parmigiani and colleagues (2003).

b. Validation of RNA Expression Levels. In order to be convinced that changes in gene expression associated with aging are biologically relevant, validation is required. The first step is to ensure that the RNA levels measured by the microarray experiment can be validated by an independent technique. Real-time quantitative RT-PCR and northern blotting are the primary methods used to validate candidate gene expression changes. Quantitative RT-PCR is considered the “gold standard” due to its sensitivity, reproducibility, and large dynamic range. The development of high throughput technologies for quantitative RT-PCR including the Taqman Low Density Arrays by Applied Biosystems has led to the possibility of independent validation of many candidates from initial screening experiments.

c. Biological Validation. Once candidate genes and pathways have been identified and validated, it is important to determine whether the observed gene expression changes are biologically relevant. This has been accomplished to varying degrees. Confirmation of changes at the protein level is crucial, particularly in aging experiments, because of changes in RNA and protein stability and changes in translational efficiency with age (Brewer, 2002;
Ekstrom et al., 1980), and thus gene expression changes may not correspond to protein levels. Furthermore, measurement of biochemical activity of any candidate genes is also extremely important because of potential alterations in post-translational modification, inactivation, and degradation of proteins with aging. These caveats suggest that microarray studies are useful for the generation of hypotheses, but, particularly in aging, investigators must test whether the changes observed at the RNA level result in functional changes before clear biological conclusions can be drawn. It is worth noting, however, that although biological relevance is desirable, it is not a prerequisite for the identification of biomarkers of aging or longevity. In order for a gene expression biomarker to be useful, all that is required is a high degree of correlation and reproducibility.

An additional level of biological validation is using the gene candidates identified by microarrays to test hypotheses about the aging process. This is most ideally done through genetic manipulation of the model organism. For example, genes that are upregulated in long-lived organisms should increase life span when overexpressed by genetic manipulation, if they are functionally relevant. Likewise, functionally relevant genes that are down-regulated in long-lived organisms should increase life span when expression or function is decreased [e.g., by deletion or RNAi]. This type of phenotypic validation has been carried out with much success in C. elegans and represents the most convincing demonstration of microarrays as a tool to study the aging process to date (Murphy et al., 2003). Particularly in more advanced organisms, the timing, tissue specificity, and quantitative level of upregulation and downregulation of identified key regulatory genes will likely play a role in the extent of phenotypic validation of candidate longevity-associated genes. As biological validation of microarray results from aging experiments becomes more common in C. elegans and other model systems, it will be possible to more accurately assess the gene expression changes that represent important aging-related biochemical pathways conserved through evolution.

B. Informatics Approaches to Gene Expression Data in Aging

The analysis of microarray data involves several sequential and parallel steps, as shown in Figure 11.3. The first three steps have been covered in some detail in previous sections. In the following sections, we focus on the pre-processing and analysis phases.

1. Preprocessing: Diagnostics and Normalization

After a set of microarrays has been hybridized and scanned, the images from the scanner need to be preprocessed before performing any statistical analysis. The preprocessing involves visual inspection of the scanner images (often TIFF files), spot quantification, slide diagnostics and quality control, and normalization.

Spot quantification software depends on the type of microarray. For commercial oligonucleotide arrays, the manufacturers offer their own quantification software. Amersham Codelink slides require the Codelink Expression Analysis program. Agilent slides use Agilent’s Feature Extractor program. Affymetrix GeneChips require using GCOS. For two-color cDNA arrays, there are a variety of open-source programs as well as commercial programs (please refer to The Institute of Genomic Research’s Web site for open-source
offerings at http://www.tigr.org). For commercial programs, choices include Axon’s GenePix, BioDiscovery’s Imagene, or CISRO’s Spot, to name a few. Most of these programs allow visual inspection of the slide images to check for defects or damage, and many produce results of similar quality.

After the scanner images have been quantified, the next main step in pre-processing involves some type of diagnostics and/or quality control. Most of the spot quantification programs mentioned previously also perform quality control tasks, such as calculating mean signal strength, background thresholds, and control spot statistics. From these types of calculations, poor quality slides can be identified and excluded.

Normalization, the process of removing the uninteresting variability within the quantified images, also requires specialized software. There are many programs available that will take a spot quantification file as input, and perform normalization as well as other statistical tasks discussed in the next section. The specialized programs mentioned above have these statistical capabilities in addition to their image-analysis features. However, there are also open-source programs available that allow for the custom analysis of data. For example, Bioconductor (http://www.bioconductor.org)
allows the input of raw quantification data. There are many normalization routines available for any type of microarray platform, from Affymetrix to two-channel cDNA arrays. Bioconductor can also perform more sophisticated multichip normalizations, such as robust multiarray analysis, RMA, GC-RMA, variance stabilization, VSN, and dChip. Each of these methods takes a different approach to adjusting signal intensities to account for nonspecific hybridization, optical noise, and between-array variations. RMA is explained in more detail by Irizarry and colleagues (2003b), where it is shown that for Affymetrix arrays, subtracting the mismatched probes from the perfect match probes results in an exaggerated variance. Their RMA method does a background adjustment that ignores the mismatch probes. GC-RMA is explained in more detail by Wu and colleagues (2003), where it is shown that using the GC content of the mismatch probes improves the background adjustment. VSN is explained in more detail by Warner and colleagues (2002), where variance stabilizing transformations are used to normalize the microarray data. dChip is explained in more detail by Li and colleagues (2003), where a model-based expression analysis is used that normalizes Affymetrix array data based on an invariant set of genes. Each of these methods performs well on a variety of performance metrics. For further comparisons, see http://affycomp.biostat.jhsph.edu for details about a list of benchmarks for Affymetrix GeneChip expression measures.

Preprocessing of microarrays is discussed in more detail in publications by Parmigiani (2003) and Yang and Speed (2002). For the normalization of Affymetrix arrays, RMA has been demonstrated to be more robust than other available programs for identification of spike-in RNA samples of known concentrations [Irizarry et al., 2003a]. GC-RMA, which uses information about the GC content of the oligonucleotide sequences, is able to improve on the RMA algorithm to maintain precision while improving accuracy (Wu and Irizarry, 2004).

2. Statistical Methods for Identifying Differential Gene Expression

Generally speaking, there are four steps in the identification of differentially expressed genes. First is the choice of the appropriate statistical model, which is used to calculate the average intensities of gene expression for each gene across replicates and the sample variance for each gene. Appropriate models will be suggested by the experimental design and can include mixed effects models if higher-order structure is present in the data, such as groupings of covariates (Churchill, 2002). Second is the calculation of the test statistic. If only two groups are to be compared, then t-tests are useful. For more groups, some type of analysis of variance approach is more appropriate. See for example Churchill (2002), Kerr and Churchill (2001a,b), Kerr and colleagues (2000), Lee and colleagues (2000b), Parmigiani and colleagues (2003), and Yang and Speed (2002).

For data with more structure, such as balanced and unbalanced data, or when the within-group correlation is important in grouped data, mixed effects models are important (see Pinheiro and Bates, 2000). Several investigators prefer a modified version of the standard t statistic that uses information borrowed from all the genes on the array to estimate the individual gene variance [Efron and Tibshirani, 2002; Smyth, 2004; Storey and Tibshirani, 2003a]. This modified test statistic approach is useful for prevention of calculating spuriously significant genes. As previously reported (Qin and Kerr, 2004), in the analysis of spike-in
experiments on two-color cDNA arrays, it has been found that \( t \)-tests performed the worst among test statistics for correctly identifying the rankings of the spike-in genes, whereas modified versions of the \( t \) test were much more robust. Furthermore, it was found that for some data sets, the simple median of the log ratio across arrays performed best at correctly ranking the spike-in genes (Qin and Kerr, 2004).

The calculation of unadjusted p-values is the third step. This involves the calculation of the null distribution for the test statistics and the selection of rejection regions (symmetric or one-sided). The fourth step requires some reasonable approach to controlling the number of falsely positive genes. When testing a hypothesis, one can make either a Type I error (calling the gene significant when it is not, a false positive) or a Type II error (calling the gene not significant when it is, a false negative). However, in microarray analysis, multiple hypotheses are being tested, so it is not clear how best to specify the overall error rate. As pointed out by Storey and Tibshirani (2003b), there are a spectrum of choices. At one end are unadjusted p-values, which result in far too many false positives (if you have an array with 50,000 genes and a p-value cutoff of 0.01, there are possibly 500 genes that are false positives). At the other end is the standard Bonferroni correction to control the family-wise error rate, which is far too conservative and results in large numbers of false negatives. What has been found to be most useful in the microarray context is the False Discovery Rate (FDR), or the positive False Discovery Rate (pFDR) or q-value. With this approach, the investigator can control the number of false positives in the number of genes called significant, rather than controlling the number of false positives out of all the genes present, by examining the distribution of p values. A detailed discussion of these issues can be found in Storey and Tibshirani (2003a).

Software programs that do most or all of the steps described above are now present in many commercial analysis programs. One popular program is SAM (Statistical Analysis of Microarrays; see Storey and Tibshirani, 2003a,b). However, identification of differential gene expression is a very active research field. To stay most current with rapid changes in the field, we also recommend using the microarray statistical research tool Bioconductor, where many new algorithms first appear.

3. Data Visualization

The visualization of large data sets in order to discovery underlying patterns is an important aspect of microarray analysis. The goal is to use a dimension reduction algorithm such as Principle Component Analysis, Singular Value Decomposition, or multidimensional scaling to capture the essential variations in the data set in just two or three dimensions. However, an inherent danger is present in such a reduction of dimensionality, as valuable information may be lost (Slonim, 2002).

Another important goal in microarray analysis is the assignment of biological samples into groups based on their expression patterns. This process of assignment is broken out into two approaches: unsupervised and supervised learning. In unsupervised learning, also known as clustering, the groups or classes are discovered from the data, as they are not known beforehand. In supervised learning, also known as classification, the groups or classes are already known or are predefined. The task is then to predict the class of a new set of experiments.

Classification methods fall into three main groups: class comparison, class prediction, and class discovery. A discussion of statistical issues of classifiers such as types of classifiers, usefulness and limitations, classifier accuracy
improvement, and performance of five main classifiers (k-nearest neighbor, naïve Bayes, logitBoost, random forests, and support vector machines) is given by Dudoit and Fridlyand (2003a,b). In Dudoit and Fridlyand (2003a), these classifiers are applied to several real data sets, and their performance is assessed. They showed that the simpler methods, such as k-nearest neighbor and naïve Bayes, were competitive with the more complex methods and are advisable for the more inexperienced user. However, they also showed that for larger data sets, the more complex methods performed best.

Clustering methods, which can be more difficult than classification, are used to group genes into clusters with similar behaviors, and usually begin with not knowing how many groupings there are in the data. Methods include hierarchical clustering, k-means clustering, self-organizing maps, and model-based approaches. Often it is useful to filter the data so as to cluster only a useful subset of genes. Such filtering can include simple nonspecific filtering such as removing low-intensity genes or more sophisticated Principle Component Analysis that is used to reduce the high-dimensionality of the data set to generate gene lists that account for most the differential changes. (Reviews of relevant clustering issues, algorithms, and software can be found in Chipman et al., 2003, Do et al., 2003, and Sebastiani et al., 2003). When repeated measurements are taken into account, it is possible to achieve more accurate and stable clusters (Yeung et al., 2003). Keep in mind that clusters will always be found by the algorithm, even if there are no true clusters in the underlying data. Choosing the most appropriate clustering method will depend on the data, design, and goals of the research (Slonim, 2002).

A useful way to think about clustering methods is that they are different ways of looking at a data set. Each one gives a somewhat different view of a complex problem, and, therefore, each is useful in its own way.

Many open-source software packages such as Bioconductor, TIGR MultiExperiment Viewer, Eisen Lab’s Cluster (Eisen et al., 1998), as well as commercial programs such as Silicon Genetics GeneSpring®, Iobion’s GeneTraffic®, SAS® Microarray, Insightful’s S + ArrayAnalyzer™, Rosetta Resolver®, and VizX Lab’s GeneSifter™, to name a few, have a wide variety of clustering capabilities.

4. Gene Ontology Mining, Pathways Analysis, and Systems Biology

It is now possible to use microarray gene expression data to identify groups of genes in common gene ontology categories and thereby uncover biological processes and pathways. This visualization and analysis ability requires specialized software, such as the open-source programs GoMiner and GenMAPP (see Dahlquist et al., 2002, and Zeeberg et al., 2003). The Gene Ontology (GO) consortium (http://www.geneontology.org) has attempted to construct a standardized structure for functional categorization of genes. While there is a strong need for such a standardized categorization, the assignments are still changing as more research is conducted on gene function.

In addition to the open-source software, there are also commercial software programs that allow for the visualization of biological pathways, gene regulation networks, and protein–protein interactions, such as Iobion’s Pathway Assist and Ingenuity System’s Pathways Analysis. These programs are based on scans of the biomedical literature, and this is a changeable, active area of current research.

At the systems biology level of analysis, there is the open-source software program Cytoscape, which is used in conjunction with protein–protein, protein–DNA, and genetic interactions
5. Expression Databases and Meta-Analyses

High-throughput technologies such as DNA microarrays generate enormous amounts of data. This necessitates the use of data management systems for the storage and retrieval of experiments. Both open-source programs as well as commercial products are available. Most, if not all, of these data management systems also have a data analysis capability. These data management and data analysis systems come in client/server form or as standalone, Web-based, Windows, Macintosh, or various flavors of Unix/Linux. Of particular importance in choosing data management and data analysis systems is the overall cost of the software. Open-source systems have low to zero initial cost, but maintenance and support usually require more experienced bioinformatics and/or programming staff. Commercial systems usually have a much higher initial cost, but that cost usually covers user training and a support help desk. Furthermore, commercial vendors can also customize their products for some additional cost, whereas customization of open-source software is a job for your programming staff. Commercial products tend to be much more user friendly, coming equipped with easy-to-use graphical user interfaces. Open-source products, especially ones such as Bioconductor, have a much larger availability of statistical algorithms for use on complex experimental designs.


Commercial systems include Silicon Genetics’ SpringCore™, Iobion’s GeneTraffic®, SAS® Microarray, Insightful’s S + ArrayAnalyzer™, Rosetta Resolver®, BioDiscovery’s GeneDirector™ and VizX Lab’s GeneSifter™, to name a few.

Public databases for the storage and retrieval of microarray data are available. These include the NIH’s Gene Expression Omnibus GEO at http://www.ncbi.nlm.nih.gov/geo/ and the European Bioinformatics Institute’s ArrayExpress at http://www.ebi.ac.uk/ arrayexpress/. These public repositories are becoming more widely used as more scientific journals require microarray data to be deposited in publicly accessible databases.

A new area of opportunity for microarray data analysis is the integration of microarray data generated by different research groups on different array platforms (Moreau et al., 2003). Furthermore, access to microarray databases also affords opportunities for meta-analyses of cross-species comparisons of expression profiles that allows the study of biological processes and global properties of expression networks (Shah et al., 2004). An ongoing effort at the NIA DNA Array Unit is the development of a Web-based database of biological pathways [http://bbid.grc.nia.nih.gov], which is used to relate gene expression studies to complex biological processes. In addition, a second project includes a Web-based database of the genetics of common complex diseases.

III. Biological Studies

A. Gene Expression in Yeast

The budding yeast *Saccharomyces cerevisiae* has been used as a model system to investigate two fundamentally different types of cellular aging processes.
The study of chronological aging involves maintaining cells in a metabolically active, nondividing state and monitoring the decrease in viability with time, perhaps akin to the aging of postmitotic cells in mammals (Fabrizio and Longo, 2003; MacLean et al., 2001). Replicative aging, in contrast, is defined by the number of mitotic cycles completed by a mother cell prior to senescence (Mortimer and Johnston, 1959).

Several genes have been identified that determine chronological or replicative life span; however, the relationship, if any, between these two aging processes remains unclear.

Yeast represents an attractive model for using microarrays to study the aging process. With a doubling time of less than two hours, yeast is amenable to both classical genetic as well as high-throughput approaches. Much of the microarray technology development was carried out in studies of this organism (e.g., by Chu et al., 1998; DeRisi et al., 1997; Eisen et al., 1998; Lashkari et al., 1997; Shalon et al., 1996; Spellman et al., 1998), resulting in a large body of knowledge regarding appropriate design and an abundance of publicly available expression data (Horak and Snyder, 2002). In addition, metabolic and protein interaction pathways are relatively well characterized compared to other model systems (Barr, 2003). Also, unlike the case in multicellular eukaryotes, the potential complications arising from tissue heterogeneity and cell specificity do not apply to studies in yeast (see section II.A.2).

It is worth noting that one significant difference between microarray studies of aging in yeast compared to other systems is that almost all have been carried out using spotted cDNA arrays. The relative advantages and disadvantages of this type of platform are discussed in section II.A.4.a.

1. Gene Expression Changes Associated with Replicative Age in Yeast

As in other model systems, microarray analysis has been used as a tool to identify gene expression changes associated with old age in yeast. In theory, the design of such an experiment is identical to similar studies in worms, flies, and mice: RNA is obtained from old organisms and compared against RNA obtained from young organisms. Attempting to obtain a pure population of replicatively aged yeast, however, presents a unique challenge not present in other models. The median life span of a typical lab strain is approximately 25 generations (Jazwinski, 1993). Thus, in order to obtain a single mother cell aged to the population median, that cell must be physically separated away from her \( \sim 2^{25} \) “progeny” cells. For lifespan analysis, micromanipulation is used to separate daughter cells from 40 to 50 mother cells per strain; however, it is not feasible to use micromanipulation as a method to obtain enough cells for even one microarray experiment without extensive, and possibly biased, RNA amplification. Two technologies are currently available to obtain large populations of aged cells: magnetic sorting and elutriation. In the magnetic sorting procedure, an unsorted population of cells is treated with biotin then allowed to grow for several generations prior to addition of streptavidin-coated magnetic beads (Park et al., 2002; Smeal et al., 1996). Subsequent magnetic separation allows enrichment of aged mother cells, which specifically retain biotin on their cell walls. Elutriation is a centrifugal technique that separates cells based on cell size (Helmstetter, 1991). Generally, \( G_0 \) daughter cells are the smallest cells in a population, and older cells are the biggest. Thus, a population enriched for bigger cells by elutriation also tends to be enriched for aged cells (Woldringh et al., 1995). Both magnetic sorting and elutriation, however,
suffer from the drawback of significant contamination by daughter cells, a potential source of experimental noise that must be considered in any gene expression study of “old” versus “young” yeast.

Two studies have described the transcriptome of replicatively aged yeast. In one study, microarray analysis was carried out on “young” (0–1 generation) or “old” (7–8 generation) wildtype cells obtained by magnetic sorting (Lin et al., 2001). In addition, young and old cells from a 20 percent shorter-lived sip2Δ mutant or a 15 percent longer-lived snf4Δ mutant (Ashrafi et al., 2000) were examined. From this analysis, it was concluded that gluconeogenesis and glucose storage increase as cells age, suggesting a metabolic shift away from glycolysis and toward gluconeogenesis. Microanalytic biochemical assays were used to verify changes in enzyme activity consistent with the microarray results. In addition to questions regarding the purity of the aged cell population, however, the validity of defining 7–8 generation cells as “old” must be questioned. The median and maximum life spans of the strain background used in this study (S288C) are approximately 25 and 50 generations, respectively (Kaeberlein et al., 2004). Thus, the use of 7–8-generation mother cells to identify gene expression changes associated with aging in yeast is analogous to using 7-month mice or 20-year-old humans as the aged population for similar studies in these organisms.

In the second study of this type, elutriation was used to obtain an aged population in which 75 percent of the cells were at least 15 generations old and 90 percent of the cells were more than 8 generations old (Lesur and Campbell, 2004). Microarray analysis of aged cells relative to young cells suggested an increase in expression of enzymes associated with glucose storage and gluconeogenesis, consistent with the previous study (Lin et al., 2001). In addition, certain stress- and damage-responsive genes were also reported to be elevated in aged cells (Lesur and Campbell, 2004). Although the average age of the “old” cells is a significant improvement over the first study of this type, it is still well below the population median. Further, both studies use an arbitrary two-fold cutoff to identify differentially expressed genes and suffer from a lack of rigorous statistical analysis. Future studies of this type should attempt to address both of these weaknesses. The fact that both studies suggest an age-associated shift toward glucose storage at ages below the median life span of the population, however, is suggestive that major metabolic changes occur early in the yeast life span. It will be of interest to determine whether these changes are retained, or perhaps enhanced, at later replicative ages.

2. Gene Expression Profiles of Long-Lived Yeast Strains

In addition to age-associated gene expression studies, microarrays have also been used to examine gene expression changes that correlate with extreme longevity in yeast. These types of experiments are simpler to perform than those described above. Because the proportion of aged cells present in a logarithmically growing yeast culture is exceedingly small (less than 1 per 2^n cells, where n equals replicative age), RNA can be harvested from an unsorted population of long-lived cells and compared against RNA from an unsorted population of wildtype cells. In theory, this approach provides an opportunity to identify molecular mechanisms of enhanced longevity in individual long-lived mutants. In addition, comparison of multiple long-lived mutants has the potential to identify gene expression biomarkers of longevity.

This type of analysis has been successfully performed using genetic models of longevity as well as environmental models of longevity, such as CR by growth on low glucose. In one study, microarray analysis was carried out on
two models of CR in yeast: growth on low glucose and deletion of the gene coding for hexokinase, HXK2 (Lin et al., 2002). Based on the observed gene expression changes, it was suggested that CR of yeast cells results in a transcriptional shift from fermentation to respiration. These findings were verified by follow-up experiments showing that oxygen consumption is elevated by CR. In addition, based on the overlap between gene expression changes observed in cells lacking HXK2 and cells grown on low glucose, 124 putative gene expression biomarkers of CR were reported. It should be noted, however, that no evidence has been presented to suggest that the observed gene expression changes play a causal role in CR-mediated lifespan extension.

To date, a large-scale comparative microarray analysis of multiple long-lived mutants is lacking. At least two additional studies have compared individual gene expression data sets from cells with enhanced life span against the CR data sets described above. In one case, the transcriptional changes associated with high external osmolarity showed significant overlap with the CR data set (Kaeberlein and Guarente, 2002). In the other study, gene expression changes associated with addition of SSD1-V, which increases mean replicative life span by approximately 75 percent, showed no significant overlap with CR (Kaeberlein and Guarente, 2002). These results were interpreted to suggest that osmotic stress response, but not SSD1-V, is likely to promote longevity by a mechanism similar to CR. In both of these studies, as well as the CR study described above, only two data sets for each genotype were obtained, and an arbitrary two-fold cutoff was used to classify genes with significantly altered expression. Rigorous statistical analysis and additional data sets would improve confidence in the individual genes reported. The use of gene expression changes across a subset of genes to place long-lived mutants into genetic pathways, however, is an approach with the potential for broad applicability as additional data sets are obtained.

3. Chronological Aging in Yeast

To date, the use of microarrays to investigate yeast aging has been largely confined to studies of replicative aging. The reasons for this dichotomy are unclear, as chronological aging would seem to present a system amenable to microarray technology. Chronological life span is determined by culturing cells into stationary phase and monitoring the percentage of cells that retain viability over time (Fabrizio and Longo, 2003; MacLean et al., 2001). Unlike the case with replicative age, studies examining gene expression changes associated with increased chronological age are trivial in design. In fact, one of the pioneering microarray studies examined the gene expression changes associated with the yeast diauxic shift, a transition from logarithmic growth to stationary phase (DeRisi et al., 1997). A similar design, but with timepoints spaced over several weeks, could be used to identify gene expression biomarkers of chronological age. As with replicative aging, microarray studies comparing chronologically long-lived mutants to wildtype cells could also be used to potentially identify gene expression biomarkers of chronological longevity. In this regard, it would be of particular interest to determine whether there are significant similarities between the transcriptional changes associated with enhanced chronological life span and those associated with enhanced replicative life span.

4. Mutation of Orthologous Genes in Yeast

In addition to directly studying the function of genes that regulate chronological or replicative life span, an alternative
approach is to use microarrays to study the function of yeast orthologs of proteins that affect aging in higher eukaryotes. For example, increased expression of the heat shock transcription factor Hsf1 has been found to increase life span in worms (Hsu et al., 2003; Morley and Morimoto, 2004). Although, no chronological or replicative life span phenotype has been reported in response to Hsf1 overexpression in yeast, a recent study employed microarray technology to identify a majority of the direct transcriptional targets of yeast Hsf1 (Hahn et al., 2004). Analysis of this data to determine which Hsf1 targets have worm orthologs is likely to provide information regarding potential downstream effectors of the enhanced longevity conferred by HSF-1 overexpression. This type of approach would be particularly amenable to highly conserved pathways or genes that are likely to behave similarly in higher eukaryotes.

5. Conclusions from Yeast Studies

Microarray analysis has provided valuable insight into certain aspects of the aging process in yeast, particularly the identification of global metabolic changes associated with calorie restriction and replicative age. However, this technology has not been used to its fullest potential. The use of arbitrary fold cutoffs and lack of rigorous statistical analysis, in particular, have been limitations of studies to date. Future studies should strive to improve in these areas. Given the relative ease with which a compendium of gene expression data sets could be generated for all of the mutants reported to increase yeast life span, this should become a priority for future work. In this way, it will be possible to rapidly identify gene expression changes correlated with longevity on a genome-wide scale, something for which yeast is uniquely suited.

B. Gene Expression in Invertebrate Models

A number of studies have employed microarrays to study aging in the invertebrate model systems *Drosophila melanogaster* and *Caenorhabditis elegans*. As is the case with yeast and mammalian model systems, two general strategies have been employed to identify genes important for the aging process. In the first, microarrays have been used to look for genes that are differentially expressed in aged animals relative to young animals. In the second type of study, microarrays have been used to identify genes differentially expressed in long-lived versus short-lived animals. The merit of each of these approaches is discussed in detail in section II.A.1.

Although there is still substantial disagreement as to which specific genes are differentially expressed, and even how many genes are differentially expressed, there is an emerging consensus that certain types of genes show specific changes in gene expression with age. In particular, several studies have suggested that expression of stress-response genes is elevated in old animals, and that these types of genes are also upregulated in long-lived mutants. There is also evidence that metabolic genes are downregulated during aging. In this section, we review selected microarray studies to demonstrate these themes. We also highlight discrepancies between various studies to demonstrate the limitations of current technology.

1. Overexpression of Stress Response Genes During Aging in *Drosophila*

Several studies in *Drosophila* have reached the conclusion that stress-response genes are overexpressed during aging (Zou et al., 2000). Zou and colleagues used dual-channel cDNA arrays and probes prepared from pooled male *Drosophila* samples of diverse ages to
identify genes differentially expressed during aging. The data analysis for this early study was nonstatistical and based on ratio-dependent selection of genes followed by hierarchical clustering. From this analysis, it was reported that expression of certain key metabolic genes, including glucose-3-phosphate dehydrogenase and cytochrome c, are decreased with age. An age-associated upregulation of some stress-related genes, including glutathione-s-transferase 1, was also observed. To test whether this represented a general response to oxidative stress, the gene expression profile of young animals treated with the superoxide-generating drug paraquat was obtained and compared to that of aged flies. Intriguingly, approximately one-third of genes differentially expressed in response to the oxidative challenge were also differentially expressed with normal aging.

More recently, several other studies have reached similar conclusions. Pletcher and colleagues (2002) used Affymetrix arrays containing the majority of Drosophila open reading frames to look for differentially expressed genes. For this study, RNA was collected only from female flies and analyzed in a pooled fashion. Samples from both control-fed and calorically restricted animals were collected at multiple time points. A sophisticated statistical analysis with false discovery rate adjustment was used to identify differentially expressed genes. Nine percent of genes on the array were differentially expressed at one or more timepoints during aging. Differentially expressed genes were then mapped against the GO gene functional ontology (http://www.geneontology.org/). Based on this analysis, stress response, antibacterial, and serine protease inhibitor functional categories were upregulated during aging, and oogenesis genes were downregulated during aging. In calorically restricted flies, genes involved in the cell cycle, DNA repair, DNA replication, protein metabolism, and protein degradation were downregulated. Fifty percent of age-dependent changes in gene expression were ameliorated by long-term CR.

Landis and colleagues (2004) performed a microarray study with the aim of identifying gene expression biomarkers of aging in Drosophila. As with Pletcher and colleagues (2002), Landis and colleagues (2004) used Affymetrix arrays coupled with a sophisticated statistical analysis to identify differentially expressed genes. RNA was obtained from male flies and pooled at a variety of aging timepoints between 10 and 61 days (at the 61-day timepoint, 50 percent of the cohort was surviving). Using Significance Analysis of Microarrays (SAM), 7 percent of genes on the array were differentially expressed at one or more timepoints, in good agreement with the Pletcher study. They also observed upregulation during aging of stress-response genes, including antioxidant genes, antibacterial genes, and some heat shock proteins (hsp 22). In addition, genes coding for enzymes in the purine biosynthetic pathway were upregulated, whereas protease, proteasome, and metabolic genes were downregulated.

Landis and colleagues (2004) also exposed young flies to 100 percent oxygen to induce antioxidant genes. They observed that there was a 33 percent overlap between genes differentially expressed in response to oxygen and aging. In agreement with the fact that Drosophila life span scales with the temperature at which they are raised, the age-associated upregulation of antibacterial genes also scaled with temperature, suggesting that these genes may be good aging biomarkers in Drosophila.

2. Overexpression of Stress Response Genes in Long-Lived Invertebrates

The studies described thus far indicate that stress-response genes, such as antioxidant genes and antibacterial genes, are
induced during aging in flies. Kang and colleagues (2002) observed that treating flies with the histone deacetylase inhibitor PBA increased both mean and maximum life span in two strain backgrounds. They used EST spotted nylon membrane microarrays to identify genes upregulated in PBA-treated flies. Genes upregulated in the long-lived animals included stress-response genes such as MnSOD, glutathione-s-transferase, and several chaperonins.

Several important studies using *C. elegans* have also tried to identify genes differentially expressed in short-versus long-lived animals [McElwee et al., 2003; Murphy et al., 2003]. These studies are based on the observation that an IGF/insulin-like signaling pathway functions in early adult *C. elegans* to regulate life span.

Hypomorphic mutations in the DAF-2 insulin-like receptor double life span and increase the stress resistance of mutant worms. Mutations in the downstream DAF-16 FOXO transcription factor block the effects of DAF-2 mutations. McElwee and colleagues (2003) used cDNA microarrays, spotted with the majority of *C. elegans* genes, to identify genes differentially expressed between young (1 day) DAF-2 and DAF-2/DAF-16 pooled worm samples. In both cases, the worms had additional mutations to render them sterile. In the long-lived DAF-2 worms, upregulation of mitochondrial superoxide dismutase sod-3, as well as several heat shock proteins, was observed. In addition, substantial overlap between the differentially expressed genes and genes with DAF-16 binding sites in their promoters was reported. In order to verify the relevance of observed gene expression changes for increased life span in DAF-2 animals, RNAi was used to knock down the activities of the differentially expressed ins-7 gene, and this lengthened life span.

A similar strategy was employed by Murphy and colleagues (2003). In this study, they collected pooled samples of sterile worms from a variety of aging timepoints as well as day 1 DAF-2 versus DAF-2/DAF-16 worms. These samples were hybridized to spotted cDNA arrays containing the majority of the *C. elegans* genome. A nonstatistical analysis plus hierarchical clustering was used to identify genes that showed differential patterns of expression. In particular, genes upregulated in DAF-2 versus DAF-2/DAF-16 were selected as potential longevity-enhancing genes. Such genes included stress-response genes such as gst-4, sod-3, catalase genes, small heat shock proteins, and antibacterial defense genes. Importantly, using RNAi to inhibit the activity of these genes in DAF-2 mutants resulted in a decrease in life span, suggesting that their upregulation is functionally related to enhanced longevity.

These studies also identified the ins-7 gene product as the likely ligand of the IGF/insulin-signaling pathway. ins-7 expression is decreased in the long-lived DAF-2 mutant but increased in DAF-2/DAF-16 double mutants. RNAi against ins-7 in a DAF-2/DAF-16 background resulted in an increase in life span.

The study of Murphy and colleagues marks an important advance in the field, as it demonstrated convincingly that gene expression changes observed by microarray can lead to the identification of functionally relevant regulators of longevity. In addition, this study enhanced a biological model for the mechanism by which hormonal signals produced in specific cells can act globally to regulate life span in *C. elegans*.

3. Bioinformatic Analysis of Cross-Species Gene Expression Changes During Aging

Collectively, these microarray studies argue that, in general, stress-response genes are upregulated in old animals and that long-lived animals have elevated
expression of stress-response genes. Also, in several of these studies, metabolic genes were noted to be downregulated during aging. In an attempt to synthesize the data from flies and worms, McCarroll and colleagues (2004) used a bioinformatic approach to compare genes differentially expressed during aging in both organisms. McCarroll first identified orthologous gene pairs across the two species and then used microarrays to analyze gene expression during aging in pooled Drosophila heads and in pooled worms at various ages. Orthologous gene pairs showed a limited ($r = .144$), but highly statistically significant ($P < 10^{-11}$), correlation during aging. Perhaps more importantly, McCarroll then mapped the orthologous gene pairs to GO functional categories. GO term mapping of age-associated gene expression changes in worms and flies also demonstrated a significant overlap in functional categories. Although it is important to note that the majority of differentially expressed aging genes were unique to each species, statistically significant enrichment for 14 GO categories was observed. This is many more significant GO categories than would be expected by chance, as permuting the underlying data and redoing the GO analysis showed a false positive rate of 1.4 +/- .91 GO categories. Many of these GO categories were involved in general metabolism and showed a coordinate repression in early adulthood when the IGF/insulin pathway is known to begin regulating life span in C. elegans. This study was the first to attempt to demonstrate the existence of broad phylogenetically conserved similarity in the types of genes showing altered transcription with age as well as specific orthologous gene pairs with similar age-associated profiles.

4. To Pool or Not to Pool Invertebrates?

Due to the small size of the experimental animal, both Drosophila and C. elegans microarray studies have generally pooled a large number of whole animals to study gene expression at each timepoint. This has advantages and disadvantages, as described in section II.A.3.b. A major disadvantage of this approach in invertebrate systems is that changes in gene expression in individual tissues are averaged out over all of the tissues in the body. For example, if a gene were strongly differentially expressed in one small organ but not differentially expressed in other tissues, it is likely that this gene would be missed in current studies.

Another disadvantage of pooling animals is that gene expression differences in individual animals are lost. To eliminate this problem, Golden and Melov (2004) analyzed gene expression changes during aging in individual C. elegans.

5. How Many Genes Are Differentially Expressed During Aging in Invertebrate Model Systems?

We have cited evidence that there are broad similarities in the gene expression changes associated with aging in Drosophila and C. elegans, as well as multiple conserved gene expression changes among orthologous gene pairs. However, there are still substantial disagreements between studies, even within the same species. For example, there is relatively poor agreement on which genes are differentially expressed during aging. There is also disagreement on the magnitude of differential gene expression during aging. For example, using modern Affymetrix arrays and sophisticated statistical analysis, Landis and colleagues (2004) and Pletcher and colleagues (2002) observed 7 to 9 percent of genes as being differentially expressed during aging across multiple timepoints. In contrast, rigorous statistical analysis of the data from Jin and colleagues (2001) suggests that the majority of variance in their expression...
data was due first to strain differences, then to sex differences, and only last to age-specific differences. This observation implies that the magnitude of gene expression changes during aging is relatively modest. Only 1 percent (1.0 percent) of genes in the Jin study showed age-dependent statistically significant differences in expression, compared to 7 to 9 percent in the other studies.

Some of the differences between the studies can be attributed to the differences in experimental design. However, by plotting P-value versus ratio change for the genes using a volcano plot, Jin and colleagues (2001) demonstrated an important aspect of their result: none of the genes identified as showing a statistically significant change in gene expression during aging showed even a two-fold ratio change, and many ratio changes were as little as 1.2-fold.

It is important to note that modern statistical array analysis tools identify genes with very modest ratio changes as differentially expressed. Furthermore, many of the studies we have reviewed do not even show the distribution of observed ratios for their data. This is unfortunate, as the biological significance of, for example, a 1.2-fold ratio change in gene expression is presently unclear. Furthermore, many array data normalization procedures can cause small consistent shifts in the ratio data. This could lead to affected genes being falsely identified as differentially expressed. Thus, the significance of statistically significant but tiny changes in gene expression must be questioned.

It can also be asked whether microarrays are missing key aspects of changes in gene function during aging in both D. melanogaster and C. elegans. As noted above, in most studies, relatively small gene expression changes have been observed in aging. However, biosynthetic activity in both invertebrate and vertebrate models plunges during aging, with decreases in both synthesis and degradation of mRNA and protein. For example, protein synthesis decreases by as much as 60 to 90 percent during aging in Drosophila (Arking, 1998). Such large changes may be highly significant for the aging phenotype and yet are not detected by microarrays, at least in part because current microarray studies examine only relative, not absolute, changes in gene expression.

C. Gene Expression in Rodent Models

Although studies of yeast, nematodes, and fruit flies yield important insight into evolutionarily conserved pathways of aging, the rodent model better represents aging in a complex mammalian system, and still offers time, space, and economic benefits compared to larger mammals. Additionally, the availability of long-lived genetic mouse models (such as dwarf mice deficient in growth hormone signaling), CR, and transgenic and allele replacement mice have allowed for comparisons of gene expression patterns associated with life-span extension in the mammal.

Use of a mammalian model, however, does present additional concerns. Unlike studies of lower eukaryotes, mammalian studies of gene expression changes with age are intrinsically linked to tissue type, as gene expression variability between tissues overwhelms the changes observed with aging. For most laboratories, it is not economically feasible to pursue microarray studies on all tissues, or even several tissues of interest, and, thus, most of the reported studies focus on transcriptional profiles in one or two tissues that may or may not be representative of global gene expression changes. Tissues examined to date include brain (Blalock et al., 2003; Lee et al., 2000a; Preisser et al., 2004; Prolla, 2002; Prolla and Mattson, 2001; Weindruch and Prolla, 2002; Weindruch et al., 2002), skeletal muscle (Lee
et al., 1999; Tollet-Egnell et al., 2004; Weindruch et al., 2001, 2002; Welle et al., 2001), liver (Cao et al., 2001; Dozmorov et al., 2001; Dozmorov et al., 2002; Meydani et al., 1998; Miller et al., 2002; Tollet-Egnell et al., 2004; Tsuchiya et al., 2004), heart (Bronikowski et al., 2003; Csiszar et al., 2003; Edwards et al., 2003, 2004; Lee et al., 2002; Meydani et al., 1998), kidney (Preisser et al., 2004), duodenum and colon (Lee et al., 2001), adipose tissue (Higami et al., 2004; Tollet-Egnell et al., 2004), and submandibular gland (Hiratsuka et al., 2002).

The current body of literature employing microarrays to examine gene expression associated with aging in mammals is relatively young and largely descriptive. However, Helmberg (2001) and Weindruch and colleagues (2002) discuss the enormous potential utility of this approach. The possible outcomes from this work include: (1) insights into the fundamental causes of aging by, for example, the identification of biochemical pathways altered with age; (2) development of tools and biomarkers useful in the evaluation of aging interventions; and (3) perhaps even to define “individual genomic risk constellations” useful in the treatment and management of aging-associated conditions. This review highlights some of the seminal papers to date that employ microarray technology to advance our understanding of aging in mammals.

1. Gene Expression Changes Associated with Age

Lee and associates were among the first to report changes in gene expression profiles with age in mouse tissues, using microarrays to study gene expression patterns in brain (neocortex and cerebellum) and skeletal muscle from adult (5-month) and old (30-month) C57Bl/6NHsd mice (average life span 30 months), as described in a series of papers (Lee et al., 1999; Lee et al., 2000a; Weindruch et al., 2001, 2002; Prolla, 2002). [The CR component of these studies will be discussed later.] For these studies, the authors compared three animals per group using 6347-gene oligonucleotide arrays. The data were analyzed using pairwise comparisons (nine total comparisons) with Pearson correlation coefficients calculated for individual animals and a fold-change cutoff was used to classify genes as significantly upregulated or downregulated.

The principal finding of these studies is that only a small percentage of genes in each tissue (approximately 1 percent) were upregulated or downregulated by at least two-fold (1.7 fold in the neocortex), indicating that aging is unlikely to result from widespread gene expression changes of large magnitude. Furthermore, the authors observed little overlap among the individual genes altered in the tissues examined, although they did observe coordinate induction of complement cascade members and cathepsins in the neocortex and cerebellum with age. Cathepsins may be of particular interest as they are involved in the processing of amyloid precursor protein and are upregulated in Alzheimer’s diseased brains. Of the genes reported to be upregulated with age in the tissues examined, the greatest proportion fell into the functional categories of inflammation and stress response, whereas decreased expression was observed for genes involved in metabolism and biosynthesis.

Although these early studies are somewhat limited by their lack of rigorous statistical analyses and validation, they were among the first to illuminate trends that seem to be recurrent in subsequent microarray studies—namely, that relatively few transcriptional changes are of great magnitude, and that transcriptional profiles across tissues and even across subregions of a particular tissue...
are markedly different. Furthermore, the initial findings in these studies, that genes involved in stress response and inflammation seem to increase with age while genes involved in protein turnover and structural maintenance decline, are also echoed in later studies.

Other studies have also analyzed age-associated gene expression changes (young versus old) in conjunction with gene expression changes associated with enhanced longevity. For example, Dozmorov and colleagues (2001) measured gene expression changes with age in livers of 5-, 13-, and 22-month-old mice (n = 3 to 4 mice per group). This study utilized both control and Ames dw/dw dwarf mice, but presently this discussion will focus on the analysis of the subset of genes altered with age. The authors measured gene expression levels using Atlas 588-gene cDNA membranes, 323 genes of which were removed from analysis due to low levels of expression or proximity on the array to highly expressed genes. Although the authors did observe large changes in the expression ratios of many of the remaining 265 genes, a closer examination revealed that high fold-change values were correlated with high variability and were likely to be false positives. Thus, of the 265 remaining genes, the authors report only four genes altered in the wildtype mice between 5 and 22 months, and only three genes altered between 5 and 13 months. Two of these seven genes, however, were directly involved in insulin signaling, IGFBP1 and IGF receptor 2, notable in particular due to the repeated implication of insulin signaling in longevity studies of lower invertebrates.

Edwards and colleagues compared cardiac gene expression profiles in young (5-month), middle-aged (15-month) and old (25-month) mice at 0, 1, 3, 5, and 7 hours after a single intraperitoneal injection of paraquat (N = 3 for each age and timepoint) (Edwards et al., 2003, 2004). These authors measured expression levels of 9,977 genes using high-density oligonucleotide arrays and found an age-dependent decrease in the cardiac transcriptional response to the paraquat challenge, particularly a decrease in the stress-response pathways signaling through MAPKKK and JNK, and a decreased induction of the DNA damage-induced gene GADD45. The authors also note a shift in the spectrum of oxidative stress genes induced at different ages, with induction of glutathione-S-transferase A3 specific to young mice, glutathione peroxidase 1 and peroxiredoxin 4 specific to middle-aged mice, and superoxide dismutase 1 specific to old mice.

A recent study by Blalock and colleagues (2003) has attempted to link age-dependent expression changes with measurable functional consequences. They describe age-dependent transcriptional profiles associated with cognitive impairment in the rat hippocampus, in particular the hippocampal CA1 region. The authors trained young (4-month), middle-aged (14-month), and old (24-month) male rats on two memory tasks, the Morris spatial water maze and the object memory task, and the hippocampal CA1 region of each animal was subsequently harvested for expression analysis on individual Affymetrix oligonucleotide arrays (one chip per animal, N = 10 animals per group). The expression data were analyzed for both aging effects (ANOVA) and for cognition effects (Pearson’s test), and although most of the gene expression changes were initially evident in the middle-aged group, impaired cognition was not clearly manifest until late life. The aging- and cognition-related genes identified represent familiar categories such as oxidative stress, inflammation, decreased mitochondrial function, and altered protein processing, but also include genes involved in downregulated early response signaling, cholesterol synthesis, lipid
and monoamine metabolism, and other likely brain-specific categories such as activity-regulated synaptogenesis, upregulated myelin turnover, and structural reorganization genes. This study clearly benefits from the statistical power of a relatively large number of samples (\( N = 10 \) for each group), but more importantly, the inclusion of an intermediate timepoint and the correlation with a functional outcome in the study design allow the authors to put forth an integrative model of brain aging in which gene expression changes observed in early adulthood trigger subtle changes resulting in cumulative cognitive deficits not evident until a much later date.

2. Attenuation of Age-Related Expression Changes by Caloric Restriction in Rodents

The early studies by Lee and coworkers comparing genomic expression profiles in young and old brain (Lee et al., 1999; Prolla, 2002) and skeletal muscle (Lee et al., 1999; Weindruch et al., 2001) also examined the effect of CR on the expression profiles of tissues from old (30-month) animals. In general, the authors noted that CR (initiated at 2 months) selectively prevented many of the age-related increases in inflammatory and stress-response genes while having little effect on expression of genes involved in neuronal growth and plasticity in the neocortex and cerebellum. In skeletal muscle, CR shifted the expression profile toward increased energy metabolism, increased biosynthesis, and increased protein turnover, more similar to that of younger animals.

The reversal of age-dependent expression changes by CR seems to occur whether the CR is of short or long duration. The same group (Lee et al., 2002) reported in a separate paper that CR initiated at middle-age (14 months) resulted in a 19 percent global inhibition of the age-related gene expression changes observed in 30-month-old mouse hearts (Lee et al., 2002). Cao and colleagues employed a long-lived F1 hybrid strain of mice, comparing liver gene expression profiles of young (7-month) versus old (27-month) animals fed ad libitum or on a CR diet, and also included a short-term CR cohort consisting of 34-month-old control mice placed on a CR regimen for 4 weeks (Cao et al., 2001). Gene expression changes observed with age were consistent with previous studies, reflecting increased inflammation, stress, and fibrosis with reduced expression of genes involved in apoptosis, xenobiotic metabolism, and DNA replication and cell-cycle. However, the observed changes in gene expression were attenuated by both long-term CR and 4-week short-term CR in old mice. These results imply that CR begun late in life and instituted for a short time can shift gene expression patterns for a subset of genes to a more youthful profile.

However, it may be that the short-term effects of CR on gene expression are tissue and/or age-dependent. Higami and colleagues (2004) investigated the influences of short-term and long-term CR on gene expression in white adipose tissue and report gene expression changes associated only with long-term CR. The authors compared four groups of 10- to 11-month-old male C57Bl6 mice (\( N = 5 \) per group): non-fasted controls, fasted for 18 hours before death, short-term caloric restriction for 23 days, or long-term caloric restriction for 9 months. For this study, the authors employed high-density oligonucleotide arrays with over 11,000 genes. Compared to the control mice, only a few transcripts were differentially expressed in the fasted and short-term CR groups, whereas 345 transcripts were found to be significantly altered by long-term CR, the majority of which were directly involved in metabolism or insulin signaling. The discrepancy between results of the Higami study and
the Cao study likely reflects the different tissues investigated as well as the timing of the short-term CR—in the Cao study, the short-term CR was initiated in much older mice than in the Higami study (34 months versus 9 months).

Indeed, Dhahbi and colleagues (2004) demonstrate that CR begun relatively late in life (at 19 months) begins to increase the mean time to death as well as mean and maximum life spans within 2 months of initiation and is accompanied by a rapid and progressive shift toward a hepatic transcriptional profile associated with long-term CR. (Other studies, however, have failed to demonstrate a beneficial effect of late-onset CR, for example, Lipman et al., 1995, 1998). For the Dhahbi study, the authors used high-density oligonucleotide arrays to compare hepatic gene expression in control mice with that of mice on a CR regimen for 2, 4, or 8 weeks (N = 3 to 4 animals per group). Additionally, a cohort of long-term CR mice was returned to a control diet for 8 weeks. Microarray analysis of livers from 19-month-old control mice switched to CR for 2, 4, or 8 weeks revealed a pattern of early and sustained gene expression changes in response to CR, with early gene expression changes occurring within 2 weeks, intermediate gene expression changes occurring between 4 and 8 weeks, and late gene expression changes occurring after 8 weeks of CR. Furthermore, analysis of the hepatic gene expression profile of mice shifted from long-term CR to a control diet demonstrated that 90 percent of the gene expression effects of long-term CR were reversed within 8 weeks. Thus, the authors’ findings imply a temporal and phenotypic link between CR-induced longevity and genomic expression changes, but these changes may be limited to late onset CR.

The impact of other late-onset dietary interventions on global gene expression has also been evaluated. In particular, the effects on cardiac gene expression profiles in mice given dietary supplementation with alpha-lipoic acid (LA) or coenzyme Q(10) (CQ) started at middle age (14 months) has been examined and compared to CR as a positive control [Lee et al., 2004]. In contrast to CR, supplementation with LA or CQ had no impact on longevity or on the spectrum of observed tumors when compared with control mice on an isocaloric diet. Global analysis of 9977 genes demonstrated that LA, CQ, and CR mitigated age-dependent gene expression changes related to cellular and extracellular structure and protein turnover, but CR was the only intervention to affect gene expression related to energy metabolism. The authors conclude that, although supplementation with alpha-lipoic acid or coenzyme Q(10) induces a gene expression profile indicative of reduced cardiac oxidative stress, CR is much more effective at inhibiting the aging process in the heart, likely due to the observed changes in energy metabolism.

3. Gene Expression Changes in Long-Lived Dwarf Mouse Models

Genetic mutations in the growth hormone signaling pathway and targeted disruption of the growth hormone receptor in the mouse have given rise to various dwarf mice that display remarkable increases in life span. As the primary effector of growth hormone signaling is insulin growth factor 1 [IGF-1], there is considerable interest in whether the longevity observed in these mice results from a mechanism fundamentally similar to that of CR. If true, similar gene expression patterns associated with CR and dwarf models may reveal a transcriptional profile of longevity.

Miller and colleagues (2002) examined the influence of CR on gene expression in the livers of both 9-month-old control and growth-hormone receptor knockout mice [GHR-KO] (n = 8 per group) using 2352-gene cDNA arrays. In this analysis,
CR was reported to significantly alter mRNA levels for 352 genes. Although the GHR-KO genotype had little impact on gene expression in control-fed animals, the gene expression changes associated with CR were significantly diminished in the GHR-KO mice, pointing toward an interaction between the GHR-KO genotype and CR. The genes with altered expression patterns in this study were compared to those identified in a study of another long-lived dwarf mouse model, the Snell dwarf, and expression of 29 genes was found to be similarly altered in both studies. These findings lend strength to the hypothesis that various models of increased longevity share similar underlying mechanisms.

As previously mentioned, Dozmorov and colleagues (2001) conducted an aging study of hepatic gene expression in control and Ames dwarf mice, comparing 5-, 13-, and 22-month-old animals using the 588-gene Atlas cDNA arrays. Although very few genes (7) survived the statistical significance testing to demonstrate an age-dependent change, of these, none were found to be attenuated by the Ames dwarf genotype between 5 and 13 months or between 13 and 22 months of age. A separate study by the same group investigated hepatic gene expression in 6-month-old Snell dwarf mice using the Atlas 2352-gene cDNA array set (Dozmorov et al., 2002). From this analysis, the authors report several gene expression changes associated with the dwarf genotype; however, it remains to be seen which of these early-age (6-month) changes are functionally associated with longevity.

In summary, microarray analyses of gene expression changes in aging rodent tissues have revealed a trend of increased expression of genes involved in inflammation and stress response with age with a corresponding decrease in expression of genes involved in protein turnover and structural maintenance. Several studies suggest that these observed gene expression changes are modulated to some extent by interventions that retard aging, such as CR and/or dwarf phenotype, and thus may provide insight into the molecular mechanisms associated with aging. Furthermore, that the gene expression changes are similar in nature to those observed in nonmammalian models potentially reflects a universal aging process at the transcriptional level, and the availability of long-lived rodent models will prove essential for elucidating these mechanisms in complex mammalian systems.

D. Primates

1. Primate Literature—In Vivo

To date there are few microarray studies reporting gene expression changes with age in primates, particularly in vivo studies. This undoubtedly is due to the relatively recent advent of this technology coupled with the time and logistical requirements in obtaining primate or human tissue samples for analysis, especially for longitudinal aging studies. There are, however, several studies in primates of note, with more recent studies reflecting an increased sophistication in the use of microarray profiling to derive and test hypotheses in the study of aging.

An early attempt to detect age-related gene expression changes in human samples made use of a previously existing database of gene expression profiles from colon adenocarcinoma and normal colon samples (Kirschner et al., 2002). Although the original data were not part of an aging study, the samples (n = 16) were derived from donors ranging in age from 35 to 85 years. The samples had been hybridized onto Affymetrix high-density oligonucleotide arrays, with approximately 6,800 genes represented, and initial data assessment was
conducted using Affymetrix software before loading into the public database. To determine the effect of age—if any—on gene expression, the authors calculated a correlation coefficient between mRNA expression and age of donor for each gene, the significance of which was computed using a t-test. In this manner, only nine genes were found to be significantly altered with age in the normal tissue samples, of which one-third were likely to be false positives as determined by random permutation. In the tumor samples, 12 genes were found to be altered with age, again with one-third likely false, and the overlap between the normal and tumor samples was only three genes.

Another early study measured gene expression in young (age 13 and 14, n = 2) versus old (age 62 to 74, n = 3) human retinas (Yoshida et al., 2002). The RNA extracted from these samples were hybridized to glass cDNA microarray slides with 2,400 genes, 80 percent of which were of neuronal origin. Most genes were unchanged with age; only a small number of (24) genes displayed differential expression, with these representing energy metabolism, stress response, cell growth, and neuronal transmission/signaling. Although this study is limited by a very small and unequal sample size as well as by the small number of genes on the slide, it does represent one of the earliest microarray studies to investigate aging in human tissues.

The National Primate Research Center at the University of Wisconsin, Madison, has maintained a colony of rhesus monkeys (Macaca mulatta) for conducting longitudinal studies of aging in primates, and a cohort of these monkeys are part of a long-term CR study in primates. Kayo and colleagues (2001) reported an expression analysis of 7,070 genes from the vastus lateralis muscle of these rhesus monkeys, analyzing the data by pairwise comparisons between young and old, AL and CR. In a comparison of young (7- to 11-year-old) and old (25- to 27-year-old) monkeys (N = 3 each group), aging resulted in a selective induction of transcripts involved in inflammation, stress response and neuronal factors, with a downregulation of genes involved in mitochondrial electron transport and oxidative phosphorylation. The CR component of the study examined gene expression in middle-aged monkeys (age 19 to 21 years, N = 3 each), fed normally or on a CR regimen for 9 years at the time of biopsy. CR induced an upregulation of genes largely representing structural components and growth regulation, and the authors also observed a downregulation of genes involved in mitochondrial bioenergetics. Interestingly, the authors found little or no evidence for an inhibitory effect of adult-onset CR on the age-dependent changes in gene expression; of the 34 genes upregulated or downregulated with age in the middle-age and old monkeys, only three were found to be significantly altered by CR. Thus, it may be that, in primates, the benefit of late-onset CR is limited. The full impact of adult-onset CR on the life span of these animals is as yet unknown.

A recent study focusing on gene expression in Alzheimer’s diseased brains is nevertheless noteworthy to the aging field, not only because of the increased incidence of Alzheimer’s disease (AD) with age, but also due to findings of a correlation of genomic profiles with disease severity, particularly with early or incipient AD (Blalock et al., 2004). The authors correlated hippocampal gene expression with severity of AD based on antemortem MiniMental Stage Exam (MMSE) score and post-mortem measurements of neurofibrillary tangles (NFTs) and Braak stage scoring. Subjects were assigned to four groups: control (N = 9), incipient AD (N = 7), moderate AD (N = 8), and severe AD (N = 7). Extracted RNA samples were analyzed using high-density
oligonucleotide arrays (>14,000 genes), and after initial data analysis identifying AD-related genes by Pearson's correlation with MMSE and NFT, the authors employed expression analysis systematic explorer (EASE, a modified Fisher's test) to test statistically for co-regulation of genes in a common pathway or process. In this manner, the authors identified 3,413 genes as AD-related genes across all 31 subjects, with genes correlating more strongly with MMSE than with NFT. The authors then examined expression of these genes in only the control and incipient AD groups to identify early markers of AD disease, or incipient AD-related genes, and these genes were analyzed with EASE to determine over- or under-represented categories. Using this approach, the authors report that early or incipient AD is characterized by transcriptional reprogramming and cell growth in the hippocampus, with an unexpected upregulation of tumor-suppressor genes, and also by a downregulation of bioenergetic pathways. Although these findings, possible only with a global analysis of gene expression, lead to a better understanding of the early etiology of AD, they also reflect the powerful utility of microarrays in gaining information on fundamental changes associated with any age-related or progressive disease, which in turn directs research efforts toward better-targeted therapeutics and interventions.

Lastly, a study illustrating the power of microarray profiling to generate testable hypotheses is reported by Lu and colleagues in a survey of human brain gene expression profiles from individuals ranging in age from 26 to 106 years [Lu et al., 2004]. The authors analyzed RNA transcripts of post-mortem prefrontal cortex samples (N = 30) using high-density oligonucleotide arrays (~11,000 genes), and age-related genes were determined by Spearman rank correlation. Hierarchical clustering of these age-related genes and pairwise comparisons of all the samples (by Pearson correlation) reveal clusters of relatively homogenous expression in individuals under 42 and over 73 years of age, with genes upregulated in young (<42) being downregulated in old (>73) and vice versa. The data also clearly reflect considerable heterogeneity of gene expression in the middle years. The greatest number of genes upregulated with age were stress-response and DNA repair genes, whereas downregulated genes included synaptic transmission and vesicular transport. This coordinated downregulation of a defined cluster of genes and subsequent induction of antioxidant and repair genes led the authors to hypothesize that the observed downregulation results from oxidative damage to the promoters of certain genes that are prone to such damage. To test this hypothesis, the authors developed a real-time PCR assay measuring DNA damage in specific DNA sequences based on the resistance of DNA cleaved at apurinic sites to amplification, thereby allowing quantitation of DNA damage from a ratio of PCR products in cleaved versus uncleaved DNA templates. The authors assayed the promoters of 30 selected genes in individual brain samples and demonstrated an age-dependent increase in damage to the promoters of downregulated genes such as mitochondrial ATP synthase alpha, calmodulin 1, sortilin, and calbindins 1 and 2. As proof of principal, the authors demonstrated that oxidative damage (via FeCl₃ and H₂O₂ treatment) reduced expression of the tau gene in cultured neuronal cells, and that this damage and subsequent reduction in mRNA expression was prevented by concomitant overexpression of a human base-repair excision enzyme. Furthermore, the genes downregulated with age in the human brain samples displayed an increased vulnerability to oxidative DNA damage in cultured cells when compared to genes which are stable or upregulated with age, as determined by quantification of
promoter damage and luciferase reporter assays. Taken together, these data support the authors’ model that DNA damage in the promoters of certain susceptible genes precipitate their downregulation with age, with the hypothesis formulated by clever interpretation by the authors of the global expression profiles generated with microarray technology.

2. Primate Literature—In Vitro

Cell culture systems offer several advantages in studies of the cellular physiology aging process, while also being subject to the well-known controversy as to the relatedness of in vitro to in vivo senescence. Although cellular senescence, which is highly correlated with telomere shortening, has not been directly linked to mammalian aging, it has been argued that cellular senescence may underlie the functional decrements present in aging tissues (Bird et al., 2003). Gene expression in cellular senescence has primarily been addressed using cDNA arrays. The cell culture system has the great advantage that gene expression changes can be assayed on a relatively homogeneous cellular population. Therefore, it is notable that a high degree of heterogeneity has been identified in the senescence-associated gene expression patterns between cell types even assayed on the same array platform (Bortoli et al., 2003; Shelton et al., 1999). Differences between platforms further complicate the comparison of results across experimental systems. Additionally, there has been little statistical analysis and almost no validation of candidate genes with a quantitative independent technique to determine exact genes involved in senescence. However, despite these problems, some patterns are emerging. Genes associated with the extracellular matrix, cell–cell signaling are reported as categories of changed genes (Minagawa et al., 2004; Shelton et al., 1999), and these are remarkably close to the gene expression changes observed in organismal aging. However, these associations have all been established without quantitative methods for identifying relative appearance of functional categories as compared to the gene expression platform used. As methods for categorizing genes and identifying over- or under-represented categories are refined, we will be able to determine whether these patterns hold true across heterogeneous cultured cell populations.

In addition to seeking candidate genes associated with cellular senescence that may be biomarkers of aging, chromosomal position of gene expression changes has also been investigated. It has been hypothesized that the genes associated with replicative senescence may localize to telomere-proximal regions. However, a lack of preferential expression or repression of telomere-proximal genes has been reported (Allen et al., 2004; Chen et al., 2004; Minagawa et al., 2004). This result suggests that in vitro senescence is not due to dysregulation of gene expression genes caused by proximity to shortened telomeres.

IV. Conclusions, Future Directions, and Challenges

The technologies and tools to support the use of microarrays for global analysis of gene expression have perhaps matured from infancy, but clearly remain in rapid development. Especially in applications to yeast and invertebrate models, we have begun to see the promise of this approach for elucidating expression profiles associated with aging. More importantly, studies such as those of McCarroll and colleagues (2004) have demonstrated that inter-species comparisons of gene expression can discover common (“public”) transcriptional changes in aging, if only at the level of functional classification. Similarly, inter-comparisons between
different models of life-span extension, such as caloric restriction and the growth hormone/IGF-1 axis, are likely to elucidate the differences and commonalities between mechanisms of longevity in a more comprehensive fashion than might otherwise be possible. Gene expression profiles can define sets of genes that serve as biomarkers of aging and predictors of longevity, both of which are greatly needed for characterization of aging phenotypes. As published databases of expression results increase in number and quality, the power of meta-analyses increases, so that discovery of complex transcriptional relationships, or coexpression networks, becomes ever more feasible (Stuart et al., 2003). Discovery and quantitation of these networks will be very important to increase the sophistication and reduce the complexity of gene expression data analysis, as smaller numbers of networks can substitute for long lists of individual genes. Quantitation of transcriptional pathways also brings transcriptomics to an interface with metabolomics, improving the understanding of changes in the programming of cell function. The limitations of transcriptomics must always be kept in sight, particularly that messenger expression may not reflect protein level and post-translational modification of proteins. For a more complete understanding, proteomics must be studied. Ultimately, using a full suite of tools for global analysis of “systems biology,” we can look forward to a deeper and more integrated understanding of the molecular biology of aging.

References


in blood and mortality in people aged 60 years or older. *Lancet*, 361, 393–395.


Storey, J. D., & Tibshirani, R. (2003b). SAM thresholding and false discovery rates for


I. Introduction

A. The Why, What, and How of Biological Modeling

Three intersecting processes are making the application of mathematical and computer modeling increasingly important in the biological sciences. First, biology itself has become much more of an informational science, as a result primarily of the development of genomic (based on advances in gene sequence and expression data) and post-genomic (based on advances in proteomic and functional data) sciences. Our capacity to answer questions ranging from cell and molecular function through to evolutionary genetics requires an increasing ability to acquire, store, and manipulate large volumes of raw data. This requirement has called upon biologists to develop the necessary computational skills and understanding.

Second, there is a realization that complex biological processes cannot be understood through the application of ever-more reductionist experimental programs alone. There needs to be some integration of the mass of data and insight from study of the detailed mechanisms at the level of the physiological “system.”

Third, the sophistication and power of desktop computer hardware has increased to a point where the kind of model that two decades ago might have required an overnight run on a large mainframe computer can now be done in the individual scientist’s lab or office with a response time that makes possible a much more interactive way of working.

Alongside these changes is the development of a different perception of the role and value of computer modeling in biomedical research. To many scientists who have trained and worked in environments where modeling has not been a part of the scientific toolkit, the nature and scope of computer modeling is still unclear. Many see models as essentially descriptive, begging the question “Why bother?” when the real answer will be revealed in time by experiment. Others have been indoctrinated with the widespread—but largely
false—idea that as soon as a model has more than two or three parameters, it can “explain” anything, resulting in a suspension of belief that models, particularly of complex systems, can be of any real use at all. Fortunately, the increasing dialog between modelers and experimentalists is beginning to break down these barriers of misunderstanding and giving rise to new interactions that are likely to change the way a great deal of science will be done in the coming decades. This new approach is commonly being described as systems biology. This chapter reviews how computer modeling is developing within the context of the biology of aging.

The distinctive advantages of modeling a biological process with the rigor that is needed to build a computer model are as follows:

1. Model building requires that verbal hypotheses be made specific and conceptually rigorous. Before a mathematical model can be formulated, the investigator must specify each element of the model and how it interacts with other elements.

2. Starting to build a computer model may help to highlight gaps in current knowledge. The process of specifying a mathematical model will highlight any important unknowns. Sometimes these can be represented as variables yet to be estimated or determined.

3. The process of model development might lead to the recognition of a gap that needs to be filled by further experimental investigation, which may be fundamental to understanding a complex system. Thus, modeling can be useful even if the gap means that a model cannot yet be completed.

4. Computer models yield quantitative as well as qualitative predictions. A hypothesis can be tested much more rigorously by a model that permits quantitative predictions to be made. In aging, where multiple mechanisms might be at work, it often happens that data are broadly consistent with a hypothesized mechanism, but modeling can show that the magnitude of the effect is too small to explain aging on its own.

5. Modeling can result in improved experimental design, especially where the system embodies the potential for complex interactions. Complexity is very hard to deal with experimentally but is relatively straightforward in a computer model. Models are thus ideal for analyzing complex interactions prior to experimental tests. In extreme cases, modeling may actually reveal that because of interactions within complex systems, a proposed experiment would be inconclusive.

6. Modeling can provide a low-cost, rapid test bed for candidate interventions, thereby enabling a more predictive approach and effecting significant savings in time and money.

To the non-modeler, the science of biological modeling can easily appear to involve the application of the same set of skills to a very diverse range of problems, and in much the same kind of way. A facility with numbers, knowledge of computer programming, and some understanding of the biological system seem all that is required. In reality, the range of approaches and skills in computer modeling is broad, involving a significant diversity of skills and research subdisciplines. Later sections of this chapter examine the different kinds of computer modeling, how they are performed, and what they can achieve. If the integration between theoretical and experimental science is to take place as fast and as effectively as is needed, researchers from both communities will need to learn more about each other’s methods of working. Curiously, there is more similarity between the methods of working of experimenters and modelers than is usually recognized. The experimenter has to decide which factors to include and vary in the study in order to address the
hypothesis most efficiently and directly; (2) spend a large amount of the total effort of the study on controls in order to reduce the possibility of artifact; (3) be careful in framing the conclusions from the study so as not to extrapolate beyond what the results support; and (4) be conscious of planning the study within a constrained budget of time, money, and human resources. So does the modeler. A number of texts expand on some of the issues raised here [e.g., Hilborn & Mangel, 1997].

In parts of this chapter, we describe the mathematical, statistical, and computational approaches that have been brought to bear on understanding the aging process. Because these approaches will be unfamiliar to some readers, we have explained the terms and basic concepts as clearly as possible. It is not feasible, however, to include all of the explanation that would be necessary to equip the reader new to these approaches with a complete knowledge base. We have therefore had to find a balance between explanation and concision. At all relevant points within the text, we give references to texts where the reader can find more detailed explanation.

B. How Computer Models Have Been Instrumental in Solving Biological Problems

In many biological domains, it is difficult to see how a clear understanding of key processes could be gained without mathematical modeling. This includes, for example, the study of population dynamics in ecology, disease transmission in epidemiology, and population genetics and life history theory in evolutionary biology. In other domains, single examples exemplify the insights that mathematical investigation can provide, such as in the study of cardiac fibrillation in physiology. Modeling is making an increasingly important contribution in the relatively new field of systems biology, which aims, in part, to bridge molecular biology and physiology by capitalizing on the large amount of post-genomic data currently being generated. Many biochemical networks involve nonlinear components, which means that relying on intuition is not reliable. In a recent essay, Lander (2004) provides an excellent example of how modeling has helped reveal the details of a molecular mechanism that was proving difficult to understand from a purely experimental approach: the role of the segmentation polarity genes in maintaining the segmentation pattern during Drosophila development [von Dassow et al., 2000]. Although details differ, a similar process is shared by all insects, and to a lesser degree in vertebrates.

Drosophila embryonic development can be described as a three-stage process. In the first stage, maternally expressed mRNA enters the Drosophila oocyte within the ovary, and following translation, a polarized protein gradient is established. In the second stage, gap and coordinate genes are expressed in response to the protein gradient, which in turn govern the periodic expression of pair-rule genes. The final stage involves the segment polarity genes. A repetitive pattern of en (engrailed) and wg (wingless) expression is established based on the pair-rule genes. The embryo then undergoes cellularization, and the pattern of en and wg expression is transferred to an intracellular context involving signals between morphologically distinct bands. The gap and pair-rule gene products fade, en and wg expression is maintained via a complex network of transcription factors and intracellular signals, and the segmented structure is retained during substantial morphological change.

Von Dassow and colleagues (2000) constructed a mathematical model based on current information about the segment
polarity gene network to test whether it could maintain a stable segmented structure. The structure was represented as a series of connected cells, each cell populated with the focal mRNAs, proteins, and protein complexes. The known intracellular and intercellular relationship between interacting molecules was then represented as a series of differential equations, with simulation used to provide information on the temporal variation in concentration of all the molecules. Although a substantial body of information existed to construct the network, it was insufficient to fully quantify the model—50 parameters were unknown. The approach taken was to run simulations with many randomly chosen parameter sets, with each parameter bounded within realistic limits. Interestingly, no solution could be found that satisfactorily reproduced the observed stable segmented pattern.

Attention was then turned to the network structure, and it was realized that modifications were needed to capture the known biology, namely the asymmetry in signaling to neighboring cells anterior and posterior. With appropriate modifications in place, the differential equations were updated and the process of simulation with randomly chosen sets of parameters was repeated. The observed segmented structure was now reproduced with surprising ease. The important result was that it was the fine detail of the network structure that was key in determining the system behavior and not exact parameter values. Interestingly, this may reflect an evolutionary adaptation as minor variations, such as mutations in components of the network or environmental fluctuations affecting levels of signals, would not unduly affect the segment structure and subsequent development. The major message was that the segment polarity genes represent a “robust developmental module” that ensures the formation of an appropriate pattern even across distantly related insect species in which earlier stages of development differ.

II. Why Aging Particularly Needs Models

Recent years have seen rapid progress in the science of aging. A key factor in this progress has been the interaction between evolutionary (why?) and mechanistic (how?) lines of research, which gives shape to the likely genetic basis of aging and to the mechanisms that may be involved (Kirkwood & Austad, 2000). This has helped overcome a situation where the field was dominated by a plethora of rival theories, with little effective dialog between them. In particular, the disposable soma theory (Kirkwood, 1977; Kirkwood & Austad, 2000) suggests that aging is caused ultimately by evolved limitations in organisms’ investments in somatic maintenance and repair rather than by active gene programming. This predicts that aging is due to the gradual accumulation of unrepaired random molecular faults, leading to an increasing fraction of damaged cells and eventually to functional impairment of older tissues and organs. Genetic effects on the rate of aging are, in this view, mediated primarily through genes that influence somatic maintenance and repair.

Although the idea of aging as a buildup of damage is straightforward in principle and supported by a growing range of data, it presents a number of distinctive challenges (Kirkwood et al., 2003). First, it predicts that there are multiple mechanisms that cause aging, instead of just one or a few. Second, it predicts that aging is inherently stochastic—that is, it is modulated to an important degree by chance. Extensive evidence points to an important contribution in aging that arises from chance variations, which are not explained by genetic or environmental factors [Finch &
Kirkwood, 2000). A particularly clear example of the role of chance in aging is the threefold range in life span (see Figure 12.1) and the apparently stochastic age-related cell degeneration of individual worms in isogenic populations of *Caenorhabditis elegans* reared under uniform laboratory conditions (Herndon et al., 2002; Kirkwood & Finch, 2002). Third, since multiple mechanisms contribute to aging, a high level of complexity is to be expected. For all of these reasons, there is exceptional need in aging research for the use of computer models to help integrate findings from different lines of experimental work.

Although the multiplicity of aging mechanisms is now widely acknowledged, the reductionist nature of experimental techniques means that, in practice, most research is still narrowly focused on single mechanisms. This is where computer modeling can make a major contribution. By allowing for interaction and synergism between different processes, models reveal that the predicted effects on the system are often much greater than when mechanisms are considered one at a time. Furthermore, models can highlight important differences between the upstream mechanisms that set a process in train and the end-stage mechanisms that dominate the cellular phenotype at the end of its life. For example, a gradual accumulation of mitochondrial (mt)DNA mutations, occurring over years, might lead to a steady increase in the production of reactive oxygen species (ROS) and a gradual decline in energy production [Kowald & Kirkwood, 1996]. However, although the buildup of mtDNA mutations initiates the process, what ultimately destroys the cell is that eventually a threshold is reached where homeostatic mechanisms collapse. The end-stage of the cell's life span is dominated by dramatic biochemical changes, such as an accumulation of damaged protein. Experimental study of the latter effect, or even of the former cause, in the absence of a quantitative model to link the two would find it hard to establish the connection.

Another benefit of integrative model building is that it is well suited to take account of the fact that many of the key reactions involved in normal cell maintenance and metabolism do not act in isolation—rather, they belong to a network of activity. When the activity of one enzyme changes, all connected metabolite pools and enzyme activities may be altered. In some cases, there may be redundancy in pathways, which provides buffering against damage, whereas in other cases, the effect of damage may be propagated.

Another important area for modeling is to understand the actions of genes that affect the rate of aging. Over the past decade, scores of genes have been identified that affect aging in yeast, nematodes, fruit flies, and mice, and there is growing interest in genes affecting human longevity (Gems & Partridge, 2001; Jazwinski, 2000; Larsen, 2001; Lithgow, 1998; Tan *et al*., 2004). Experimental data are beginning to reveal the interactions of these genes within pathways.
that control the aging rate, and there is evidence that several of the most important genes are those that affect basic cellular processes, such as insulin and insulin-like growth factor (IGF) signaling, which are strongly conserved across the species range (Gems & Partridge, 2001; Rincon et al., 2004). Nevertheless, we are a long way from understanding the interactions between these effects. These studies need also to take account of the intrinsic stochastic nature of gene regulatory networks.

III. Different Approaches to Modeling Biological Systems

A. Descriptive Versus Predictive

A descriptive model describes a process or behavior that has already been observed. A predictive model predicts the behavior of a system not previously observed. A valid descriptive model is often easier to develop but it has less value than a model with predictive power. However, descriptive models are useful for highlighting gaps in our current knowledge. A model may start out as being descriptive but can then be used to predict outcomes when parts of the system are perturbed. For example, a descriptive model of a metabolic pathway of proteins under known conditions can be used to predict protein functions under different circumstances or in different species.

A predictive model provides quantifiable as well as qualitative predictions. The value of quantifiable predictions is that a hypothesis can be tested much more rigorously. In aging, where numerous mechanisms might be at work, data are often broadly consistent with a hypothesized mechanism, but modeling can show that the magnitude of the effect is too small to explain aging on its own. Another advantage of predictive models is that they can provide a low-cost, rapid test bed for candidate interventions.

B. Simple Versus Complex

Biological systems are complex and involve the interrelationships of many different “species,” where species can refer to molecules, cells, tissues, or organisms. A model is a description of the system. The “art” in building a good model is to capture the essential details of the biology, without burdening the model with nonessential details. Every model is to some extent a simplification of the biology, but it is valuable in taking an idea that might have been expressed purely verbally and making it more explicit. Nevertheless, the question still remains: what level of complexity should be incorporated in the model?

At the most basic level, a model must be able to capture the desired inputs and outputs of a system. This is where a clear, prior specification of the problem to be addressed is as essential in modeling as it is in experimentation. For instance, if we wish to investigate how an increase of ATP will affect the production of ROS by mitochondria, then obviously these elements must be included into the model. Other elements, such as the dynamics of a cell cycle, are likely to be excluded. However, it is at this point that the modeler needs to exercise caution and to bear in mind the opportunities that are available to include further factors in the model than could easily be added to an experiment. The choice of which reactions should be left out of the model—since the activity of one enzyme may conceivably affect all connected metabolite pools—has no easy or universal answer. Some modelers prefer to start simple and add further detail as required; others prefer to recognize greater complexity from the outset. Either way, a modeler should always develop a model with as much direct biological input as possible.
Another key factor in the modelers’ decision-making process is the time scale of processes involved. Molecular processes often need to be modeled on a time scale of seconds or less; outcomes affecting aging develop in months or years; evolutionary changes occur over generations. Models that seek to integrate across levels present particularly challenging problems that need to be addressed in defining the aims and scope of the project.

When modeling a process as complex as aging, an unfortunate side-effect is that very quickly, the mathematical representation can become exceedingly complex. Although the modeler may be comfortable with each and every detail of the model, the reader may be presented with an indecipherable collection of symbols. Conversely, an oversimplification of the systems may lead to the justified claim that the model does not represent the structure under consideration. Thankfully, part of this problem is being overcome with the introduction of standard methods for describing models used throughout the biological community, such as the Systems Biology Markup Language (SBML, described in more detail later). When a standard has been decided, this enables generic tools to be developed that aid the understanding of models. For instance, an SBML-aware visualization tool should accept any SBML-encoded model and return a graphical representation of it.

A barrier that limits the amount of complexity that can be included in a model is computational power. Put simply, do we have a computer powerful enough to calculate a solution to our model? It is relatively easy to construct a simple model that when simulated could take weeks to finish. With the yearly increase in processor power, models that would have taken weeks of computational time 5 years ago can now be solved in a matter of minutes. Other exciting avenues include the emergence of the GRID—the new generation of hardware/software computer networking that is designed to facilitate the sharing of data and compute resources over a network. A benefit of the GRID is the harnessing of idle computer power. For instance, whereas a model may take weeks on a single 500-MHz processor, if 50 machines, say in a university computer laboratory, which are idle for 12 hours per day were set to the task, a properly formulated model could take hours.

C. Discrete Versus Continuous

When modeling biological processes, it is often helpful to treat time as a discrete quantity divided into a number of intervals. For instance, when dealing with the cell cycle of the budding yeast \textit{Saccharomyces cerevisiae}, it may natural to deal in terms of generations (Gillespie \textit{et al.}, 2004; Sinclair, 2002).

Although some types of system naturally lend themselves to discrete-time modeling, it is important to consider any distortion that may be introduced. In the yeast cell cycle example, a mother cell produces on average 24 daughter cells; however, the time taken to form a daughter cell gradually increases. So if the events being modeled were directly affected by the interbudding interval, the model may be of limited validity if only discrete generations were considered.

D. Deterministic Versus Stochastic

A model can be generally classed as deterministic or stochastic. A deterministic model is one that takes no account of random variation and therefore gives a fixed and precisely reproducible result. It can be solved by numerical analysis or computer simulation. Deterministic models are often mathematically described by sets of differential equations. Deterministic models are appropriate when large numbers of individuals of a
species are involved and the importance of statistical variations in the average behavior of the system is relatively unimportant. However for many biological systems, this assumption may not be valid.

To illustrate the concepts, let us consider perhaps the simplest of molecular reactions, spontaneous degradation. The ordinary differentiation equation for the degradation of a species \( X \) at rate \( k_1 \) is given by

\[
\frac{dx}{dt} = -k_1 x .
\]  

(1)

Because this is a simple equation, it can be solved exactly to give the deterministic solution

\[
X(t) = X(0)e^{-k_1 t}
\]  

(2)

where \( X(0) \) is the initial amount of species \( X \).

A stochastic model should be used when either the number of a particular species is small or when there is reason to expect random events to have an important influence on the behavior of the system. Often, a stochastic model will be more appropriate when we need to take account of species as discrete units rather than as continuous variables, and particularly when the numbers of a particular species may become small. It may also be necessary to take account of events occurring at random times. The essential difference between a stochastic and deterministic model is that in a stochastic model, different outcomes can result from the same initial conditions.

A stochastic model is formulated in terms of probabilities and is constructed by considering the probability that an event occurs during a small time period. Formulating the model in this manner enables us to calculate the probability that the population is of size \( X \) at time \( t \), \( P_X(t) \). Because the model is reasonably simple, the exact stochastic solution can be obtained:

\[
P_X(t) = \left( \frac{X(0)}{X} \right)e^{-k_1 t(1 - e^{-k_1 t})X(0) - X}
\]  

(3)

where \( X \) takes the values between 0 and its initial amount \( X(0) \).

In most biological systems, the number of species involved and the interactions between them mean that for stochastic models, an analytical solution—that is, one that can be obtained by purely algebraic formulae without using a computer—will not be feasible. In these cases, computer simulations of the stochastic kinetics are used. A simulation keeps track of the number and state of each species over time. Therefore, it is necessary to carry out repeated simulations and then look at the distribution of results to get a picture of the central tendency, the dispersion, and outliers. This process is called Monte Carlo simulation.

Figure 12.2 shows a stochastic realization of a spontaneous degradation
reaction and its deterministic counterpart. For a given starting condition, a single stochastic realization may differ considerably from its deterministic counterpart. This is particularly important in models when the number of a particular species becomes low and the species may or may not become extinct. Figure 12.2 shows a clear example of a species reaching a very low concentration but never becoming totally extinct in the deterministic model, whereas in the stochastic model, the species becomes extinct and the time of extinction varies in different realizations. In the modeling of epidemic diseases within a host population, where it may matter greatly whether and when the first or last infective individual dies or recovers, the difference between stochastic and deterministic models can be very marked. Similar considerations can arise in gene regulatory networks with respect to the random association of transcription factor complexes.

Both the deterministic and stochastic methods have their respective advantages and disadvantages. The modeler should determine which method is more suitable to the task at hand (it may sometimes be both) and use that which is appropriate.

E. Software Tools

There are many ways to develop a model, from using traditional programming languages such as C, Fortran, and Java to mathematical packages such as Mathematica, Matlab, and R. Conventional publication of computer models is generally restricted to the presentation of a few key predictions, although it is common to allow the reader to download the computer program, or source. In order to replicate the same predictions as were published, the program must first be downloaded and any necessary algorithm libraries installed. To actually use the model for further investigation generally requires a significant degree of computing skill.

As noted above (in section III B), there has recently been an increasing effort to agree on a standard method for representing mathematical models. One standard that has been widely adopted is the Systems Biology Markup Language (SBML). SBML provides a computer-readable format for representing models of biochemical reaction networks. Although SBML is human-readable, it is intended that it will usually be other software that would both read and write any models. This is analogous with the now widespread use of HTML for Web documents. Although a human can read HTML source documents, these are intended primarily for reading by a browser, such as Internet Explorer, that transforms the HTML code into a more easily read document on screen.

Currently there are over 60 groups using the SBML standard (see Hucka et al., 2003). Some tools, such as CellDesigner, have been created that enable models to be constructed using a drag-and-drop approach. Using this approach, a user creates species that are assigned to graphical nodes. The nodes are then connected up using arrows to denote reactions (see for example Funahashi et al., 2003).

Other tools, such as JigCell, allow the user to construct a model using chemical equations combined with a spreadsheet approach (Allen et al., 2003). Using this approach is perhaps more useful when dealing with large and complex models, whereas the graphical approach is particularly useful when constructing a model for the first time.

SBML is not the only standard that has emerged. For example, the Petri Net Markup Language (PNML) and CellML are similar efforts in creating standards (see Lloyd et al., 2004; Weber & Kindler, 2003). Although each standard focuses on
different aspects of the model, they are not mutually exclusive. Hence, an effort is being made to create tools that allow the transformation from one language to another.

F. Validation of Models

Once we have constructed a model and are satisfied with its behavior, we need to test the model against observations from the biological system that it represents. This process is called validation. First, it is necessary to test the model to see whether it fits the data that we already have. Any discrepancies here need to be addressed. Once the model has been validated in this way, we should then test the model against data that were not used to estimate parameters for our model. If the model predictions and the observed data are not in close agreement, then the modeler needs to study the model to try and find where the discrepancies arise. This could mean modifying the model or adding further detail to the model. This is an important step as it may highlight that the current knowledge of the system is insufficient and that further experimental work should be carried out. Once modifications to the model have been made, the model is tested again, as shown in Figure 12.3.

Another aspect of validation is sensitivity analysis to assess how varying model parameters affect the model outcomes. There are two reasons why this is useful. First, we may be interested in some particular parameters—for example, the rate of degradation of a protein and how this might affect the buildup of damaged protein. Second, some of the parameter rates might not have been accurately determined and so it is important to see how sensitive the model is to small changes in these. The size of changes to make for each parameter depends on how well the parameter was determined initially. In the case of parameters that have been estimated from data, a multiple of the standard error is appropriate. For parameter values that have been guessed, a guess at the percentage reliability is also required. Caution needs to be taken when parameter estimates are correlated because if one parameter estimate is changed, some of the others might have to be changed too.

IV. Currently Available Models of Aging

Existing work reflects the variety of current models in aging research, which range from detailed modeling of individual intracellular mechanisms to higher-level modeling required to address the fundamental problem of why aging should occur. We will not attempt to be wholly comprehensive in this review but will illustrate the breadth of coverage and the different methodologies employed.

A. Intracellular Mechanisms

There are a large number of models currently available that focus on individual intracellular mechanisms. Currently, the
models that relate directly to aging have been mainly concerned with telomere shortening, the accumulation of somatic mutations, and the accumulation of defective mitochondria.

1. Telomere Models

Telomeres are repetitive DNA sequences found at both ends of linear chromosomes. In telomerase-negative cells, telomeres shorten with each cell division, and this process eventually causes cells to enter a state of replicative senescence. One cause of telomere shortening is the end-replication problem caused by the inability of DNA polymerases to replicate a linear DNA molecule to its very end. In the 1990s, models were developed to try to explain replicative senescence in human fibroblasts based solely on the end-replication problem (Arino et al., 1995; Levy et al., 1992). Later models included additional mechanisms of telomere shortening. Rubelj and Vondracek (1999) modeled abrupt telomere shortening due to DNA recombination or nuclease digestion. It has been found that an increase in oxidative stress accelerates the rate of telomere shortening due to an accumulation of single-strand breaks in telomeric DNA (von Zglinicki et al., 1995; von Zglinicki et al., 2000). More recent models have included this additional mechanism and found that the models predict that oxidative stress plays an important role [Proctor & Kirkwood, 2002, 2003]. For example, simulations showed that increasing the level of ROS led to fewer cell divisions on average. Space does not permit detail of all of the current models of telomere shortening, but the interested reader may refer to the references for further models [Aviv et al., 2003; den Buijs et al., 2004; Golubev et al., 2003; Hao & Tan, 2002; Olofsson & Kimmel, 1999; Sidorov et al., 2004; Tan, 1999a,b, 2001].

2. Somatic Mutations

The role of somatic mutations in aging is an area of particularly active research, following new methods for measuring DNA modification and repair. The somatic mutation theory was first proposed several decades ago after experiments showed that irradiation shortened life span in animal models and induced features of premature aging (Henshaw et al., 1947; Lindop & Rotblat, 1961).

Szilard (1959) proposed a mathematical model that assumed that recessive mutational “hits” in diploid organisms would accumulate so that a cell could continue to function until one pair of genes had both received a “hit.” Holliday and Kirkwood (1981) developed a deterministic model of the accumulation of recessive mutations in human fibroblast populations. A stochastic model of the same processes was later developed [Kirkwood & Proctor, 2003], which also considered the possibility that there may be synergistic interactions between mutations.

3. Mitochondria Models

The free-radical theory of aging proposes that ROS, which are constantly generated through normal cell metabolism in the mitochondria, cause aging by damaging membranes, proteins, and DNA (Harman, 1956). The mitochondrial theory of aging proposes that an accumulation of defective mitochondria is a major contributor to the cellular deterioration that underlies the aging process (Harman, 1972). Studies have shown that defective mitochondria accumulate with age to a greater extent in post-mitotic tissues (Cortopassi et al., 1992; Lee et al., 1994), although it has recently been reported that high levels of mitochondrial defects are observed in aged human colon (Taylor et al., 2003). In addition, several studies have shown that muscle
fibers are taken over by a single form of mutant mtDNA (Brierley et al., 1998; Müller-Höcker et al., 1993).

Hypotheses to explain the apparent “clonal expansion” of mutant mtDNA are a replication advantage for the mutant mtDNA; slower degradation of mutant mitochondria (de Grey, 1997); and random intracellular drift. Mathematical models have been developed to explore quantitative predictions from these ideas. Kowald and Kirkwood (2000) developed a deterministic model based on de Grey’s hypothesis. Other models are based on the idea of random intracellular drift (Chinnery & Samuels, 1999; Elson et al., 2001).

4. Chaperone Models
Molecular chaperones have an important role in helping to maintain protein homeostasis within cells. It has been observed that the induction of heat shock proteins, a major class of chaperones, is impaired with age and that there is also a decline in chaperone function. Although there are a few mathematical models on the role of heat shock proteins in the cell, to date only one model has looked at the role of chaperones in the aging process (Proctor et al., 2005). This model describes how heat shock proteins are upregulated after an increase in intracellular stress and can be used to investigate the effect of stress on protein homeostasis.

5. Network Models
A few models exist that show how different mechanisms interact synergistically (Kowald & Kirkwood, 1994, 1996; Sozou & Kirkwood, 2001), examples being the interactions of defective mitochondria, aberrant proteins, free radicals, and scavengers in the aging process (Kowald & Kirkwood, 1996); and the interactions of telomere shortening, oxidative stress, and somatic mutations in nuclear and mitochondrial DNA (Sozou & Kirkwood, 2001). We are currently engaged in a major effort, the Biology of Ageing e-Science Integration and Simulation (BASIS) project, to develop interactive models that can network a variety of individual processes together in a flexible, user-friendly manner (Kirkwood et al., 2003). One of the aims of the BASIS project is to allow models of individual mechanisms to be linked together to form a “Virtual Aging Cell” (Proctor & Kirkwood, 2003).

B. Tissue Models
The functional properties of an aging organ or tissue can become compromised, even if most of the cells are in good working order. Mathematical models are required to help us try to understand how a fraction of damaged cells can lead to altered tissue function. Early models were motivated by the fact that cultured human diploid fibroblasts cannot be grown indefinitely in culture (Hayflick, 1972). These models were based on the commitment theory, the idea that cells become irreversibly committed to senescence while still outwardly healthy (Holliday et al., 1977; Kirkwood & Holliday, 1975).

Recently, extensive experimental data has been generated on intrinsic age changes that affect the function of intestinal stem cells in aging mice (Loeffler et al., 1993; Martin et al., 1998a,b). This has led to a number of mathematical models (e.g., Gerike et al., 1998; Loeffler et al., 1993; Meineke et al., 2001). Another model based on data from muscle-derived stem cells has also been developed (Deasy et al., 2003).

Another tissue system that has been extensively studied and modeled is the population of T cells and their role in immunosenescence (Luciani et al., 2001; Romanyukha & Yashin, 2003).
C. Organism Models

There are only a very limited number of models dealing with whole-organism aging, and these are limited to unicellular organisms. Budding yeast, *Saccharomyces cerevisiae*, is commonly used to study cellular aging. Accumulation of extrachromosomal ribosomal DNA circles (ERCs) appears to be an important contributor to aging in yeast, and a mathematical model has been developed to examine this process (Gillespie *et al.*, 2004). Another interesting model contrasts regulatory with stochastic processes in genetic segregation during division as a mechanism for the aging observed in the asexually reproducing ciliate *Stylonychia* (Duerr *et al.*, 2004).

D. Population Models

An area of research that has contributed to a fundamental understanding of why aging occurs is life-history theory—a theory that essentially deals with schedules of growth, survival, and reproduction maximizing Darwinian fitness (Kirkwood & Austad, 2000; Kirkwood & Rose, 1991; Partridge & Barton, 1993). Many classic life-history papers included an investigation of senescence (Cole, 1954; Fisher, 1930; Hamilton, 1966; Williams, 1957). Another modeling approach is represented by the disposable soma theory (Kirkwood, 1977; Kirkwood & Rose, 1991), founded on the principles of optimality theory (Parker & Maynard Smith, 1990). The disposable soma theory suggests that aging arises as part of an optimal life history due to tradeoffs in resource allocation between investment in reproduction and maintenance affecting long-term survival. There is much data in general support of the existence of such tradeoffs, including in humans (Lycett *et al.*, 2000; Westendorp & Kirkwood, 1998). More complex general life-history models that have incorporated measures of an organism’s state such as size in addition to age have been analyzed using techniques such as dynamic programming, simulated annealing, and Pontryagin’s maximization principle (Abrams & Ludwig, 1995; Blarer & Doebeli, 1996; Cichon, 1997; Clark & Mangel, 2000; Houston & Mcnamara 1999; Schaffer, 1983; Teriokhin, 1998; Vaupel *et al.*, 2004).

The following specific issues have attracted attention using approaches and modeling techniques drawn from the domains of life-history theory, demography, and population genetics.

1. Dietary Restriction

Dietary restriction is observed to cause slowing of aging and extension of life in many species. One hypothesis is that animals have evolved a response to temporary fluctuations in resource availability, in which energy is diverted from reproduction to maintenance functions in periods of food shortage, thereby enhancing survival and retaining reproductive potential for when conditions improve. A detailed quantitative development of this hypothesis using a dynamic resource allocation model revealed that the effect could be the result of the suggested evolutionary process provided that the following conditions were satisfied: (1) there is a substantial initial cost to reproduction, and (2) juveniles are at a disadvantage during periods of food shortage (Shanley & Kirkwood, 2000). An alternative approach is presented using a dynamic energy budget (van Leeuwen *et al.*, 2002). Recently, metabolic control analysis has been used to help identify an increased proton leak in the mitochondrial inner membrane as one possible mechanism whereby ROS production is reduced (Lambert & Merry, 2004).

2. Negligible Senescence

Species that exhibit negligible senescence are of particular interest (Finch, 1998) and
are well represented in species that continue to grow after maturation, so-called indeterminate growth. A dynamic model that explicitly included size as a state variable predicts that this should be the case, and indeed predicts that some species should show negative senescence [Vaupel et al., 2004]. Interestingly, conditions for non-aging can be found in a general life-history model in an analysis of vitality: a term that combines the declining fecundity and increasing mortality characteristic of senescence [Sozou & Seymour, 2004].

3. Gompertz, Mortality Plateaus, and Heterogeneity

Most species exhibit an exponential increase in mortality with age that can be described by the Gompertz model or by the Gompertz-Makeham model that has an extension to include extrinsic sources of mortality [Golubev, 2004]. One problem in the acceptance of this model is the observation that in large laboratory populations, mortality rates appear to plateau at later ages [Vaupel et al., 1998]. A number of models have been proposed to account for this pattern, such as an evolutionary trade-off [Mueller & Rose, 1996; Mueller et al., 2003], a combination of mutation accumulation and pleiotropy [Charlesworth, 2001], a state-based approach [Mangel & Bonsall, 2004], and individual heterogeneity within the population [Pletcher & Curtsinger, 1998].

4. Human Menopause

The rapid reproductive senescence associated with menopause in human females occurs well in advance of general somatic senescence and poses an interesting evolutionary problem. The reproductive life span of human females is limited, as in almost all other mammals, by a finite pool of oocytes established in the developing fetus. Menopause is manifest when this pool is near to exhaustion, and mechanistic models have focused on the follicular dynamics [Faddy et al., 1992] and problems of fetal loss that increase in frequency as the oocyte pool is depleted [O’Connor et al., 1998]. Other modeling has focused on determining whether at some age mothers may increase their fitness by diverting investment from continued reproduction to existing offspring and grand-offspring [Hawkes et al., 1998; Lee, 2003; Peccei, 1995; Rogers, 1993; Shanley & Kirkwood, 2001]. To date the results have not been conclusive, but given the importance of intergenerational assistance—for example, as seen in the two-fold improvement in mortality for infants with a living grandmother [Sear et al., 2002] combined with the particularly high risk of mortality in childbirth for human females—an evolutionary explanation remains a clear possibility and the development and testing of further models appears likely.

V. Models, Data Collection, and Experimental Design

Models are developed based on the collective understanding of the scientific community regarding the underlying mechanisms driving the processes of interest. The related activity of designing experiments to provide data to falsify or refine the models is essentially just a formalization of “the scientific method.” However, there are a number of issues that arise in the context of complex dynamic models that make this procedure far from straightforward in practice. Models are typically concerned with underlying mechanisms that are difficult or impossible to measure directly through conventional experimental procedures. Consequently, they often contain parameters [such as various kinds of rate constants] whose values are not known.
If a model cannot make predictions regarding quantities that can (at least in principle) be measured experimentally, then it is not falsifiable, and in an important sense is not “scientific.” There is therefore a requirement to develop models that are predictive, as this then affords an opportunity to compare the model predictions with experimentally determined reality. As well as providing the opportunity for falsification, this predictive behavior also potentially allows “calibration” of model parameters by finding combinations of parameters that reduce the discrepancy between the model predictions and reality.

These kinds of “inverse” problems have long been recognized in the physical and engineering sciences, and there is a large literature concerned with attempts to solve them. The problem can be understood generally as follows: a complex system has a range of inputs, and based on these, produces a range of outputs. The inverse problem is to find a set of inputs to the system that closely matches a given set of outputs [desired target or experimental observation]. In the context of an attempt to match an experimentally observed history of a physical system, attempts to solve the inverse problem are often referred to as history matching or calibration, of which more is presented in the next section. Such calibration techniques are generally applied post hoc, after the experimental data have been collected. However, in the context of biological modeling, there is often an opportunity to go back to the lab to collect appropriate data for model validation and calibration.

A question then naturally arises as to exactly what data should be collected, and how much. In the context of complex biological models, such experimental design questions are difficult to tackle within a formal statistical framework, but it is relatively straightforward to justify some guiding principles. First, it is necessary to collect data that are (or can be) predicted by the model. This may require model refinement but is necessary; otherwise, there is nothing to link the model and the data that is collected. Second, one must collect data corresponding to model predictions that are sensitive to underlying model assumptions (structural and otherwise). This is necessary on the grounds of falsifiability. Third, it is necessary to gather measurements that help to answer questions of key scientific interest. For example, if a key model parameter of interest is the degradation rate of a particular protein, then measurements should be taken on data that are sensitive to the choice of rate rather than data that are relatively robust to this choice. This will ensure that the data collected provide useful information. Fourth, enough data should be gathered to ensure that an adequate assessment can be made of inter- and intra-experimental variation; otherwise, there is no way to be sure that the data are representative and that the model is not being calibrated to fit an atypical data set. If the model being calibrated is stochastic, more data are probably required, as it is likely to be necessary to reliably determine a good approximation to the full probability distribution of key observables. These basic principles are fairly self-evident, but effectively operationalizing them for a complex dynamic simulation model is not necessarily easy. However, there are many excellent texts on experimental design that can provide further guidance [see for example Clarke & Kempson, 1997; Cochran & Cox, 1992; and Mead, 1988].

VI. Parameter Inference
A. The Calibration Problem
Many approaches have been applied to the calibration problem within the traditional engineering context. First, a
(deterministic) computer simulator (model) of the physical (or biological) system under study is built, so that for a given set of inputs, a corresponding set of outputs may be computed. Next, a simple measure of “distance” for how far the output is from the desired match is defined. The input space is then searched for a set that minimizes the distance measure. This optimization problem can be approached using a variety of techniques for multivariate function minimization—for example, steepest descent, Newton methods, conjugate gradients, simulated annealing, genetic programming, and so on (Koza et al., 2001). Such minimization techniques can work well if the computer simulator is very fast, but for a simulator of a large, complex system, which may take hours or days for a single run, such naive approaches generally fail.

Simple search methods are very wasteful of information. Typically, a very small number of runs are used to decide on a new “best guess” for the target input parameters, and then all existing information is discarded as the search continues from this new input set. In contrast, statistical approaches to the calibration problem attempt to use all available runs from the simulator in order to infer a model for the relationship between the inputs to and outputs from the simulator. In this context, it is not necessarily optimal to always evaluate the simulator at the current “best guess” at the optimal input set, but instead to evaluate at an input set that gives the most information regarding the relationship between the inputs and outputs in the vicinity of the predicted optimal set. Thus, such statistical approaches to calibration need to combine both non-parametric statistical inference techniques and experimental design algorithms in order to effectively solve the problem (Sacks et al., 1989).

A range of different approaches can be taken to carrying out statistical inference for model parameters given data, and these correspond to different schools of statistical thought. Classical frequentist approaches seek to construct estimators that are a function of the data and have desirable properties (such as consistency and lack of bias) under repeated sampling. However, even leaving aside the serious philosophical objections many people have to the repeated sampling framework at the heart of frequentist inference, there are many practical difficulties associated with applying such techniques in the context of complex dynamic models. Consequently, few statisticians would consider a frequentist framework in this scenario.

Approaches based on the likelihood function of the data provide a more powerful and natural way of addressing the simulation model parameter inference problem. These can be divided into two main camps. The first is the maximum likelihood school, which attempts to be “objective” by using only the likelihood function of the data and seeks combinations of parameters, which makes the data as likely as possible conditional on those parameters. The likelihood (or log-likelihood) function is used as a way of “scoring” the goodness of fit, which can then be optimized. Although this sounds straightforward in principle, the likelihood function is typically not analytically tractable for complex models, and this introduces a variety of complications.

The second camp is the Bayesian school, which corrects the conditioning from data on parameters to parameters on data, and thereby seeks parameters that are likely given the data. This is done at the expense of introducing prior distributions into the problem but has a range of benefits as a result. These include the fact that the resulting framework is fully probabilistic, and that probabilistic information regarding all parameters can be obtained from the posterior distribution. The use of priors is also valuable, as they
help regularize the problem and allow the modeler to incorporate information regarding realistic parameter ranges into the inference algorithm. A further benefit of the Bayesian framework is that because it is probabilistic, powerful computational algorithms may be naturally applied to problems where the likelihood is analytically intractable. Markov chain Monte Carlo (MCMC) algorithms (Gamerman, 1997) use stochastic simulation techniques to obtained realizations from the (complex) posterior distribution, which are then used to draw inferences about model parameters. For more information about Bayesian inference, see Bernardo and Smith (2000), O’Hagan and Forster (2004), and references therein.

B. Statistical Approaches to Simulation Model Calibration

Although non-Bayesian approaches to the calibration problem are possible, the complexity and dimensionality of the problem, together with the need to incorporate available expert prior information regarding, inter alia, plausible ranges for rate constants and information on data quality, mean that a Bayesian approach is particularly attractive. Typically, (in the context of deterministic processes), a model is specified in the following form:

\[ z_i = \zeta(x_i) + \varepsilon_i, \quad \zeta(x_i) = \rho \eta(x_i, \theta) + \delta(x_i) \]  

(4)

Here \( z = \{z_1, z_2, \ldots, z_n\} \) represents the available experimental data, obtained from \( n \) different experimental conditions \( x_1, x_2, \ldots, x_n \); \( \zeta(x_i) \) is the real behavior of the biological system under experimental condition \( x_i \); \( \varepsilon_i \) is the measurement error associated with the \( i \)th experiment; \( \rho \) is a bias associated with the computer simulator of the biochemical system; \( \eta(x_i, \theta) \) is the result of running the computer simulator under experimental condition \( x_i \) with the “perfect” set of calibration parameters \( \theta \); and \( \delta(x_i) \) represents model inadequacy that is independent of the calibration issue. In addition to the experimental data, there will be data \( y = \{y_1, y_2, \ldots, y_N\} \) obtained from \( N \) runs of the computer simulator (where \( N \) will typically be larger than \( n \), even in the case of an expensive simulator), where

\[ y_j = \eta(x_j^*, t_j) \]  

(5)

is the result of the \( j \)th computer experiment, \( x_j^* \) is the experimental condition associated with the \( j \)th computer experiment and \( t_j \) is the set of calibration parameters associated with the \( j \)th computer experiment (Kennedy & O’Hagan, 2001).

Note that within this framework, the computer simulator of the biological model is represented by a (deterministic) function \( \eta(\cdot, \cdot) \), which can be evaluated at any combination of experimental conditions and calibration parameters by running the simulator with the specified input. If the simulator were very fast, so that evaluating \( \eta(\cdot, \cdot) \) were cheap, then standard Bayesian inference techniques could be used in order to make direct inferences for \( \theta \) using (1), generating data of the form (2) as and when required. However, due to the expense of evaluating \( \eta(\cdot, \cdot) \) for large complex models, \( \eta(\cdot, \cdot) \) is often regarded as an unknown function, modeled using a Gaussian process. Thus inference may proceed for \( \theta \) using only the \( N \) computer simulator runs available. Bayesian inference is typically carried out using a mixture of analytic direct matrix computations related to Gaussian processes together with computationally intensive techniques, using Markov chain Monte Carlo (MCMC) methods.

Note that in the context of biological modeling, choice of the experimental conditions for the \( n \) “wet lab” experiments will often (though not always) be predetermined. However, the choice of conditions for the \( N \) computer simulator runs will be at least partly under the calibrator’s control.
and will be of key importance to the overall effectiveness of the procedure. This is a nontrivial (sequential) experimental design problem, but limited literature already exists that will provide guidance in this area (Craig et al., 1996; Currin et al., 1991; Kennedy & O’Hagan, 2001; Sacks et al., 1989).

Many complex computer codes have the facility to be run at different levels of sophistication, and hence accuracy. The BASIS simulator, for example, may be run in “exact” mode, where the simulation of the stochastic process used to model a given biochemical system is “perfect,” based on a discrete event simulation strategy similar to the Gillespie algorithm. Such exact simulation procedures are desirable but are typically very expensive to carry out. On the other hand, the system may also be run in an approximate mode, based on a time-discretization of the process, where both the accuracy of the procedure and the time taken for a run depend on the size of time-step adopted. In this case, it can often be optimal to combine a large number of fast (but less accurate) runs with a small number of slow (but accurate) runs in order to make most efficient use of computer time. There is already a sizable literature in this area; see for example Higdon and colleagues (2003), Kennedy and O’Hagan (2001), and references therein.

C. Direct Statistical Parameter Inference

The calibration techniques alluded to above work well in the deterministic context but are not completely straightforward to extend to the case of stochastic simulation models. For a stochastic model of relatively low dimension, it may be possible to make a direct attempt to carry out statistical inference for the parameters of the system given (for example, time course) experimental data on the system dynamics. Here, rather than regarding the simulator as an “unknown function,” the stochastic process corresponding to the simulator is modeled directly, and all aspects of the process that are not observed are “filled-in” probabilistically using appropriate MCMC techniques. Conditional on complete knowledge of the stochastic process, inference for any rate parameters driving the system dynamics is straightforward. The difficulty of such methods is in the construction of the MCMC algorithms to fill in the missing aspects of the stochastic process. This is very problem-specific and generally requires a fairly detailed understanding of the underlying dynamics, including the likelihood function, as well as experience in the use of MCMC algorithms. The use of these techniques for identification of biological models is still in its infancy, but see Boys and colleagues (2004), Gibson & Renshaw (2001), Golightly & Wilkinson (2005), and O’Neill (2002) for some successful examples.

VII. Conclusions

This chapter has examined the rationale for the use of computer models in studying the aging process and has reviewed the range of models that have been developed. It has also described some of the generic issues that need to be addressed in terms of the methodology of modeling. In coming years, it is likely to be essential, if aging research is to realize its potential, that modeling studies are greatly extended and that models are increasingly used to link together the pieces of the picture that are revealed by reductionist experimental techniques. These developments are an inherent part of the “new” ways of doing science that are commonly described as “systems biology.” Whether systems biology is really new or not is a matter for debate, and a spectrum of opinion can be found. What is unquestionably new is the mass of detailed information emerging at accelerating pace from functional
genomic technologies, the rapid expansion of raw computing power, the developing connectivity offered by advances in Internet (soon to be GRID-enabled) Web services, and the recruitment of increasing numbers of mathematicians, statisticians, and computer scientists into the life sciences.

Not all areas of biology need to be taken over by the systems approach, but few are likely to remain untouched by it. The biology of aging is one area where it is hard to envisage the necessary progress being made without embracing the systems approach. There are just too many mechanisms, levels of action, and experimental models for it to be realistic to anticipate integration without the use of computer models. Effecting the building of the cross-disciplinary research programs to bring this about is going to be a challenge, but it should also prove to be intellectually stimulating and fun.

References


that accounts for the effects of caloric restriction on body weight and longevity. *Biogerontology, 3*, 373–381.


