DEVELOPING COMPUTATIONAL MODELS TO UNDERSTAND AGING

by

Spencer Farrell

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
December 2021

Table of Contents

List of Tables

List of Figures

Abstract

Aging in biological organisms is a complex process, involving changes at all levels of functioning. No single pathway or mechanism is responsible for aging, leading to the current understanding that aging is due to a number of interacting biological factors. To understand this interconnected complex process, this thesis develops complex computational models of aging. Using human data I develop network models of aging, which model the aging process as a network of interacting components. These models are used to understand the network structure of different aspects of health, as well as make quantitative predictions of aging health outcomes and mortality. Using worm data I develop an aging trajectory clustering model, which describes the dynamics of worm aging with a low-dimensional latent space that exhibits simple dynamics and clear clustering. This model is used to infer distinct worm aging phenotypes. Using mice and human data, I develop a method to extract damage and repair processes in aging. This approach is used to study the effects of age and interventions on the processes of damage and repair.

This work is an attempt to build computational models of aging, and demonstrates the potential of these types models in the study of aging in the future.

List of Abbreviations and Symbols Used

$\mathbb{E}_q[f]$	Expectation over the distribution p, $\mathbb{E}_p[f] = \int f(z)p(z)dz$
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\odot$	Element-wise multiplication of vectors, $\mathbf{x} \odot \mathbf{y} = [x_1 y_1, ..., x_N y_N]$
C. elegans	Caenorhabditis elegans, a small worm used as a model organism
ACE inhibitor	Angiotensin-converting enzyme inhibitor
ADL	Activities of Daily Living
AUC	Area Under ROC Curve
BA	Biological Age
C-index	Concordance index, the proportion of pairs of individuals where the model correctly predicts which dies first.
CSHA	Canadian Study of Health and Aging
DJIN model	Dynamics Joint Interpretable Network model
ELBO	Evidence Lower Bound
ELSA	English Longitudinal Study of Aging
ER	Erdos-Renyi Network
FI	Frailty Index
FI-clin	Clinical Frailty Index
FI-lab	Laboratory Frailty Index

GEM	Griffiths, Engen, and McCloskey stick breaking process
GNM	Generic Network Model
GRU	Gated Recurrent Unit
IADL	Instrumental Activities of Daily Living
IBS	Integrated Brier Score
MCMC	Markov-Chain Monte-Carlo
MFT	Mean Field Theory
MICE	Multiple Imputation by Chained Equations
NHANES	National Health and Nutrition Examination Survey
RMSE	Root Mean Square Error
RNN	Recurrent neural network
ROC curve	Receiver Operating Characteristic curve
SDE	Stochastic Differential Equation
SSA	
t-SNE	t-distributed stochastic neighbour embedding, a dimensionality reduction technique
VAE	Variational Auto-Encoder
WNM	Weighted Network Model
WS	Watts-Strogatz network

Chapter 1

Introduction

1.1 Motivation

Aging is the age-dependent functional decline of an organism, and is the leading risk factor for almost all severe chronic diseases and disabilities [1]. Some of the most common of these diseases are cardiovascular disease, cancer, and dementia. Even if it was possible to fully treat or prevent some of these diseases, as individuals continue to age it is overwhelmingly likely that they will eventually acquire another. This suggests that instead of a "whack-a-mole" type of approach to treating these diseases one-by-one, tackling the root cause through aging must be the ultimate long-term strategy for achieving long and healthy lives [2, 3]. To achieve this goal, we must first understand aging.

Aging is a high-dimensional complex stochastic process, involving changes at all physical scales of organism functioning. No single pathway or mechanism is responsible for aging, leading to the current understanding that aging is due to a number of interacting biological factors involving damage and repair mechanisms at the fundamental level [4, 5]. Computational models are essential to make predictions or understand mechanisms within complex non-linear, stochastic, and interconnected systems — such as aging [6–8]. This thesis focuses on the development of complex computational models of aging.

Few complex models of aging have been developed to date, despite the understanding of aging as a complex system. There have been descriptive models developed to theoretically model and understand mortality [9–13], but these often do not include health, and do not make quantitative predictions. Additionally, many current approaches to analyzing multi-dimensional aging data only summarize the health state into one-dimension, such as the Frailty Index [14, 15], Frailty Phenotype [16], Biological Age [17, 18], Physiological Dysregulation [19, 20], and recent machine learning

approaches [21]. These approaches are useful to develop health measures and quantify the progression of aging, but they are not dynamical models that can simulate or predict the future health trajectories of individuals, nor do they capture the full multi-dimensionality of aging health outcomes.

This work represents an attempt to build complex computational models of aging that can be used to confront, describe, and learn from aging data. This thesis involves both descriptive models to describe and understand aging phenomena, and develops predictive machine learning models to make quantitative predictions and understand mechanisms of aging. One key challenge to developing machine learning models for application in aging, and science more generally, is interpretability [22, 23]. Interpretability is concerned with understanding *why* a model makes certain predictions, which is key to a scientific understanding of aging.

In this thesis, the philosophy of "Scientific Machine Learning" or "Theory guided data science" [24–26] is taken. In this approach, models use machine learning techniques to learn unknown components of a complex model, while incorporating ideas and concepts from the theory of aging to constrain the models. This incorporates interpretable components into flexible machine learning models.

To paraphrase Box [27], all models are at least partially wrong but some can nevertheless be useful. We cannot hope to build comprehensive models of aging, but we can build models to aid in our understanding of aging. The goal of this thesis is to develop models that can contribute towards answering three fundamental questions in aging: how can we better understand the mechanisms or causes underlying observed aging phenomena, how can we better predict outcomes at an individual or population level, and, finally, how can we better intervene to decrease mortality and to improve health during aging?

1.2 Dynamical Models of aging

In the review and perspective paper "The potential for complex computational models of aging" [23], we introduced the concept of *Dynamical* models of aging, where the dynamics of the health state is modelled as an individual ages until death. These are the type of models developed in this thesis. Here I give an overview of the current literature on dynamical models of aging.

1.2.1 Network and redundancy models of mortality

One of the first computational models of aging developed was the "Reliability Theory of Aging" developed by Gavrilov and Gavrilova [9]. The goal of this model was to explain the exponential increase in mortality rate with age, known as Gompertz' law of mortality [28, 29].

This model attempts to explain Gompertz' law in terms of the redundancy in physiological systems. This model describes aging with a system of "cells" containing many redundant components. Each component has a constant rate of failure, and a cell fails when all of its components fail. When all cells in the system fail, mortality occurs. Simulations of the model result in mortality rates vs age that are nearly exponential [9], similar to Gompertz' law.

The work of Vural *et al.* [30] builds off of this work, but instead considers a *network* of redundant components that can fail (damage) or repair. Where the model of Gavrilov and Gavrilova considered independent components that can fail, this network model introduces interactions. In this model, components can randomly damage, or damage when the majority of their connected neighbours damage. This model is used to fit the mortality rate vs age curves of various model organisms, including C. elegans (worms), Drosophilia (fruit flies), medflies, beetles, mice, and tahr. This model has been used for the theoretical study of repair in aging [31] and used to theoretically explain effect of aging interventions on C. elegans [32].

Figure 1.1a) shows the basic structure of redundancy and network models of aging. Components of the network fail or damage (red), leading to the accumulation of damage, until death. Nodes can interact as indicated by the connections in the network, propagating damage throughout the network. Although health can be measured with the number of damaged components in these models, these models instead focused on mortality. We consider a similar style of network model in Chapters 3 and 4, but considering both health and mortality.

The Stochastic Process Model (SPM) of Yashin *et al.* [33–36] is a different type of network model of aging. This model describes aging in terms of continuous stochastic dynamics of the health state, coupled in a pairwise network. The current state of health then is used to compute the mortality rate. In this model the state of health

is not a set of components that can damage or fail, but a real-valued vector evolving with age according to stochastic differential equations. While the mathematical form of these models include pair-wise network interactions, the model has not been interpreted as such because only small datasets with one or two health variables have been used with this model. We return to this model and greatly expand upon it in Chapter 5.

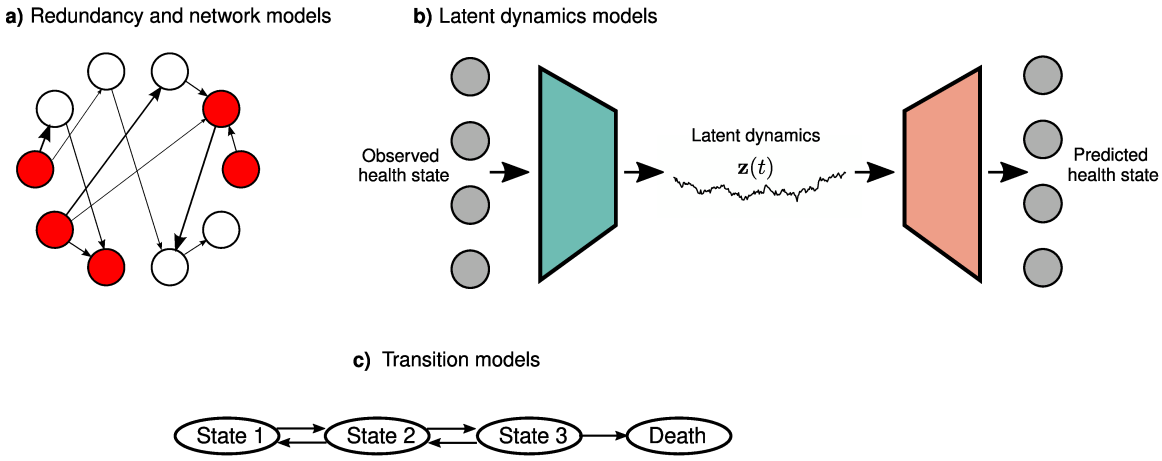


Figure 1.1: **Dynamical models of aging.** **a)** Models of redundancy with damageable components. Undamaged components (white) can damage (red), with potential interactions (arrows) propagating the damage. **b)** Latent variable dynamics models. The observed health state is mapped to a latent state $\mathbf{z}$ with an *encoder* (blue trapezoid). Dynamics are run on these latent variables. Latent variables are then mapped back to the predicted health state with a *decoder* (orange trapezoid). **c)** Transition models describe transitions between discrete health states. Eventually transition to death occurs.

1.2.2 Dynamical latent variable models of aging

Machine learning models of aging are flexible models that learn from data. Generally, these approaches focus on learning *latent* variables to describe the dynamics of aging. The latent state is typically of a lower dimension than the observed health state, and so can compress and simplify the relevant information. Figure 1.1b) shows the basic form of these models. The observed health state is transformed to a latent health state, which evolves with age. This latent state is then transformed back into predicted health states at a later age.

Pierson *et al.* [37] infer a set of *rates* of aging, which describe the deterministic

linear evolution of latent variables with age. Avchaciov *et al.* [21] describe aging with stochastic one-dimensional latent variable dynamics. These models simplify the dynamics of aging, reducing the dynamics to just the key components.

One of the difficulties with latent variable models is that in general, they will be difficult to interpret since the latent variables will often be a complex combination of the relevant information from all of the observed health variables. To solve this problem, we can include *structure* in the latent variables, encouraging them to extract desired features of the dynamics of aging. In Chapter 7, I develop a latent variable clustering approach. This model clusters the latent dynamics of aging, inferring distinct trajectories of aging.

1.2.3 Transition models

Transition models describe aging in terms of a set of discrete states representing distinct health states, and transitions between them, until a final transition to death. These approaches simplify the dynamics of aging in terms of a finite set of states. Figure 1.1c) shows an example of these models. Both forward and background transitions between the states can occur. The redundancy and network models from Figure 1.1a) can also be thought of as transition models, since components transition from healthy to damaged states.

Mitnitski *et al.* [38–41] model transitions between discrete health states and mortality. This is done by modelling the transitions between different numbers of binary health deficits (i.e. damageable health components). However this approach was simplistic, and only parameterized the mean number of binary health deficits after a period of time, rather than explicit transition rates. Oswal *et al.* [42] model transitions between movement states of C. elegans — where healthy worms exhibit vigourous movement. This model is used to separately describe movement speed and lifespan and the effects of interventions.

In Chapter 8 we directly infer the rates of damage and repair transitions from longitudinal data. This approach is used to probe the processes of damage and repair in aging.

1.2.4 Biological models

Other approaches model specific aspects of aging biology. Karin *et al.* [43] model the dynamics of the accumulation of senescent cells. Senescent cells cease dividing and accumulate with age, and can cause chronic inflammation, increasing the risk of aging-related diseases. They use this model to provide an explanation for the exponential increase of the mortality rate in Gompertz' law.

Podolskiy *et al.* [13] model the critical dynamics of gene regulatory networks. Gene regulatory networks govern the expression of genes, determining cell function. Their work studies the link between the instability of these regulatory networks and mortality rate, and how this can give rise to Gompertz' law.

This thesis does not involve such specific biological models that focus on a single system, and instead focuses on more general models that incorporate multiple aspects of aging health.

1.3 Outline and contributions

This thesis is in publication format with five related papers in separate chapters. Two of these papers are published, one is under review, and two are in preparation for submission. Each chapter is preceded by background information to thematically link the chapters within the thesis.

Chapter 2 provides background to understand the contents of the thesis. It covers concepts in aging research and an introduction to the modelling techniques used throughout the rest of the thesis.

Chapters 3, 4, and 5 form a sequence of papers about the structure of physiological interaction networks in aging. In Chapter 3 we use a previously developed [11, 44] descriptive network model of aging and indirectly probe the network structure, to describe the hierarchical organization of damage propagation in aging. This work is published as "Probing the network structure of health deficits in human aging", authored by Spencer Farrell (myself), Arnold Mitnitski, Olga Theou, Kenneth Rockwood, and Andrew Rutenberg in *Physical Review E* **98** 032302 (2018) [45]. Two review articles were also written involving previous work [11, 44] with this model, where I was a co-author [46, 47].

In Chapter 4 we build from the descriptive network model in Chapter 3 by making it predictive of individual health and mortality. Using cross-sectional aging data, we infer small networks of binary health variables. This work is published as "Generating individual aging trajectories with a network model using cross-sectional data", authored by Spencer Farrell, Arnold Mitnitski, Kenneth Rockwood, and Andrew Rutenberg in *Scientific Reports* **10** 19833 (2020) [48]. A review article was written about the potential of computational models of aging, including the work of Chapter 4 and foreshadowing the work of Chapter 5, published as "The potential for complex computational models of aging ", authored by Spencer Farrell, Garrett Stubbings, Kenneth Rockwood, Arnold Mitnitski, and Andrew Rutenberg in *Mechanisms of Ageing and Development* **193** 111403 (2021) [23].

In Chapter 5 we develop a new kind of network model, involving a larger network of continuous health variables. This model is fit using a large longitudinal aging dataset. This model is predictive of individual aging trajectories, and infers a network describing the interactions between the health variables. This work is accepted for publication as "Interpretable machine learning for high-dimensional trajectories of aging health", authored by Spencer Farrell, Arnold Mitnitski, Kenneth Rockwood and Andrew Rutenberg in *PLoS Compuational Biology*.

Chapters 3, 4, and 5 form a large portion of the work in the thesis with the central theme of modelling aging in terms of interaction networks of health variables. A summary of this work is in Chapter 6.

Chapter 7 is a follow-up to Chapter 5, building off of this work. This work is not yet ready for submission, but is an extension of the types of models discussed in Chapter 5. We explore an alternative to the models involving networks of interactions between the health variables developed in the preceding chapters, using latent variables. This model infers *clusters* of distinct aging trajectories.

In Chapter 8, we study the processes of damage and repair in aging. We use mouse and human data and develop a new method to extract damage and repair from longitudinal data. This work explores the effects of age and interventions on these processes. This work is in preparation for submission, authored by Spencer Farrell, Alice Kane, Elise Bisset, Susan Howlett, and Andrew Rutenberg.

Chapter 9 returns to the three fundamental questions discussed above: how can

we better understand the mechanisms or causes underlying what we observe, how can we better predict outcomes at an individual or population level, and, finally, how can we better intervene to decrease mortality and to improve health during aging? This chapter discusses the progress this thesis has made on these questions, and potential future work.

Other publications I authored or co-authored during my PhD are not included in this thesis: work on understanding deficit binarization in the Frailty Index, where I was a co-author [49], and two first author publications on the statistical physics of Single-file diffusion of particles in narrow channels and pores [50, 51].

Chapter 2

Background

This chapter provides useful background for the rest of the thesis. I do not attempt to survey all aspects of aging research, data analysis, or machine learning, just some key aspects and concepts.

2.1 Measures of aging

Aging involves the accumulation of damage that leads to eventual death [4, 52]. This process is stochastic and multi-dimensional, involving changes at all physical scales of an organism [53, 54]. At the fundamental biological level, key hallmarks or pillars of aging have been identified [4, 5]. However, how these interact and contribute to observed functional decline is largely unknown [54–56].

The different aspects of aging are studied with biomarkers describing the relevant features [54, 56, 57]. At the lowest biological level, these markers consists of 'omics data such as genomics, epigenomics, proteomics, and others [58–60]. These measures are used to study aging at the biological level and have been used to identify molecular pathways involved in aging, but large longitudinal studies containing these measurements are rare. Many blood and urine biomarkers of aging have been identified including inflammatory markers, immune markers, markers of oxidative stress, and other common blood test markers [57]. Clinical health deficits that measure physical functioning and disability are also used as markers of aging [14, 16]. These measures are the more readily apparent aspects of aging, such as difficulty walking, difficulty performing daily activities, and organ dysfunction or disease.

2.2 Summary measures of aging

Since aging involves many changes in organism functioning, it is useful to have "overall health" measures that can measure the progression of aging [61]. Since humans

have long life-spans, this is especially important for evaluating the impact of aging interventions without waiting the entire duration of an individuals life [61]. Many methods to measure the progression of aging have been developed. These methods differ by both the method of estimation, and the aspects of health measured. In this thesis these methods are referred to as "summary measures" of aging, since they summarize aging health into a single dimension. However, these have also been called "composite biomarkers", since they integrate many biomarkers together [57]. Summary measures can be interpreted as one-dimensional *latent* variables describing aging.

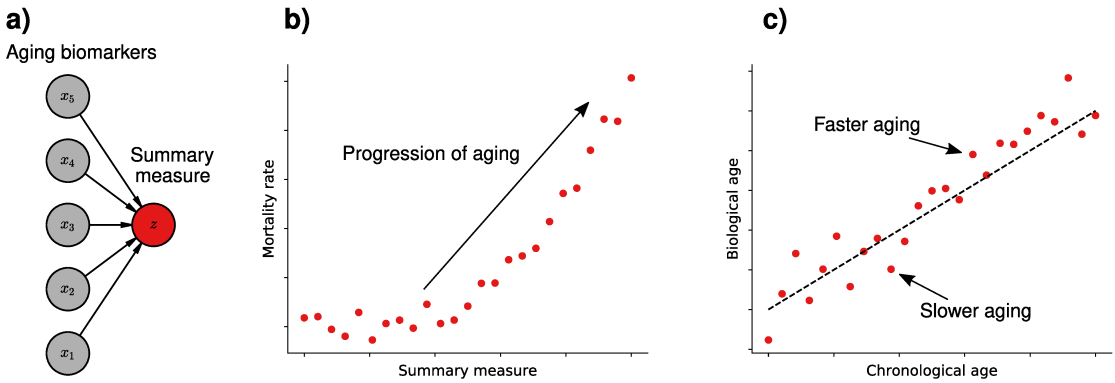


Figure 2.1: **Summary measures of aging.** **a)** The creation of a summary measure of aging. Various biomarkers are combined into a single summary. For the Frailty Index, the input $\mathbf{x}$ biomarkers are binary-valued health deficits. For biological age, these are typically continuous-valued molecular biomarkers. **b)** Summary measuring of aging aim to summarize the progression of aging. As the progression of aging proceeds as measured by the summary measure, the mortality rate increases. **c)** Biological age is a particular summary measure that has a clear interpretation when compared with chronological age. When biological age is above chronological age, the individual is said to have faster aging. When biological age is below chronological age, the individual is said to have slower aging.

One method that is of particular interest in this thesis is the Frailty Index (FI), developed by Arnold Mitnitski and Kenneth Rockwood at Dalhousie University [62]. The FI was developed to measure physical frailty, a state of increased vulnerability to adverse health outcomes [63]. The FI is measured as the fraction of binary health deficits $d_i \in \{0, 1\}$ an individual has out of a set of total potential N deficits, $f = \sum_{i=1}^{N} d_i/N$. These health deficits are usually binary-valued health variables that increase in prevalence as individuals age [15]. However in some cases multi-valued

health deficits are used, where there are multiple intermediate values of the deficit representing partial damage, e.g. $d_i \in \{0, 0.5, 1\}$ or $d_i \in \{0, 0.25, 0.5, 0.75, 1\}$. In all cases, the FI has a lower bound of 0 and an upper bound 1, although the observed maximum in observed data is typically around 0.6-0.8 [15, 64–68].

Different versions of the Frailty Index have been developed to measure different aspects of health. This is achieved by choosing health deficits that capture the desired features of health. Versions of the FI have been developed for physical or functional health (called FI-clin) [14, 15, 69], for blood and urine test biomarkers (called FI-lab) [49,70–72], and for mice [73]. For example, FI-clin typically contains deficits such as "Difficulty walking", "Require help bathing", and "High blood pressure", while FI-lab typically contains blood markers variables such as "Glucose", "Cholesterol", and "Hemoglobin" with cut-points to binarize based on healthy ranges [49]. Other related measures of frailty exist such as the Fraily Phenotype [16] and Clinical Frailty Scale [74], although this thesis uses the FI as a measure of frailty.

A common summary measure is "biological age" (BA). Estimates of BA are constructed by building a model to predict chronological age, with aging biomarkers as covariates [17]. The predicted age is called biological age, and represents a mapping between the observed biomarkers, and the expected value of age given the value of these observed biomarkers. Positive differences between BA and chronological age indicate accelerated aging, and negative differences between BA and chronological age indicate slowed aging. BA can be estimated with any type of biomarker data, but has had recent success with epigenetic data and other 'omics [18, 58–60, 75–77]. The advantage of biological age over other summary measures are its ease of interpretation, as it has units of age, and it can be constructed from different types of biological data to create distinct summary measures, summarizing different aspects of health [78].

2.3 Model organisms in aging

Since humans live long lives and it is difficult to do controlled studies, aging is often studied with model organisms [79]. The advantages of animal models of aging include more closely controlled experiments, genetic manipulation, and easier to perform interventions. The majority of this thesis uses human aging data, however Chapter

7 uses data from C. elegans, and Chapter 8 uses data from mice.

C. elegans is a species of roundworm often used as a model organism due to its short lifespan, ease of study, ease of manipulation, and well understood biology [80]. In aging research, techniques have been developed to perform large high-throughput experiments to track time-series of health and lifespan of large numbers of worms by automated imaging [32, 81–83]. This feature of C. elegans is exploited in Chapter 7, where we make use of C. elegans time-series data.

While C. elegans enables high-throughput experiments, direct translation between C. elegans and humans can be difficult [84]. It is more difficult to do large studies with mice than C. elegans, however mice are used in aging research due to their similarly to humans in comparison to other model organisms. This makes mice a useful organism for the study of aging interventions [85, 86]. In Chapter 8, we use mice data that includes interventions that have previously been shown to slow down functional decline during aging [87, 88].

Many other organisms are utilized in aging research, each with their own benefits and downsides [89], however this thesis only uses data from humans, worms, and mice.

2.4 Resilience and robustness in aging

In aging, the quantities of resilience and robustness can be distinguished. Broadly, resilience corresponds to the ability to recover after deviation from a healthy physiological state, and robustness corresponds to the ability to resist a deviation from a healthy physiological state [90–94]. Resilience has been observed by measuring the health state as a function of time following an acute stressor [95–97]. Measurement of resilience can be of the length of time it takes to recover, or the deviation from the original baseline physiological state after a less than full recovery. There has been less work on the observation of robustness, although proxies such as the ability to resist onset of disease have been used [20].

Figure 2.2 demonstrates these concepts. a) shows the perturbation of a physiological state variable after a stressor, followed by recovery. Higher resilience results in a faster recovery to baseline or a final recovered health state that is closer to the original baseline. b) shows the perturbation of a physiological state variable after a stressor,

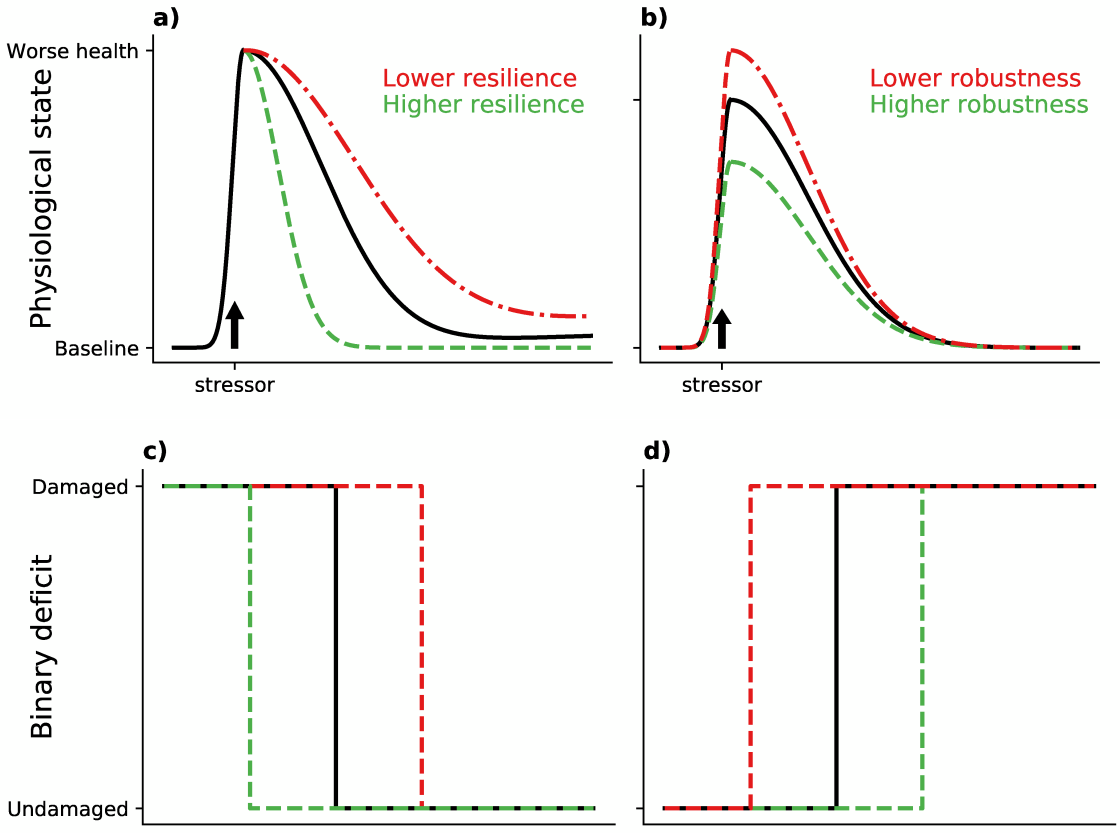


Figure 2.2: **Resilience and robustness.** **a)** Example of resilience. A stressor causes the physiological state to deviate from the health baseline, and resilience measures the ability to return to the baseline. **b)** Example of robustness. A stressor causes the physiological state to deviate from the health baseline, and robustness measures the ability to resist this deviation. **c)** Binary resilience is measured by time to transition back to an undamaged or healthy state. **d)** Binary robustness is measured by time to transition to a damaged state.

with robustness reducing the magnitude of the stressors effect on the physiological state. A higher robustness would result in a smaller deviation from baseline.

In this thesis, measures of resilience and robustness are developed for binary health deficits. Figure 2.2c) shows how resilience for a binary health deficit is interpreted. Higher resilience results in a faster repair transition from a damaged to an undamaged state. Figure 2.2d) shows how robustness for a binary health deficit is interpreted. Higher robustness results in a slower transition to a damaged state, or remaining undamaged altogether.

With longitudinal data, robustness and resilience can be observed in binary health

deficits through damage $0 \to 1$ and repair $1 \to 0$ transitions. The evolution of resilience and robustness is explored with age in Chapter 8 in both mice and humans, and with interventions using longitudinal binary health deficits.

2.5 Complex networks

Since aging involves interconnected physiological systems, it is natural to model this process with complex networks. A complex network is a set of *nodes* connected together by *edges* or *connections*. In this work, nodes represent health variables and connections between the nodes represent causal interactions between the health variables.

The entire N node network can be described by an *adjacency matrix*, $\{a_{ij}\}_{i,j=1}^{N}$. The elements of this matrix indicate the presence of connections between the nodes. Matrix elements a_{ij} take value 1 if there is connection from node j to node i, and is 0 otherwise. *Undirected* networks have a symmetric adjacency matrix with two-way connections $a_{ij} = a_{ji}$, *directed* networks have an asymmetric adjacency matrix, with connections potentially only going in one direction, e.g. j to i but not i to j.

The degree of a node is the number of edges connected to it, i.e. the number of neighbours $k_i = \sum_j a_{ij}$. Note that for a directed network, this is the *in-degree*, and alternatively the *out-degree* can be defined by $\sum_i a_{ij}$. Another important quantity is the average nearest neighbour degree, which is the average degree of the nodes connected to a node $k_{i,nn} = \sum_{j,l} a_{ij}a_{lj}/k_i$ (defined for a symmetric network).

Instead of specifying the full $N \times N$ adjacency matrix, a simple way to characterize large complex networks is by their degree distribution $p(k)$, representing the probability that a randomly selected node has a degree k. Additionally, networks can be characterized by their nearest neighbour degree distribution $p(k'|k)$, which is the probability that a randomly selected node that is connected to a node of degree k has degree k'. The average degree of the network and average nearest neighbour degree of a node of degree k follow from these distributions,

$$\langle k \rangle = \sum_k kp(k), \tag{2.1}$$

$$k_{nn}(k) = \sum_{k'} k'p(k'|k). \tag{2.2}$$

A common complex network structure, and used in this thesis in Chapter 3, is the Barabási-Albert preferential attachment network structure [98]. This network structure has a power-law degree distribution $p(k) \sim k^{-\alpha}$, with many low degree nodes and few high degree nodes. An algorithm is available to generate these networks with arbitrary exponent α and average degree $\langle k \rangle$ [99, 100]. This network is called a "preferential attachment" network due to the iterative generation process; a node is added at each iteration by connecting to existing nodes with probability proportional to the degree of the existing node. This process leads to networks with a few high degree nodes and many low degree nodes

An important characteristic of complex networks discussed in this thesis is the *assortativity* of a network. Assortativity is the extent to which nodes in the network are connected to nodes of a similar degree. An *assortative* network has many nodes connected to other nodes of a similar degree. A *disassortative* network has many low degree nodes connected to high degree nodes, leading to few big "hub" nodes connected to many low degree nodes. The assortativity of a network can be measured by observing the average degree of the neighbours of a node of degree k, $k_{\mathrm{nn}}(k)$. For assortative networks, $k_{\mathrm{nn}}(k)$ increases with k, while it decreases with k for disassortative networks. This is demonstrated in Figure 2.3.

The assortativity of networks generated by preferential attachment depends on their exponent α, with $\alpha < 3$ being disassortative, and $\alpha > 3$ being assortative in the large N limit [101]. However, we can arbitrarily modify the assortativity while preserving the degree distribution by rewiring connections [102].

These network structures are explored in the context of a complex network model of aging in Chapter 3.

2.6 Information entropy and mutual information

Shannon entropy or *information* entropy, represents the amount of uncertainty in the possible outcomes in a random variable. Given a random variable Y and its probability distribution $p(y)$, the information entropy of this variable is computed in the same way as the Gibbs entropy in statistical physics ($k_b = 1$),

$$H[Y] = -\sum_{y} p(y) \log p(y). \tag{2.3}$$

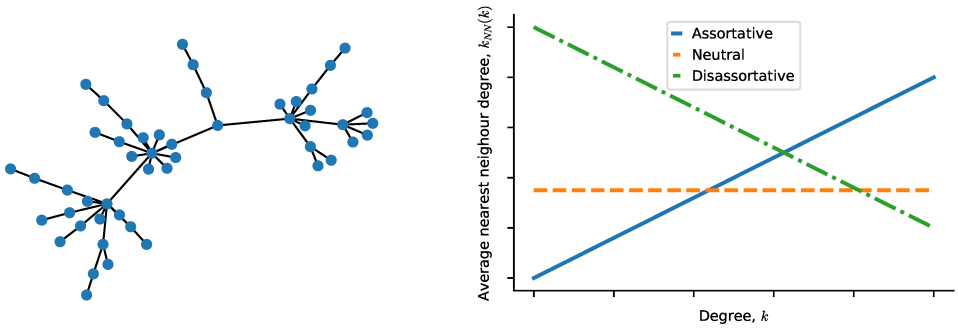


Figure 2.3: **Scale free networks and network assortativity.** Left, a Barabási-Albert scale-free network with average degree 2 and exponent 3. There are a few "hub" nodes with many connections, and many low-degree nodes. Right, characterizing the assortativity of complex networks. Networks with increasing average nearest neighbour degree vs degree are assortative. Networks with decreasing nearest neighbour degree vs degree are disassortative.

When Y is continuous, the sum becomes an integral over y.

The entropy can be interpreted as a measure of the width or "dispersion" of the distribution $p(y)$. For example, the entropy of a normally distributed variable is monotonic with the variance, $H \sim \log \sigma^2$.

With a second variable X, the conditional entropy of Y given X is,

$$H[Y|X] = -\sum_x p(x) \sum_y p(y|x) \log p(y|x), \tag{2.4}$$

$$= \sum_x p(x) H[Y|X = x]. \tag{2.5}$$

The specific conditional entropy $H[Y|X = x]$ is the entropy of Y conditioned on a specific value x of X. The conditional entropy $H[Y|X]$ is the average of this specific conditional entropy over all possible values of X. Note that $H[Y] \geq H[Y|X]$, since knowledge of X can only reduce our uncertainty in Y. If X provides no information about Y, $H[Y] = H[Y|X]$.

With the conditional entropy, we can define the *mutual information* between X and Y,

$$I(X;Y) = H[Y] - H[Y|X]. \tag{2.6}$$

The mutual information is the *reduction in uncertainty* in Y when knowing X. This can be used as a measure of non-linear association between variables. If two variables

are strongly related, the mutual information between them is large. Note, the mutual information is symmetric, $I(X;Y) = I(Y;X)$.

Care needs to be taken when computing entropy from noisy data. For a continuous-valued variable Y, estimating $H[Y]$ by binning to estimate $p(y)$ then applying Equation 2.3 is known to be a poor estimate of the entropy [103]. Instead, an approach based on ordered sample-spacings is used in this thesis [104–107].

Mutual information is used to characterize the amount of information in the Frailty Index in Chapter 3.

2.7 Feed-forward neural networks

In many areas of data-based modelling, there are unknown functions that need to be fit or *learned*. In some cases we have knowledge of the particular problem that restricts functions to a specific form, for example in physics often the form of a function is known up to the value of physical constants, which can be then fit from data. However in many problems, no specific form of the function is known, and so the unknown function to be fit needs to be flexible enough to capture the desired behavior. Neural networks are one such method of building arbitrarily flexible functions, with parameters that are fit or learned from the data. The field of "Deep Learning" builds models with neural networks, and has become the standard approach to building complex machine learning models [108].

Neural networks consist of a series of linear transformations, each followed by a non-linear *activation* function. For an input vector $\mathbf{x} \in \mathbb{R}^N$, a feed-forward neural network produces an output $\mathbf{y} \in \mathbb{R}^M$ by transforming the input with a linear transformation and non-linear activation function between each intermediate *layer* $\mathbf{h}_l \in N^l$. Applying all layers, this produces an output in the following way,

$$\mathbf{h}_0 = \mathbf{x}, \tag{2.7}$$

$$\mathbf{h}_l = g_l(\mathbf{W}_l \mathbf{h}_{l-1} + \mathbf{b}_l),\ \mathbf{W}_l \in \mathbb{R}^{N_l \times N_{l-1}},\ \mathbf{b}_l \in \mathbb{R}^{N_{l-1}},\ l = 1...L, \tag{2.8}$$

$$\mathbf{y} = \mathbf{h}_L. \tag{2.9}$$

The functions g_l are the non-linear activation functions, $\mathbf{W}_l$ are the weight matrices, and $\mathbf{b}_l$ are the bias vectors. These weight matrices and bias vectors are free parameters

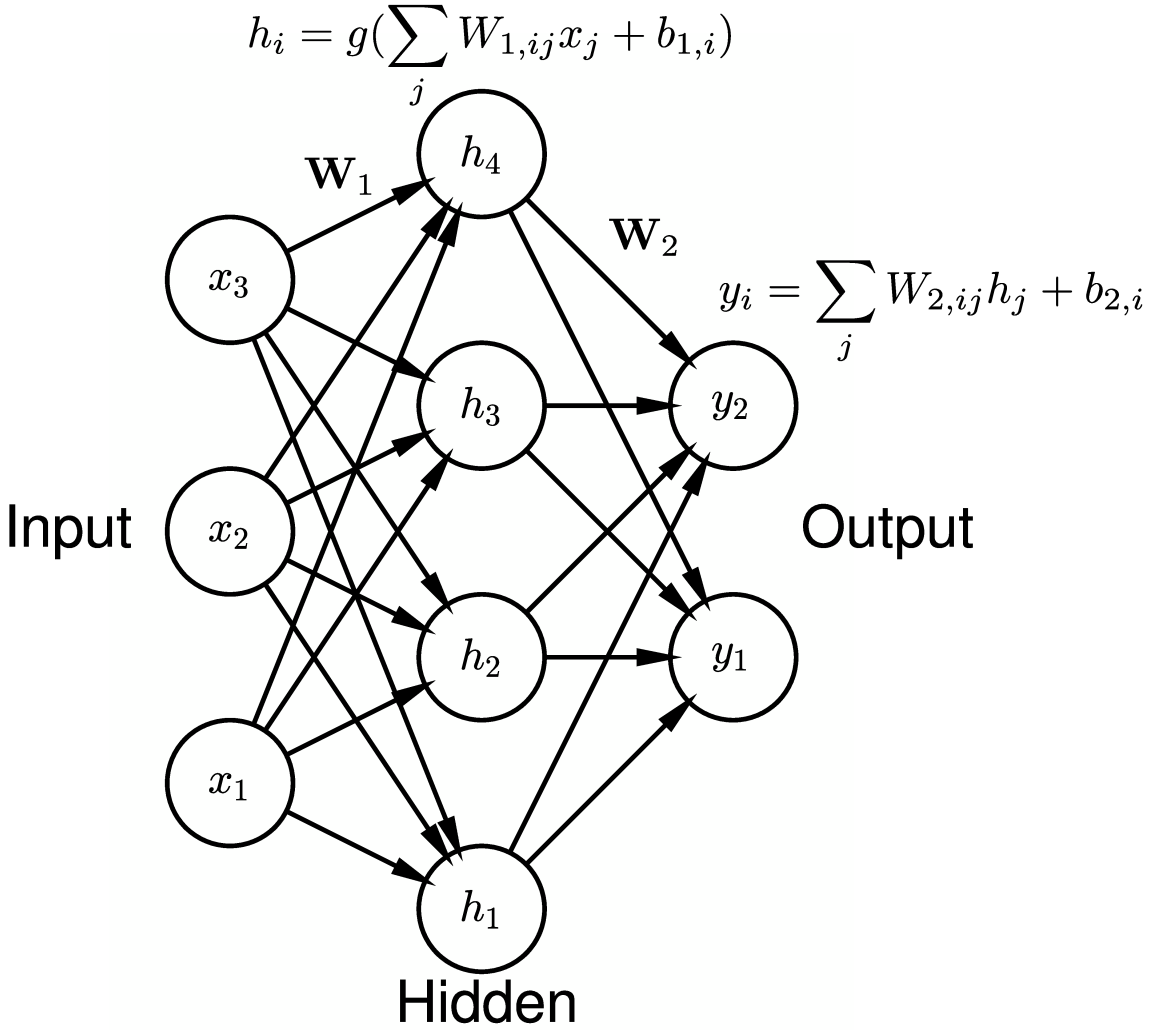


Figure 2.4: **Feed-forward neural network.** Feed-forward neural network with input $\mathbf{x} \in \mathbb{R}^3$. This neural network has a single hidden layer of size 4, $\mathbf{h} \in \mathbb{R}^4$. The weight matrices for the layers are $\mathbf{W}_1$ and $\mathbf{W}_2$, and the bias vectors are $\mathbf{b}_1$ and $\mathbf{b}_2$. The output is $\mathbf{y} \in \mathbb{R}^2$.

to be learned from data. Typically activation functions of the form $g(\mathbf{z}) = \max(0, \mathbf{z})$ or similar are used for the hidden layers (called ReLU, or Rectified Linear Units). For the output layer, the activation function is used to map the output to a desired specific domain, for example a sigmoid $\mathbf{y} = 1/(1 + e^{-\mathbf{z}})$ can be used for $\mathbf{y} \in [0, 1]^M$, a soft-plus $\mathbf{y} = \log(1 + e^{\mathbf{z}})$ can be used for positive outputs $\mathbf{y} \in (0, \infty)^M$, or a linear activation $\mathbf{y} = \mathbf{z}$ for real-valued outputs $\mathbf{y} \in (-\infty, \infty)^M$. A diagram of a feed-forward neural network is shown in Figure 2.4.

Increasingly flexible functions can be represented in this way by increasing the depth of the network L, or increasing the width of the layers $\{N_l\}_{l=1}^{L}$. With this form, arbitrary functions can be learned, given a large enough neural network [109]. In this thesis, neural networks are used to parameterize unknown functions within larger models of aging. For an example of this application in other parts of science, neural networks have been used to represent unknown terms in partially-known differential equations, and the neural network parameters are then fit from data [24]. This approach allows scientific knowledge to be included within flexible models with neural networks.

Neural networks are trained with variations of stochastic gradient descent, which requires derivatives of the output with respect to the parameters of the layers. The analytic derivatives of neural network layers are computed with the back-propagation algorithm, which uses the chain rule to recursively compute derivatives of every layer [110]. Modern computational frameworks compute these derivatives automatically, and so sophisticated differentiable models that combine neural networks with other model components can be easily implemented and trained.

For gradient-based optimization, this thesis uses the Adam optimizer [111], which is a variant of stochastic gradient descent. This method computes an adaptive learning rate from estimates of moments of the gradient, which allows it to improve gradient estimates from noisy samples.

2.8 Recurrent Neural Networks

Section 2.7 discussed feed-forward neural networks, which pass a vector input $\mathbf{x}$ through a series of linear transformation layers and non-linear activation functions to arrive at the output. Recurrent neural networks (RNN) differ by possessing an

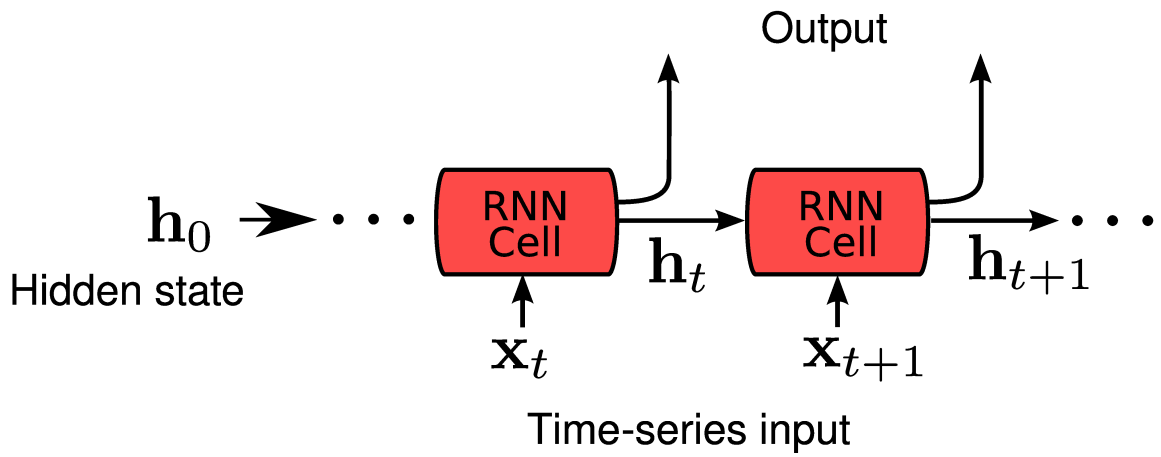


Figure 2.5: **Recurrent neural network.** A time-series or sequence $\{\mathbf{x}_t\}_{t=1}^T$ is recursively used as input in the RNN. A hidden "memory" state $\mathbf{h}_t$ allows features from past time-points to influence the output at current time-points.

internal state $\mathbf{h}_t$ (i.e. memory) that is updated for each vector $\mathbf{x}_t$ in a sequence $\{\mathbf{x}_t\}_t$. This allows unknown functions to be parameterized that depend not only on a single input, but a variable-length sequence, $f(\{\mathbf{x}_t\}_t)$.

Such RNNs consist of a *unit cell* that is applied recurrently over all elements of a length sequence $\{\mathbf{x}_t\}_{t=1}^T$,

$$\mathbf{h}_t = \text{RNN}(\mathbf{h}_{t-1}, \mathbf{x}_t), \ t = 1, ..., T, \tag{2.10}$$

given an initial internal state $\mathbf{h}_0$. The internal state $\mathbf{h}_t$ is taken as the output, and then transformed as desired with additional neural network layers and activation functions. A diagram of this process is shown in Figure 2.5. Since the same unit cell is applied recursively, an RNN can be used for datasets with variable-length sequences.

The particular structure of the unit cell depends on the type of RNN. In this thesis the Gated Recurrent Unit (GRU) is used [112], since it has a simpler structure and fewer parameters than other common forms of RNNs. The GRU uses linear transformations and non-linear activation functions to update the internal state with two different processes: selectively retaining information from the previous internal state $\mathbf{h}_{t-1}$, and including the relevant information from the current input $\mathbf{x}_t$.

2.9 Maximum likelihood fitting

The likelihood $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})$ is the probability of observing data $\mathbf{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^{M}$ and $\mathbf{X} = \{\mathbf{x}^{(i)}\}_{i=1}^{M}$ from a model with parameters $\boldsymbol{\theta}$. Maximum likelihood fitting regards the log-likelihood as a function of the parameters, $L(\boldsymbol{\theta}) = \log p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})$, then maximizes this function to find the optimal parameters. These optimal parameters result in the highest likelihood for the data for the specific model. Typically, data-points are assumed independent in the likelihood, $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_i p(\mathbf{y}^{(i)}|\mathbf{x}^{(i)}, \boldsymbol{\theta})$.

The form of the likelihood is assumed as part of the model. For example, a Gaussian likelihood can be assumed for a model with a mean computed with a neural network and constant variance,

$$p(\mathbf{y}^{(i)}|\mathbf{x}^{(i)}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}^{(i)}|\boldsymbol{\mu}(\mathbf{x}^{(i)}; \boldsymbol{\theta}), \boldsymbol{\sigma}^2). \tag{2.11}$$

The maximum log-likelihood for this model reduces to the minimum squared deviation between the observed value and the mean,

$$\text{argmax}_{\boldsymbol{\theta}}\ L(\boldsymbol{\theta}) = \text{argmin}_{\boldsymbol{\theta}}\ \sum_i |\mathbf{y}^{(i)} - \boldsymbol{\mu}(\mathbf{x}^{(i)}; \boldsymbol{\theta})|^2, \tag{2.12}$$

leading to least-squares fitting. For neural network models, the negative log-likelihood can be minimized by stochastic gradient descent.

2.10 Bayesian modelling

A Bayesian approach is used to formulate the probabilistic models discussed in Chapters 5, 7, and 8. Given a dataset $\mathbf{Y}$ and $\mathbf{X}$ and a model with parameters $\boldsymbol{\theta}$, this approach makes use of Bayes' Theorem,

$$p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}) = \frac{p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{Y}|\mathbf{X})}, \tag{2.13}$$

which formulates the *posterior* distribution of the parameters $p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X})$ in terms of the *likelihood* $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})$ (as in Section 2.9), *prior* distribution of parameters $p(\boldsymbol{\theta})$, and normalization factor $p(\mathbf{Y}|\mathbf{X}) = \int p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ known as the *model evidence*.

The posterior distribution is the main quantity of interest, representing the most probable values of the model parameters given the observed data. The Bayesian approach is appealing because the posterior distribution naturally includes uncertainty

for the model parameters with a distribution of possible values, rather than just a single estimate of parameters that maximize the likelihood. Additionally, uncertainty for model predictions $\mathbf{y}'$ at a new data point $\mathbf{x}'$, given the observed data $\mathbf{Y}$ and $\mathbf{X}$ can be also estimated by integrating the likelihood of this prediction $p(\mathbf{y}'|\mathbf{x}', \boldsymbol{\theta})$ over the posterior distribution of parameters $p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X})$,

$$p(\mathbf{y}'|\mathbf{x}', \mathbf{Y}, \mathbf{X}) = \int p(\mathbf{y}'|\mathbf{x}', \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}) d\boldsymbol{\theta}. \tag{2.14}$$

Similar to the *confidence* interval in frequentist statistics, posterior distributions are generally summarized by computing the median and the interval of the posterior containing $\alpha\%$ of the mass of the distribution, called *credible* intervals. The $\alpha\%$ credible interval of a parameter contains the true value of the unknown parameter at probability $\alpha\%$. Note that this is distinct and more intuitive than a frequentist confidence interval, where $\alpha\%$ of the confidence intervals computed from repeated samples would include the true value of the parameter.

The prior distribution $p(\boldsymbol{\theta})$ incorporates prior knowledge of the value of the model parameters. If specific knowledge of the value of the parameter is known, priors can be placed around that value. Otherwise, priors can be used to obtain well-behaved models. Typically, this is done by placing narrow priors on parameters around zero, e.g. we may not know where the exact value of a parameter will lie, but we know it is extremely unlikely its value will be $> 10^3$. This can enforce sparsity in the model, limit extreme values of parameters, and prevent overfitting by regularizing.

The Bayesian approach is also a natural approach for learning latent variable models. These approaches are discussed below.

2.11 Markov-chain Monte-Carlo sampling

The calculation of the normalization factor in Bayes' Theorem,

$$p(\mathbf{Y}|\mathbf{X}) = \int p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{2.15}$$

can only be done analytically for the simplest of models, and can be seen as analogous to computing the partition function in statistical physics. Markov-chain Monte-Carlo (MCMC) is a method of sampling the posterior distribution without requiring the computation of this normalization factor. This allows the computation of integrals of

the posterior distribution by Monte-Carlo integration in cases where the normalization factor is unknown,

$$\int f(\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})d\boldsymbol{\theta} \approx \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{\theta}_i), \;\; \boldsymbol{\theta}_i \sim p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X}), \tag{2.16}$$

where $\{\boldsymbol{\theta}_i\}_{i=1}^{N}$ are MCMC samples from the posterior $p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})$, and $f(.)$ is an arbitrary function of which we want to integrate over the posterior, e.g. Equation 2.14 where $f(\boldsymbol{\theta}) = p(\mathbf{y}'|\mathbf{x}',\boldsymbol{\theta})$.

MCMC methods are stochastic processes that generate samples in such a way that the equilibrium distribution of the process is the desired posterior distribution. In this thesis the No U-Turn Sampler (NUTS) from the Stan probabilistic programming language [113] is used for MCMC sampling in Chapter 8. NUTS makes use of derivatives of the likelihood to speed up sampling, unlike much slower random-walk based Monte-Carlo methods such as the Metropolis-Hastings algorithm [114], and so is well suited for continuous valued parameters.

2.12 Variational Bayesian Inference

In high-dimensional scenarios with models containing many parameters and/or large datasets, even MCMC is computationally intractable. In particular when models include neural networks, MCMC will not be a viable option. A computationally faster approach is to use a parametric approximation to the posterior distribution instead of sampling it with MCMC.

A parametric approximation to the posterior is made by assuming a particular form of the approximate posterior $q(\boldsymbol{\theta};\boldsymbol{\phi})$ parameterized by the new *variational* parameters $\boldsymbol{\phi}$. The goal is to optimize the variational parameters so that this approximation is close to the true posterior, $q(\boldsymbol{\theta};\boldsymbol{\phi}) \approx p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})$.

A common way to do this is to minimize the KL-divergence between the two

distributions,

$$\min_{\boldsymbol{\phi}} \mathrm{KL}(q(\boldsymbol{\theta};\boldsymbol{\phi})||p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})) = \min_{\boldsymbol{\phi}} \int q(\boldsymbol{\theta};\boldsymbol{\phi}) \log \frac{q(\boldsymbol{\theta};\boldsymbol{\phi})}{p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})} d\boldsymbol{\theta}, \tag{2.17}$$

$$= \min_{\boldsymbol{\phi}} \mathbb{E}_q[\log q(\boldsymbol{\theta};\boldsymbol{\phi}) - \log p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})], \tag{2.18}$$

$$= \min_{\boldsymbol{\phi}} \big\{\mathbb{E}_q[\log q(\boldsymbol{\theta};\boldsymbol{\phi})] - \mathbb{E}_q[\log p(\boldsymbol{\theta},\mathbf{Y},\mathbf{X})] + \log p(\mathbf{Y}|\mathbf{X})\big\}. \tag{2.19}$$

To simplify the notation, expectations over the distribution $q(\boldsymbol{\theta};\boldsymbol{\phi})$ have been written as $\mathbb{E}_q[.]$. The last term $\log p(\mathbf{Y}|\mathbf{X})$ is the difficult to calculate posterior normalization factor or *model evidence* from Equation 2.15. This term does not involve $\boldsymbol{\phi}$, and so can be dropped from the optimization. The ability to drop this difficult to calculate term is key to the variational approach. Thus, the variational approximation is found by maximizing the following function with respect to the variational parameters $\boldsymbol{\phi}$,

$$\mathrm{ELBO}(\boldsymbol{\phi}) = \mathbb{E}_q[\log p(\boldsymbol{\theta},\mathbf{Y},\mathbf{X})] - \mathbb{E}_q[\log q(\boldsymbol{\theta};\boldsymbol{\phi})]. \tag{2.20}$$

This is known as the *Evidence Lower Bound* (ELBO), since it bounds the model evidence, $\log p(\mathbf{Y}|\mathbf{X}) \geq \mathrm{ELBO}(\boldsymbol{\phi})$. This approach is equivalent to variational free-energy minimization in statistical physics, with $\mathrm{ELBO}(\boldsymbol{\phi})$ representing the negative variational free-energy and $\log p(\mathbf{Y}|\mathbf{X})$ representing the exact negative free-energy.

To see this, note that the last term in 2.20 is the entropy of $q(\boldsymbol{\theta};\boldsymbol{\phi})$, as defined in Equation 2.3, and the first term is the expectation of the negative energy from a Boltzmann distribution $p(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}) = e^{-E(\mathbf{Y},\mathbf{X};\boldsymbol{\theta})}/Z(\boldsymbol{\theta})$ up to a constant in $\boldsymbol{\phi}$, $\log Z(\boldsymbol{\theta})$ (setting $\beta = k_b = 1$). This approach can easily be applied to the Ising model by assuming an independent distribution of spins q to compute the mean-field theory solution [115].

Another way to interpret the ELBO is to write it in terms of an expectation of the model likelihood and a KL-divergence between the posterior and prior for $\boldsymbol{\theta}$,

$$\mathrm{ELBO}(\boldsymbol{\phi}) = \mathbb{E}_q[\log p(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta})] - \mathrm{KL}(q(\boldsymbol{\theta};\boldsymbol{\phi})||p(\boldsymbol{\theta})). \tag{2.21}$$

This way, the variational inference approach can be seen as maximizing the expectation of the likelihood with samples from the posterior $q(\boldsymbol{\theta};\boldsymbol{\phi})$, while including a penalty term to enforce the posterior to be close to the prior $p(\boldsymbol{\theta})$ with the KL-divergence.

The simplest way to construct a variational approximation is with a mean-field approximation – where all parameters are assumed to be independent in the posterior,

$$q(\boldsymbol{\theta};\boldsymbol{\phi}) = \prod_i q(\theta_i;\boldsymbol{\phi}). \tag{2.22}$$

For real-valued $\boldsymbol{\theta}$, a simple mean-field approach is to use one-dimensional Gaussian distributions for each $q(\theta_i;\boldsymbol{\phi})$, where $\boldsymbol{\phi}$ is the set of means and standard deviations of these Gaussians $\{\mu_i, \sigma_i\}_i$. The inference then proceeds by maximizing the ELBO for these means and standard deviations. Mean-field approximations can be progressively improved by incorporating structure into the factorization of $q(\boldsymbol{\theta})$, but how this is done depends greatly on the structure of the model.

The variational approach is computationally much faster than MCMC because the approximate posterior is found by just optimizing the variational parameters $\boldsymbol{\phi}$, rather than adequately sampling a high-dimensional parameter space.

Variational Bayesian inference is used in Chapters 5 and 7.

2.13 Latent variable modelling and variational auto-encoders

One key application of variational Bayesian inference used in this thesis is the Variational auto-encoder (VAE) [116, 117]. The variational auto-encoder forms the basis of the models developed in Chapters 5 and 7, and forms the basis of a useful generic framework for Bayesian latent-variational modelling called deep latent variable modelling.

Latent variable models aim simplify a problem by introducing a latent variable $\mathbf{z}$. For observed variables $\mathbf{x}$, latent variables decompose the likelihood,

$$p(\mathbf{x}|\boldsymbol{\theta}) = \int p(\mathbf{x}|\mathbf{z},\boldsymbol{\theta})p(\mathbf{z})d\mathbf{z}. \tag{2.23}$$

Rather than directly modelling $p(\mathbf{x}|\boldsymbol{\theta})$, a latent variable model allows the development of a simpler model for $p(\mathbf{x}|\mathbf{z},\boldsymbol{\theta})$. For example, if $\mathbf{x}(t)$ exhibits complex dynamics in time, it can be simpler to model dynamics of a latent variable $\mathbf{z}(t)$, and then transform this latent variable to the observed variables $\mathbf{z}(t) \to \mathbf{x}(t)$.

To fit a latent variable model, we need to infer the latent variables from the data $p(\mathbf{z}|\mathbf{x})$, which can be recognized as the posterior distribution of $\mathbf{z}$. So we can use Bayesian inference for these latent variable models.

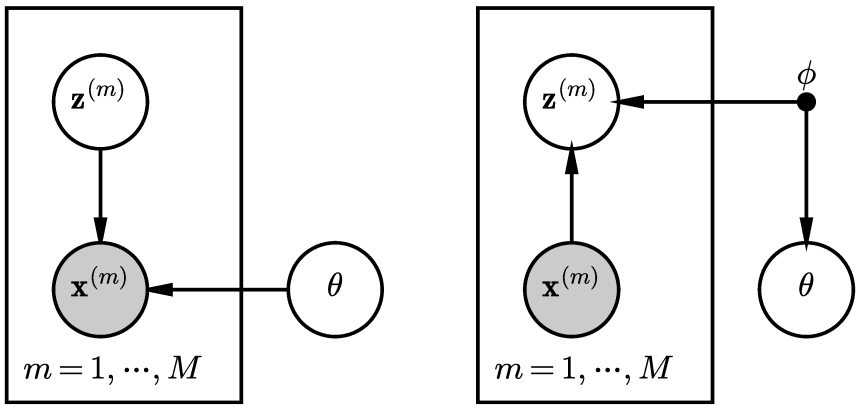


Figure 2.6: **Bayesian graphical notation for a latent variable model.** In this notation, grey nodes indicate observed variables and white nodes indicate unobserved variables i.e. latent variables). The box indicates variables that are indexed by $m = 1$ to $m = M$ (e.g. different individuals). Variables outside the box are global parameters. Arrows represent conditional dependencies in probability distributions. (Left) Generative form of the model, which shows the factorization of the joint distribution $p(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = p(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta})p(\mathbf{z})p(\boldsymbol{\theta})$. (Right) Factorization of the variational posterior approximation $q(\mathbf{z}, \boldsymbol{\theta}|\mathbf{x}; \boldsymbol{\phi}) = q(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})q(\boldsymbol{\theta}; \boldsymbol{\phi})$. The black point represents the parameters $\boldsymbol{\phi}$ to be optimized (not a stochastic variable described by a probability distribution).

Suppose we have a latent-variable model of the following form,

$$\begin{aligned}
\boldsymbol{\theta} &\sim p(\boldsymbol{\theta}), && (2.24)\\
\mathbf{z}^{(m)} &\sim p(\mathbf{z}^{(m)}), && (2.25)\\
\mathbf{x}^{(m)}|\mathbf{z}^{(m)}, \boldsymbol{\theta} &\sim p(\mathbf{x}^{(m)}|\mathbf{z}^{(m)}, \boldsymbol{\theta}), && (2.26)\\
p(\mathbf{x}^{(m)}, \mathbf{z}^{(m)}, \boldsymbol{\theta}) &= p(\mathbf{x}^{(m)}|\mathbf{z}^{(m)}, \boldsymbol{\theta})p(\mathbf{z}^{(m)})p(\boldsymbol{\theta}), && (2.27)\\
& \quad m = 1, ..., M. && (2.28)
\end{aligned}$$

This is called the *generative* form of the model, and describes how to generate outputs of observed variables $\mathbf{x}^{(m)}$, given the latent variables $\mathbf{z}^{(m)}$ and the parameters $\boldsymbol{\theta}$. Indexes m indicate separate individuals from $m = 1$ to $m = M$, so that the observed data and latent variables are individual specific, but the parameters are global. Equations 2.25 and 2.24 are prior distributions for the latent variables and parameters, and Equation 2.26 is the likelihood. Figure 2.6(left) shows a compact form to represent the structure of this model, which demonstrates the factorization

of the joint distribution in Equation 2.27.

In this model, we want to infer the unobserved latent variable $\mathbf{z}^{(m)}$ for each observed input $\mathbf{x}^{(m)}$. To do this, we need to perform Bayesian inference by computing the posterior $p(\mathbf{z}^{(m)}, \boldsymbol{\theta}|\mathbf{x}^{(m)})$. We build a variational approximation that follows the factorization shown in Figure 2.6(right). This has the form

$$q(\mathbf{z}^{(m)}, \boldsymbol{\theta}|\mathbf{x}^{(m)}; \boldsymbol{\phi}) = q(\mathbf{z}^{(m)}|\mathbf{x}^{(m)}; \boldsymbol{\phi})q(\boldsymbol{\theta}; \boldsymbol{\phi}). \tag{2.29}$$

Assuming a real-valued $\mathbf{z}$, the simplest approach is to assume independent Gaussian distributions (mean-field approximation as above) for $\mathbf{z}$. Since this distribution depends on the input $\mathbf{x}$, the means and standard deviations of this distribution need to depend on $\mathbf{x}$. These can be estimated with neural networks for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$,

$$q(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}; \boldsymbol{\phi}), \boldsymbol{\sigma}(\mathbf{x}; \boldsymbol{\phi})^2). \tag{2.30}$$

The neural network weights and biases along with the parameters of $q(\boldsymbol{\theta}; \boldsymbol{\phi})$ are the variational parameters $\boldsymbol{\phi}$ to be optimized by maximizing the ELBO as above in Equation 2.20,

$$\begin{aligned} \text{ELBO}(\boldsymbol{\phi}) \quad = \quad & \sum_m \Big\{ \mathbb{E}_q[\log p(\mathbf{x}^{(m)}|\mathbf{z}^{(m)}, \boldsymbol{\theta})] + \mathbb{E}_q[\log p(\mathbf{z}^{(m)})] - \mathbb{E}_q[\log q(\mathbf{z}^{(m)}|\mathbf{x}^{(m)}; \boldsymbol{\phi})] \Big\} \\ & + \mathbb{E}_q[\log p(\boldsymbol{\theta})] - \mathbb{E}_q[q(\boldsymbol{\theta}|\boldsymbol{\phi})]. \end{aligned} \tag{2.31}$$

The sum goes over all individuals (assuming independent individuals in the likelihood). This model is known as a *variational auto-encoder*. Generically, this model can be used as a framework for many different types of latent-variable models.

Note, often the variational posterior distribution of the global parameters $q(\boldsymbol{\theta}; \boldsymbol{\phi})$ is implicitly set to be a delta function for some or all of the parameters, $q(\boldsymbol{\theta}; \boldsymbol{\phi}) = \delta(\boldsymbol{\theta} - \boldsymbol{\phi})$. Together with a uniform prior $p(\boldsymbol{\theta})$, this has the effect of including some of the parameters $\boldsymbol{\theta}$ directly as variational parameters $\boldsymbol{\phi}$ to be optimized (rather than parameterizing a distribution of $\boldsymbol{\theta}$), simplifying the posterior [118]. This step is often done implicitly without indication.

2.14 Stochastic differential equations

A stochastic differential equation (SDE) describes the continuous-time evolution of a random variable, which takes the form of a differential equation that includes a

stochastic term. An SDE is written,

$$dx(t) = f(x,t)dt + \sigma(x,t)dB(t), \tag{2.32}$$

where f represents the *drift* term, and σ represents the strength of the diffusive noise. The stochastic component $B(t)$ is a standard Brownian motion with Gaussian increments with variance dt, $dB(t) = B(t+dt) - B(t) \sim \mathcal{N}(0, dt)$. Solutions of an SDE are realizations of the stochastic process described by the SDE.

2.15 Survival Analysis

Survival analysis is the analysis of time-to-event data. This naturally applies to the field of aging, where we have time-to-death or age-of-death data. The main quantity in survival analysis is the *Survival function* or *Survival probability*. The survival function $S(t)$ is the probability that an individual is alive at time t (e.g. has yet to have the event in question). The typical goal in survival analysis is to build models to estimate the survival function, or estimate the relative difference in the survival of distinct groups (e.g. the effect of treatment vs control for a drug).

Survival models are often best described by *hazard rates*. When the event in question is death, hazard rates can be called *mortality rates*. The hazard rate $h(t)$ is the instantaneous rate of the event occurring, and is related to the survival function in the following way,

$$S(t) = \exp\left(-\int_{t_0}^{t} h(t')dt'\right). \tag{2.33}$$

With time-to-event data, the goal is to fit the hazard rate of a survival model $h(t;\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$. However, it is common that many of the observations are censored — the period of observation for a subject concludes before the event (i.e. death) occurs, which is known as *right censoring*. Fitting a survival model requires modifying the likelihood to take this into account.

Consider a set of observation times for M different individuals $\{t_i\}_{i=1}^{M}$ and censoring indicators $\{c_i\}_{i=1}^{M}$, where the death either occurs ($c_i = 0$) or the individual is censored ($c_i = 1$). This means that t_i represents a time of death when $c_i = 0$, and is the time that the individual was last known to be alive if $c_i = 1$. Assuming a survival

model with parameters $\boldsymbol{\theta}$, the likelihood takes the form,

$$p(\{t_i\}_{i=1}^M, \{c_i\}_{i=1}^M|\boldsymbol{\theta}) = \prod_{i=1}^M [p(t_i|\boldsymbol{\theta})]^{(1-c_i)}[p(t > t_i|\boldsymbol{\theta})]^{c_i}, \quad (2.34)$$

$$= \prod_{i=1}^M h(t_i|\boldsymbol{\theta})^{(1-c_i)} S(t_i|\boldsymbol{\theta}). \quad (2.35)$$

Additionally, interval censored data is possible. In this case, observations are made at discrete time points, and it is known that the event occurs between two observation times, $t \in (t_i^L, t_i^U)$. For interval censored data, the likelihood becomes,

$$p(\{t_i^L\}_{i=1}^M, \{t_i^U\}_{i=1}^M|\boldsymbol{\theta}) = \prod_{i=1}^M p(t_i^L < t < t_i^U|\boldsymbol{\theta}), \quad (2.36)$$

$$= \prod_{i=1}^M [S(t_i^L|\boldsymbol{\theta}) - S(t_i^U|\boldsymbol{\theta})]. \quad (2.37)$$

Interval censored data is encountered in Chapter 8.

One of the most common survival models seen in the medical literature is the Cox Proportional hazards model. In this model, covariates $\mathbf{x}$ are linearly related to the log hazard rate,

$$h(t, \mathbf{x}; \boldsymbol{\theta}) = h_0(t) \exp(\boldsymbol{\theta} \cdot \mathbf{x}). \quad (2.38)$$

$h_0(t)$ is the baseline hazard rate, and can be any arbitrary function. To fit the Cox Proportional hazards model by maximum likelihood, the the baseline hazard is not required to be specified [119]. After the parameters $\boldsymbol{\theta}$ are fit, the baseline hazard can be determined non-parametrically with the Breslow estimator [120]. For a Bayesian proportional hazards model, the baseline hazard needs to be specified in terms of parameters $h_0(t; \boldsymbol{\theta})$. This is typically done with splines [121], and this approach is taken in Chapter 8.

In the proportional hazards model, a unit increase in a covariate x_i is multiplicative with respect to the hazard rate. This allows for the independent effects of covariates to be easily interpreted. For example, if there is a unit increase in covariate x_i and all others are held constant, the proportional increase of the hazard rate only depends on the parameter θ_i, $\frac{\exp(\boldsymbol{\theta}\cdot\mathbf{x}+\theta_i)}{\exp(\boldsymbol{\theta}\cdot\mathbf{x})} = \exp(\theta_i)$. This is known as the *hazard ratio* for covariate x_i.

The Cox model is used to compare with the models developed in this thesis in Chapters 4 and 5.

Chapter 3

Probing the network structure of health deficits in human aging

3.1 Background

Aging is widely considered to be the accumulation of damage with age [4, 122–124]. This suggests a model of propagating damage, where damage to a physiological system spreads to other interacting systems within the organism. In previous work, we developed a complex network model to describe this process of damage accumulation [11,44]. Other network models of aging have also been developed by other researchers [30–32]. While these other models focus on mortality, in our model we also focus on health, as measured by the Frailty Index.

In our model, nodes in a complex network represent damageable components of health. Nodes stochastically damage, which increases the damage rates of their neighbouring nodes. This results in the propagation of damage through the network, leading to eventual mortality when the two specific "mortality nodes" are damaged. In previous work, we demonstrated that this model captures population mortality rates and Frailty Index scores, and we used the model to quantify the information gained about mortality when knowing the Frailty Index [125]. This model can be used to generate a large simulated aging population, which allows the exploration of the relationships between aging, physical frailty, and mortality, which was discussed in two of our review papers [46, 47].

The nodes in this model are generic binary health attributes. These nodes are generic in the sense that they do not correspond to any specific physiological health variable, but are the abstract representation of a health component that is damaged during aging. This leads to the name of this model used in this thesis, the "generic network model" (GNM). These nodes are binary in the sense that they can either be in an undamaged state, or a damaged state. Since the nodes in the network are

generic, this is a descriptive model of aging, which does not make predictions for individuals.

The model requires the specification of the network structure. Previous work with this model used the Barabási-Albert preferential-attachment network structure [98], which contains few "hub" nodes with high degree, and many low degree nodes, forming a power-law degree distribution $p(k) \sim k^{-\alpha}$. With this network structure, the model is able to simulate populations with mortality and Frailty index scores that agree with population-level data. The only tuning of the network structure done was the adjustment of the degree distribution exponent α and the average degree $\langle k \rangle$. This motivates the questions of how important the specific structure of the network is, what are the important features of the network structure, and ultimately which network structures best capture population aging phenomena.

Distinct Frailty Indexes can be constructed for clinical markers (FI-clin) and for blood-test lab biomarkers (FI-lab). These show distinct behaviour vs age, with FI-lab having higher scores at younger ages, which suggests blood biomarker deficits precede clinical measures of frailty [71, 126, 127]. Assuming that different types of health deficits correspond to different types of nodes in the network, these two distinct Frailty Indexes allow us to probe the network. This motivates the questions of which nodes in the network correspond to FI-clin and FI-lab, and how does the network structure control the relation between these types of health deficits? This can offer an explanation of the differences between FI-clin and FI-lab in terms of the network structure.

In this chapter, we indirectly probe the network by examining the behaviour of the simulated aging population from the model for a variety of different network structures. Additionally, we develop a mean-field theory of the dynamics to specifically isolate the key structural features of the network.

This chapter presents the paper "Probing the network structure of health deficits in human aging" published in 2018 [128]. Note that the notation, figure numbers, and reference numbers have been modified from the published version for consistency within this thesis.

3.2 Introduction

Accumulation of damage is widely accepted as the mechanism behind organismal aging [52]. Even in model organisms, with controlled environment and genotype, there are large individual variations in lifespan and in the phenotypes of aging [53, 129]. While many mechanisms cause specific cellular damage [**?**], no single factor fully controls the process of aging. This suggests that the aging process is stochastic and results from a variety of damage mechanisms.

The variability of individual damage accumulation results in differing trajectories of individual health and in differing individual life-spans, and is a fundamental aspect of individual aging. A simple method of quantifying this individual damage is the Frailty Index (FI) [14, 15]. The FI is the proportion of age-related health issues ("deficits") that a person has out of a collection of health attributes. The FI is used as a quantitative tool in understanding the health of individuals as they age. There have been hundreds of papers using an FI based on self-report or clinical data, both for humans [130] and for animals [131]. Individuals typically accumulate deficits as they age, and so the FI increases with age across a population. The FI captures the heterogeneity in individual health and is predictive of both mortality and other health outcomes [69, 132–134].

In previous work we developed a stochastic network model of aging with damage accumulation [11, 44]. Each individual is modeled as a network of interacting nodes that represent health attributes. Both the nodes and their connections are idealized and do not specify particular health aspects or mechanisms. Connections (links) between neighboring nodes in the network can be interpreted as influence between separate physiological systems. In our model, damage facilitates subsequent damage of connected nodes. We do not specify the biological mechanisms that cause damage, only that damage rates depend on the proportion of damaged neighbors. Damage promotes more damage and lack of damage facilitates repair. Rather than model the specific biological mechanisms of aging, we model how damage to components of generic physiological systems can accumulate and propagate throughout an organism — ending with death.

Even though our model includes no explicit age-dependence in damage rates or mortality, it captures Gompertz's law of mortality [28, 135], the average rate of FI

accumulation [14, 66], and the broadening of FI distributions with age [64, 65]. By including a false-negative attribution error (i.e. a finite sensitivity) [44], we can also explain an empirical maximum of observed FI values – typically between $0.6 - 0.8$ [15, 64–68]. This shows that age-dependent "programming" of either mortality or damage rates are not necessary to explain these features [52].

We had chosen the Barabási-Albert (BA) preferential attachment algorithm [98] to generate our scale-free network, both due to the simplicity of the BA algorithm and due to the numerous examples of these scale-free networks in biological systems [136]. While we had constrained the scale-free network parameters with the available phenomenology, we did not examine whether other common network structures could also recover the same phenomenology. More specifically, we did not identify which observable behavior sensitively depends on the network structure.

Ideally, we could directly reconstruct the network from available data. However, the direct assessment of node connectivity from observational data is a challenging and generally unsolved problem. Nevertheless, we show here that we can reliably reconstruct the relative connectivity (i.e. the rank-order) of high degree nodes in both model and in large-cohort observational data by measuring mutual dependence between pairs of nodes. This reconstruction allows us to qualitatively confirm the relationship between the connectivity of nodes and how informative they are about mortality [44]. Specifically, we demonstrate that a network with a wide range of node connectivities (such as a scale-free network) is needed to describe the observational data.

Recently, the FI approach has been extended to laboratory [70] and biomarker data [71] and used in clinical [137, 138] and population settings [72]. Two different FIs have been constructed to measure different types of damage, F_{clin}, with clinically evaluated or self-reported data, and F_{lab}, with lab or biomarker data. Clinical deficits are typically based on disabilities, loss of function, or diagnosis of disease, and they measure clinically observable damage that typically occurs late in life. Lab deficits or biomarkers use the results of lab tests (e.g. blood tests or vital signs) that are binarized using standard reference ranges [139]. Since frailty indices based on laboratory tests measure pre-clinical damage, they are distinct from those based on clinical and/or self-report data [70, 72].

Even though they measure very different types of damage, both FIs are similarly associated with mortality [70,140]. Earlier observational studies have found (average) $\langle F_{\text{lab}} \rangle$ larger than $\langle F_{\text{clin}} \rangle$ [70,71,140]. However, a study of older long-term care patients has found $\langle F_{\text{lab}} \rangle$ less than $\langle F_{\text{clin}} \rangle$ [141]. While differences between studies could be attributed to classification differences, a large single study including ages from 20-85 from the National Health and Nutrition Examination Survey (NHANES) [72] also found that $\langle F_{\text{lab}} \rangle$ was higher than $\langle F_{\text{clin}} \rangle$ at earlier ages, but below at later ages.

The observed age-dependent relationship (or "age-structure") between F_{lab} and F_{clin} challenges us to examine whether network properties can determine similar age-structure in model data. We aim to determine what qualitative network features are necessary to explain age-structure. Our working hypothesis is that low-degree nodes should correspond to F_{lab}, just as high-degree nodes correspond to F_{clin} [11, 44].

Complex networks have structural features beyond the degree distribution. For example, nearest-neighbor degree correlations describe how connections are made between specific nodes of different degree [142]. Accordingly, we consider networks with three types of degree correlations: assortative, disassortative, and neutral [142, 143]. Networks with assortative correlations tend to connect like-degree nodes, those with disassortative correlations tend to connect unlike-degrees, and those with neutral correlations are random. We probe and understand the internal structure of these networks by examining F_{high} and F_{low}, i.e. damage to high degree nodes and damage to low degree nodes.

Since networks have many properties other than degree distribution and nearest-neighbor degree correlations, we have also constructed a mean-field theory that *only* has these properties. With it we can better connect specific network properties with qualitatively observed phenomenon, within the context of our network model.

We show how network properties of degree distribution and degree correlations are essential for our model to recover results from observational data. Doing so, we can explain how damage propagates through our network and what makes nodes informative of mortality. This allows us to understand the differences between F_{low} and F_{high}, or between pre-clinical and clinical damage in observational health data.

3.3 Methods

3.3.1 Stochastic model

Our model was previously presented [44]. Individuals are represented as a network consisting of N nodes, where each node $i \in \{1, 2, \ldots, N\}$ can take on binary values $d_i \in \{0, 1\}$ for healthy or damaged, respectively. Connections are undirected and all nodes are undamaged at time $t = 0$.

A stochastic process transitions between healthy and damaged ($d_i = 0$ and $d_i = 1$) states. Healthy nodes damage with rate $\Gamma_+ = \Gamma_0 \exp(f_i \gamma_+)$ and damaged nodes repair with rate $\Gamma_- = (\Gamma_0/R) \exp(-f_i \gamma_-)$. These rates depend on the local frailty $f_i = \sum_{j \in \mathcal{N}(i)} d_j/k_i$, which is the proportion of damaged neighbors of node i. This local frailty f_i quantifies local damage within the network. Transitions between the damaged and healthy states of nodes are implemented exactly using a stochastic simulation algorithm [144, 145]. For each step, the algorithm samples the time until the next transition from an exponential waiting-time distribution with rate equal to the sum of all transition rates. The particular transition (i.e. which node) performed is sampled from the set of all possible transitions. The probability of choosing a particular transition is determined by its transition rate. Individual mortality occurs when the two highest degree nodes are both damaged.

We generate our default network "topology" using a linearly-shifted preferential attachment algorithm [99, 100], which is a generalization of the original Barabási-Albert algorithm [98]. This generates a scale-free network $P(k) \sim k^{-\alpha}$, where the exponent α and average degree $\langle k \rangle$ can be tuned. (The minimum degree varies as $k_{\min} = \langle k \rangle/2$.) This network is highly heterogeneous in both degree k_i and nearest-neighbor degree (nn-degree) $k_{i,\mathrm{nn}} = \sum_{j \in \mathcal{N}(i)} k_j/k_i = \sum_{j,l} a_{ij} a_{lj}/k_i$.

Since we are concerned with the properties of individual nodes and groups of nodes, we use the same randomly generated network for all individuals. As a result, connections between any two nodes are the same for every individual. To ensure that our randomly generated network is generic, we then redo all of our analysis for 10 different randomly generated networks. All of these networks behave qualitatively the same, and so we present results averaged over them. Previously [44], we generated a distinct network realization for each individual.

We have used observational data for mortality rate and FI vs age to fine-tune the network parameters [11, 44]. A systematic exploration of parameters was done in previous work [11, 44]. Most of our parameterization ($N = 10000$, $\alpha = 2.27$, $\langle k \rangle = 4$, $\gamma_- = 6.5$) is the same as reported previously [44]. However, three parameters ($\Gamma_0 = 0.00183$/yr, $\gamma_+ = 7.5$, $R = 3$) have been adjusted because we now disallow multiple connections between pairs of nodes during our network generation. This simplifies analysis and adjustment of the network topology, but would also affect mortality rates (see e.g. Fig. 3.15 below) without the parameter adjustment. Other network topologies, see Sect. 3.4.4, also use this "default" parameterization unless otherwise noted.

Typically, binary deficits have a finite sensitivity [146], while our model gives us exact knowledge of when a node damages. We have modeled this finite sensitivity by applying non-zero false-negative attribution errors to our raw model FI [44]. This has no effect on the dynamics or on mortality, but does affect the observed FI scores. For any raw FI $f_0 = \sum_i d_i/n$ from n nodes, there are $n_0 = f_0 n$ damaged nodes. With a false-negative rate of q, n_q of these are overturned, where n_q is individually-sampled from a binomial distribution $p(n_q; n_0, 1-q) = \binom{n_0}{n_q}(1-q)^{n_q} q^{n_0 - n_q}$. We use $f = n_q/n$ as the corrected individual FI. Since our model f_0 tends to reach the arithmetic maximum of 1 at old ages, this effectively gives a maximum observed FI of $\langle f_{\max} \rangle = 1 - q$ [44]. We use $q = 0.4$ throughout.

3.3.2 Observational Data analysis

Observational data is typically "censored", meaning that the study ended or an individual dropped out before their death occurred, leaving no known death age. To avoid this problem, we use a binary mortality outcome e.g. $M = 0$ if an individual is alive within 5 years of follow-up, or $M = 1$ otherwise. We use 5 year outcomes throughout for observational data unless otherwise specified. We adapt this approach in our analysis of mutual information [147, 148]. Our entropy calculations will use binary entropy, $H(M|t) = -p(0|t) \log p(0|t) - H(1|t) \log p(1|t)$, which we use to calculate information $I(M; D_i|t) = H(M|t) - H(M|D_i, t)$. See also Blokh and Stambler [149], for other varieties of information analysis for observational data.

We compare our information theory results to a more standard survival analysis

with hazard ratios [150]. The hazard ratio is the ratio of instantaneous event rates for two values of an explanatory variable — e.g. with/without a deficit. A larger hazard ratio means a lower likelihood of surviving with the deficit than without. Hazard ratios are "semi-parametric", since they extract the effects of variables on mortality rate from a phenomenological mortality model. We use the Cox proportional hazards model [119]. We show below that these survival analysis techniques are consistent with our non-parametric mutual information measures.

3.3.3 High-k network reconstruction

To reconstruct network connections from observed states of nodes, we use the state of each deficit (node) at a given age t (or narrow range of ages in observational data) for each individual in the sample, and calculate the mutual information between individual deficits, $I(D_i; D_j|t)$ [151,152]. Connections in the model create correlations between nodes, so a large $I(D_i; D_j|t)$ could indicate a connection. We use data where individuals are the same age (or $\pm$ 5 years in observational data), so that time is not a confounding variable. Nevertheless, determining whether a given connection exists or not requires a threshold on $I(D_i; D_j|t)$. If we took this route, we would only assign a connection between nodes if the mutual information is above this threshold. However, we have no practical way of determining such a threshold, though attempts have been made in the past [153].

In preliminary tests with our model we have found that matching the reconstructed average degree with the exact average degree is a reliable way of determining a threshold (data not shown), but we still have no way of determining the average degree from observational data. Instead, we use a simple parameter-free method adapted from work on gene co-expression networks [154]. We construct weighted networks, with the mutual information between pairs of nodes as the strength or weight of the connections. We then calculate a "reconstructed" degree by adding the information for each possible connection to the node in the network, $\hat{k}_i \equiv \sum_{j \neq i} I(D_i; D_j|t)$ [155]. For nodes that aren't connected, $I(D_i; D_j|t) \approx 0$, while $I(D_i; D_j|t)$ is expected to be large for connected nodes. While we cannot reconstruct the actual network, we can reconstruct the rank-order degree of high-k nodes – since we find that $\hat{k}$ is roughly monotonic with the actual degree k for high-k nodes.

3.3.4 Mean-field theory of network dynamics

Here, we present a mean-field theory of our network model to understand the mechanisms underlying our model results. Our mean-field theory (MFT) is based on work on epidemic processes in complex networks by Pastor-Satorras *et al.* [156] together with ideas from Gleeson [157] that we use to incorporate mortality dynamics.

By MFT we mean a set of deterministic dynamical equations for damage probabilities of network nodes, including mortality nodes. Here, we retain the full degree distribution $P(k)$ and degree correlations $P(k'|k)$ of our stochastic network model, but drop higher-order network correlations. This allows us to identify what model behavior is controlled by the degree distribution and degree correlations. (A simpler MFT, with all nodes having the same degree, has been published [44].) With a degree distribution we then solve (see below) thousands of coupled ordinary differential equations (ODEs) with standard numerical integrators.

Instead of treating each node individually, we assume nodes of the same degree behave the same. To do this, we average the damaged probabilities $p(d_i = 1, t)$ and the undamaged probabilities $p(d_i = 0, t)$, conditioned on the damage of the mortality nodes, over all nodes of the same degree k:

$$p_{k,d_{m_1},d_{m_2}}(t) \equiv \sum_{\deg(i)=k} p(d_i = 1, d_{m_1}, d_{m_2}, t)/(NP(k)),$$
$$q_{k,d_{m_1},d_{m_2}}(t) \equiv \sum_{\deg(i)=k} p(d_i = 0, d_{m_1}, d_{m_2}, t)/(NP(k)),$$

where the mortality states are indicated by $d_{m_1}, d_{m_2} \in \{0, 1\}$, N is the number of nodes, and $P(k)$ is the degree distribution. The resulting joint probabilities are $p_{k,d_{m_1},d_{m_2}}$ and $q_{k,d_{m_1},d_{m_2}}$, for damaged and undamaged nodes respectively. These joint probabilities satisfy

$$\sum_{d_{m_1},d_{m_2}} (p_{k,d_{m_1},d_{m_2}} + q_{k,d_{m_1},d_{m_2}}) = 1, \tag{3.1}$$
$$p_{d_{m_1},d_{m_2}} = p_{k,d_{m_1},d_{m_2}} + q_{k,d_{m_1},d_{m_2}}, \text{ and} \tag{3.2}$$
$$p_{k|d_{m_1},d_{m_2}} = p_{k,d_{m_1},d_{m_2}}/p_{d_{m_1},d_{m_2}}, \tag{3.3}$$

where the first equation is a normalization condition, the second completeness, and the third Bayes' theorem for conditional probabilities. From our mortality rule of $d_{m_1}, d_{m_2} = 1$, the probability of mortality is $p_{\text{dead}} = p_{k,1,1} + q_{k,1,1}$, for any k.

The probability of a neighbor of a node of degree k being damaged (which is its local frailty f) given a particular mortality state is

$$f_{k|d_{m_1},d_{m_2}}(t) = \sum_{k'} P(k'|k) p_{k'|d_{m_1},d_{m_2}}, \tag{3.4}$$

where $P(k'|k)$ is the conditional degree distribution, or "nearest-neighbor" degree distribution. $P(k'|k)$ describes the structure of connections in the network, and can be varied independently of the degree distribution $P(k)$.

Writing exact master equations for N nodes is impractical since there would be 2^N distinct states to track, with even more distinct transition rates. As an enormous simplification, we use averaged damage and repair rates of nodes of a given connectivity k. This is our key mean-field simplification. To do this we approximate $\langle d_i d_j \rangle = \langle d_i \rangle \langle d_j \rangle$ for all nodes, and approximate the number of damaged neighbors by a binomial distribution $n_d \sim B(n_d; f_{k|d_{m_1},d_{m_2}}, k) = \binom{k}{n_d} f_{k|d_{m_1},d_{m_2}}^{n_d} (1 - f_{k|d_{m_1},d_{m_2}})^{k-n_d}$ where the average proportion of damaged neighbors will be $f_{k|d_{m_1},d_{m_2}}$. Using Eq. 3.4, we can then calculate our MFT damage and repair rates,

$$\begin{aligned} \langle \Gamma_{\pm}(f_{k|d_{m_1},d_{m_2}}) \rangle &= \Gamma_{0,\pm} \Big\langle \exp\left(\gamma_{\pm} n_d / k\right) \Big\rangle \\ &= \Gamma_{0,\pm} \Big(f_{k|d_{m_1},d_{m_2}} e^{\pm \gamma_{\pm}/k} + 1 - f_{k|d_{m_1},d_{m_2}} \Big)^k. \end{aligned} \tag{3.5}$$

The node degree is explicit in Eq. 3.5, while the degree correlation is included through the average local damage in Eq. 3.4.

Using these averaged damage/repair rates as transition probabilities, we can write a master equation for nodes with connectivity $k = k_{\min}, ..., k_{m_2} - 1$ and given the global

state of the mortality nodes:

$$\begin{aligned}
\dot{p}_{k,0,0}(t) &= q_{k,0,0}\langle\Gamma_+(f_k)\rangle - p_{k,0,0}\Big[\langle\Gamma_+(f_{m_1})\rangle + \langle\Gamma_+(f_{m_2})\rangle\Big] && (3.6)\\
&\quad -p_{k,0,0}\langle\Gamma_-(f_k)\rangle + p_{k,1,0}\langle\Gamma_-(f_{m_1})\rangle + p_{k,0,1}\langle\Gamma_-(f_{m_2})\rangle \\
\dot{q}_{k,0,0}(t) &= -q_{k,0,0}\langle\Gamma_+(f_k)\rangle - q_{k,0,0}\Big[\langle\Gamma_+(f_{m_1})\rangle + \langle\Gamma_+(f_{m_2})\rangle\Big] && (3.7)\\
&\quad +p_{k,0,0}\langle\Gamma_-(f_k)\rangle + q_{k,1,0}\langle\Gamma_-(f_{m_1})\rangle + q_{k,0,1}\langle\Gamma_-(f_{m_2})\rangle \\
\dot{p}_{k,1,0}(t) &= q_{k,1,0}\langle\Gamma_+(f_k)\rangle - p_{k,1,0}\langle\Gamma_+(f_{m_2})\rangle + p_{k,0,0}\langle\Gamma_+(f_{m_1})\rangle && (3.8)\\
&\quad -p_{k,1,0}\langle\Gamma_-(f_k)\rangle - p_{k,1,0}\langle\Gamma_-(f_{m_1})\rangle \\
\dot{q}_{k,1,0}(t) &= -q_{k,1,0}\langle\Gamma_+(f_k)\rangle - q_{k,1,0}\langle\Gamma_+(f_{m_2})\rangle + q_{k,0,0}\langle\Gamma_+(f_{m_1})\rangle && (3.9)\\
&\quad +q_{k,1,0}\langle\Gamma_-(f_k)\rangle - q_{k,1,0}\langle\Gamma_-(f_{m_1})\rangle \\
\dot{p}_{k,0,1}(t) &= q_{k,0,1}\langle\Gamma_+(f_k)\rangle - p_{k,0,1}\langle\Gamma_+(f_{m_1})\rangle + p_{k,0,0}\langle\Gamma_+(f_{m_2})\rangle && (3.10)\\
&\quad -p_{k,0,1}\langle\Gamma_-(f_k)\rangle - p_{k,0,1}\langle\Gamma_-(f_{m_2})\rangle \\
\dot{q}_{k,0,1}(t) &= -q_{k,0,1}\langle\Gamma_+(f_k)\rangle - q_{k,0,1}\langle\Gamma_+(f_{m_1})\rangle + q_{k,0,0}\langle\Gamma_+(f_{m_2})\rangle && (3.11)\\
&\quad +p_{k,0,1}\langle\Gamma_-(f_k)\rangle - q_{k,0,1}\langle\Gamma_-(f_{m_2})\rangle \\
\dot{p}_{k,1,1}(t) &= p_{k,1,0}\langle\Gamma_+(f_{m_2})\rangle + p_{k,0,1}\langle\Gamma_+(f_{m_1})\rangle && (3.12)\\
\dot{q}_{k,1,1}(t) &= q_{k,1,0}\langle\Gamma_+(f_{m_2})\rangle + q_{k,0,1}\langle\Gamma_+(f_{m_1})\rangle. && (3.13)
\end{aligned}$$

In these equations we have not shown the mortality state indices of f_k for readability, but they are the same as the associated p or q factors. We have also defined f_{m_1} and f_{m_2} as the local frailties of the first and second mortality node, respectively. We have 8 equations for each distinct degree k. The last two equations determine the mortality rate, $\dot{p}_{k,1,1} + \dot{q}_{k,1,1}$.

The mean-field model couples the dynamics of the lowest degree ($k = 2$) with all degrees up to the two highest (mortality nodes). Solving the equations requires us to explicitly determine the two mortality node degrees. While approximate calculations of the maximum degree of scale-free networks are available [158], we need the *two* highest degrees. We use $k_{m_1} = 885$ and $k_{m_2} = 768$, based on the averages from simulations of the network. Similarly, we use $k_{m_1} = 14$ and $k_{m_2} = 13$ for ER random networks and $k_{m_1} = 7$ and $k_{m_2} = 6$ for WS small-world networks. Qualitatively, our qualitative MFT results do not depend on these mortality node degrees, as long as they are sufficiently large. The minimum degree $k_{\min}$ is determined by the network topology.

Our default model uses a linearly-shifted preferential-attachment model, which has explicit functional forms for the degree distribution $P(k)$ and the nearest-neighbor degree distribution $P(k'|k)$ as $N \to \infty$ [100].

We numerically solve Eq. 3.13 for the probabilities $p_{k,d_{m_1},d_{m_2}}(t)$ and $q_{k,d_{m_1},d_{m_2}}(t)$. These then allow us to calculate the average FI,

$$\begin{aligned} \langle F(t) \rangle &= \frac{\sum_{k=k_{\text{low}}}^{k_{\text{high}}} P(k) p_{k|\text{alive}}}{\sum_{k=k_{\text{low}}}^{k_{\text{high}}} P(k)}, \\ p_{k|\text{alive}} &\equiv \frac{p_{k,0,0} + p_{k,0,1} + p_{k,1,0}}{p_{k,0,0} + p_{k,0,1} + p_{k,1,0} + q_{k,0,0} + q_{k,0,1} + q_{k,1,0}}, \end{aligned} \tag{3.14}$$

so that the average is over the surviving individuals. Our averaged damage-rates overestimate the true values, so for the same parameterization mortality occurs on a shorter timescale in the MFT. This is because rapidly damaging nodes drop out of the full model once they are damaged, but continue to contribute to the average damage rates in the mean-field model through Eq. 3.5. Because of this, when plotting MFT results we scale time by t_{scale}, the time at which every node is damaged ($p_k = 1$).

3.4 Results

We will focus on measures that can be compared between model and observational data, or that provide insight into the network structure of organismal aging. We start with observational data, to expand the observed aging phenomenology. Then we explore how our network model behaves, with a focus on how network properties determine the qualitative behavior of the model.

3.4.1 Observational Data

Dauntingly, we have three challenges for assessing network properties from observational data: human studies are small (typically with $\lesssim 10^4$ individuals) so that results will be noisy, different studies will have quantitative differences due to cohort differences and choices of measured health attributes, and we have no robust way of reconstructing networks from observed deficits so that the absolute connectivity of health-attributes is unknown. We face these challenges by focusing on qualitatively

robust behavior from larger observational studies; this will also help us to confront our results with the behavior of our generic network model.

From the American National Health and Nutrition Examination Survey (NHANES, see [159]), the 2003-2004 and 2005-2006 cohorts were combined, with up to 5 years of mortality reporting. This cross-sectional data includes one observation of age and health deficits, with either age of death or last age known to be still alive. Laboratory data were available for 9052 individuals and clinical data on 10004, aged 20+ years. Thresholds used to binarize lab deficits are found in [72]. From the Canadian Study of Health and Aging (CSHA, see [160]), 5 year mortality reporting are obtained from 1996/1997. This data is also cross-sectional, with one observation of age and health deficits, and age of death or last known age alive. Laboratory data were available for 1013 individuals and clinical data for 8547, aged 65+ years. Thresholds used to binarize lab deficits are found in [70]. By approaching both the NHANES and CSHA studies with the same approaches, we can identify qualitatively robust features of both.

Fig. 3.1 shows the average FI vs age for F_{lab} in red and F_{clin} in blue for the NHANES in the main plot and CSHA in the inset. In both studies lab deficits accumulate earlier than clinical deficits. A crossover appears in the NHANES data around age 55 after which clinical deficits are more damaged than lab deficits. A similar crossover does not appear to happen in the CSHA data.

Figs. 3.2 and 3.3 show deficits rank-ordered in information $I(M; D_i|t)$ for the NHANES and CSHA studies, respectively. These are information "fingerprints". Red points correspond to lab deficits and blue to clinical deficits, as indicated. Both types of deficits have similar magnitudes of information, although clinical deficits are typically more informative. The comparable magnitudes of mutual information for the majority of individual deficits between lab and clinical FIs is consistent with earlier analysis that found similar association between lab and clinical FIs with mortality using survival analysis [70, 72, 140].

Insets in Figs. 3.2 and 3.3 show the corresponding hazard ratio (HR) for the deficit found from a Cox proportional hazards model regression, with the deficit value and age used as covariates. This semi-parametric analysis is often done with medical data [161]. The HR tends to increase as the rank-ordered information increases, indicating

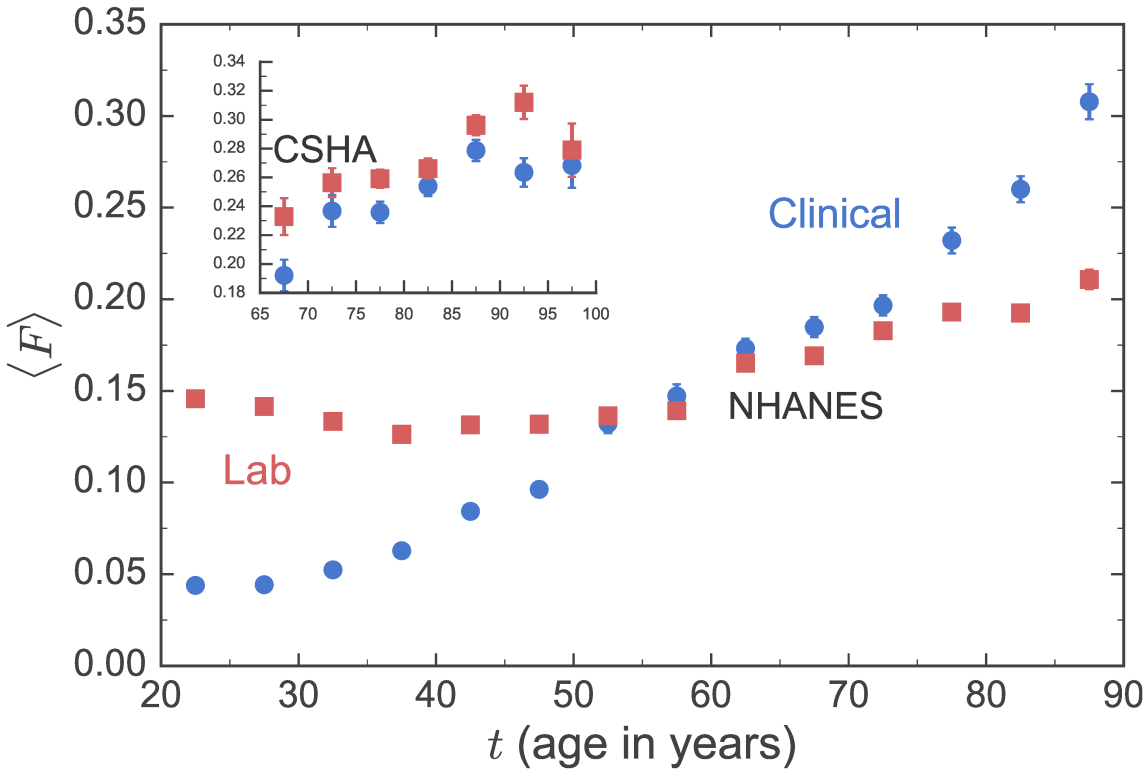


Figure 3.1: Average FI vs age t with $\langle F_{\text{lab}} \rangle$ (red squares) and $\langle F_{\text{clin}} \rangle$ (blue circles) from the NHANES dataset (main figure). The inset shows the same plot for the CSHA dataset. Error bars show the standard error of the mean. All individuals used in this plot have both F_{clin} and F_{lab} measured.

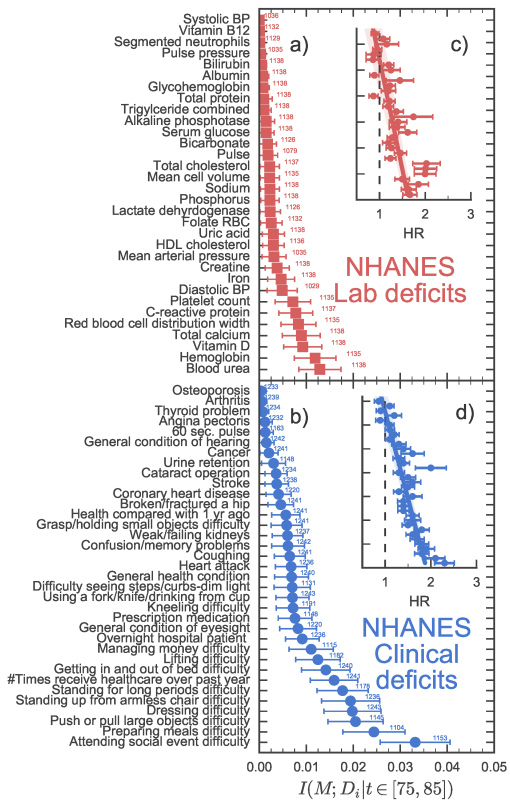


Figure 3.2: Rank-ordered deficits in terms of information $I(M; D_i|t \in [75, 85])$ for the NHANES dataset. Red points are lab deficits, blue points are clinical deficits. Error bars are standard errors found from bootstrap resampling. Small numbers next to the points indicate the number of individuals that were available in the data for the corresponding deficit. Insets show the corresponding hazard ratios for the deficits found from a Cox proportional hazards model regression, with the deficit and age used as covariates. The error bars show standard errors, and the line shows a linear regression through these points with the standard error in slope and intercept shown in a lighter color.

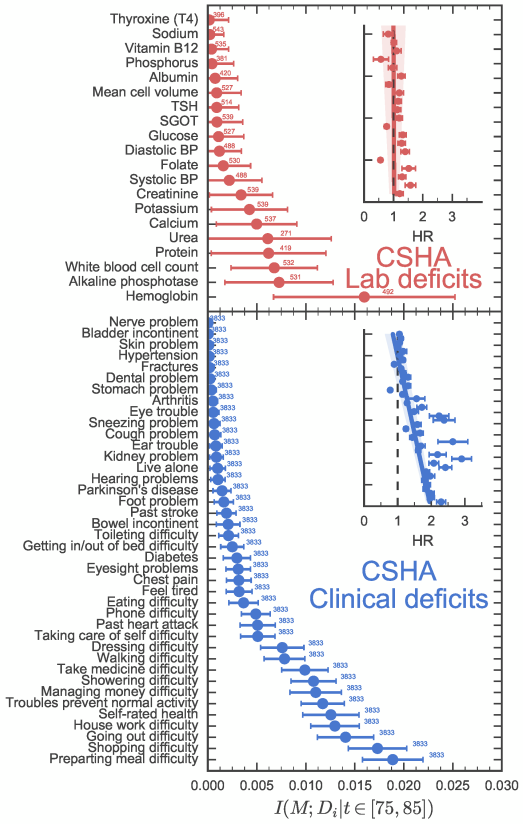


Figure 3.3: Rank-ordered deficits in terms of information $I(M; D_i|t \in [75, 85])$ for the CSHA dataset. Red points are lab deficits, blue points are clinical deficits. Error bars are standard errors found from bootstrap resampling. Small numbers next to the points indicate the number of individuals that were available in the data for the corresponding deficit. Insets show the corresponding hazard ratios for the deficits found from a Cox proportional hazards model regression, with the deficit and age used as covariates. The error bars show standard errors, and the line shows a linear regression through these points with the standard error in slope and intercept shown in a lighter color.

that our mutual-information approach is capturing similar effects. Nevertheless, we prefer mutual-information because it is non-parametric (not model-dependent) and so relies on fewer assumptions.

Our deficit-level analysis highlights the great variability of mutual information (and HR ratios) between individual deficits. We have shown that lab and clinical deficits have a range of mutual information. We further note that the top 5 - 7 most informative clinical deficits in both the NHANES and CSHA datasets measure functional disabilities or dysfunction [162]. We find that these high level deficits are the most informative of mortality, and more informative than any of the lab deficits. From this, we hypothesize that highly informative clinical deficits will also be highly connected.

We have been able to partially reconstruct the network structure of clinical measures, as detailed in Sec. 3.3.3. In Fig. 3.4, we have validated this approach with the top 32 most-connected model nodes. We use 10000 individuals for our validation, approximately the same number of people we have available in the observational studies. We know that our model information tends to increase with degree for the high degree nodes (see [44], and also Fig. 3.10 below). Fig. 3.4 shows that information also increases with the reconstructed degree $\hat{k}$, as expected for a good reconstruction. The inset showing k vs $\hat{k}$ indeed shows that the reconstructed degree is approximately monotonic with the exact degree — especially at higher k.

This means the reconstructed degree should provide a reasonable rank-order in connectivity for observational data. Nevertheless, low-degree nodes are not reliably rank-ordered. Accordingly we only attempt to reconstruct clinical $\hat{k}$ with this approach.

In Fig. 3.5, we plot information with respect to mortality $I(M; D_i|t \in [75, 85])$ for each deficit, where deficits are rank-ordered in terms of reconstructed degree $\hat{k}$. Information increases with reconstructed degree for both the NHANES and CSHA clinical data. This shows that high information deficits correspond to high connectivity in the observational data. Also, nearly all of the functional disabilities intuitively hypothesized to have a high connectivity are also found to have a large reconstructed degree.

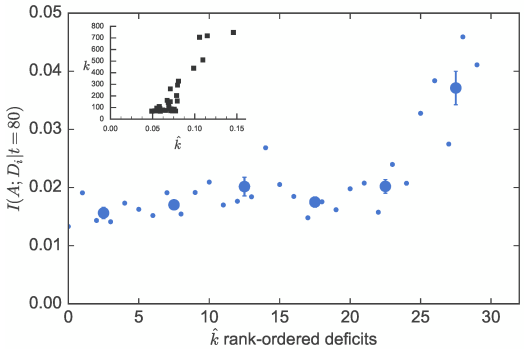


Figure 3.4: Information $I(A; D_i|t = 80)$ vs rank-ordered deficits using reconstructed degree $\hat{k}_i$, for our computational model. The top 32 most connected nodes are reconstructed with 10000 individuals. The smaller (blue circles) points show individual nodes, the larger points show a binned average, and error bars are the standard error of the mean within each bin. The inset (black squares) shows the exact degree k vs the reconstructed degree $\hat{k}$.

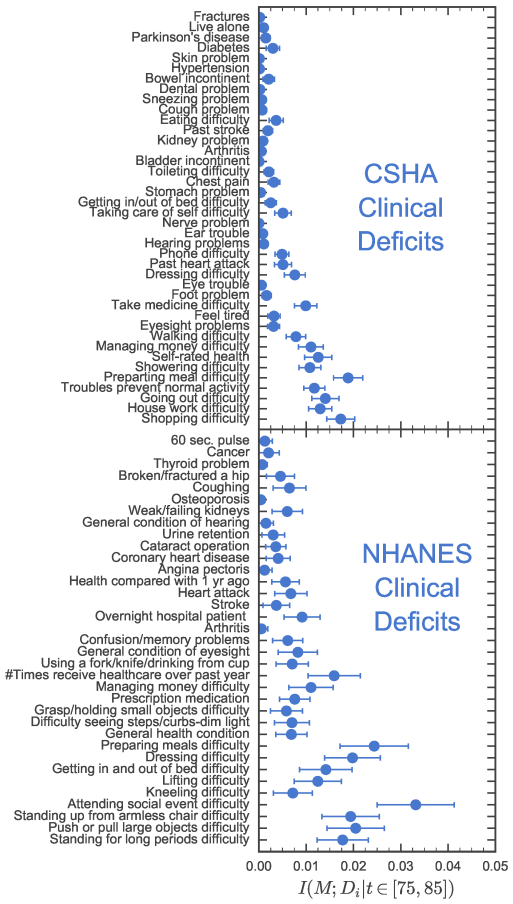


Figure 3.5: Rank-ordered clinical deficits in terms of reconstructed degree $\hat{k}$ vs information with respect to mortality $I(M; D_i|t \in [75, 85])$ for the NHANES and CSHA datasets. The reconstruction algorithm is detailed in Sec. 3.3.3. Error bars are standard errors found from bootstrap resampling.

3.4.2 Model Age-structure

We saw, in Fig. 3.1, that pre-clinical (lab) damage accumulates before clinical damage in observational data. This is a qualitatively robust observation, seen in both NHANES and CSHA observational data. We also observed, in Fig. 3.5, that (in terms of rank order) highly connected clinical deficits were more informative than less connected deficits. We expect that health-attributes assessed by laboratory tests are less connected than the high level functional attributes assessed clinically. We hypothesize that F_{lab} and F_{clin} should behave *qualitatively* like collections of low or high degree nodes, respectively, within our network model of aging.

We construct two distinct FIs to capture the difference between well-connected hub nodes and poorly connected peripheral nodes. We measure low-degree damage by constructing $F_{\text{low}} = \sum_i d_i/n$ from a random selection of $n = 32$ nodes all with $k = k_{\min} = 2$. Similarly, we measure high-degree damage with F_{high} from the top 32 most connected nodes (excluding the two most connected nodes, which are the mortality nodes).

Fig. 3.6 shows the cumulative average degree of damaged nodes $\langle k_{\text{dam}} \rangle = \langle \sum_{i=0}^{N} k_i d_i / \sum_{i=0}^{N} d_i \rangle$ vs age t. Error bars represent the standard deviation between 10 different randomly generated networks. They are each comparable to or smaller than the point size, indicating that the age-structure represents the network topology rather than a single network realization.

For a uniform network or for damage rates independent of the degree of a node, we would expect $\langle k_{\text{dam}} \rangle = \langle k \rangle$ for all ages t. However, we see the average degree of damaged deficits start at $\langle k \rangle$, with an initial decrease until around age 25 and then an increase back to $\langle k \rangle$ — implying damage does not uniformly propagate through the network.

Initially damage is purely random, so $\langle k_{\text{dam}}(0) \rangle = \langle k \rangle$. Nodes with degree $k_i < \langle k \rangle$ are being damaged when $\langle k_{\text{dam}} \rangle / \langle k \rangle$ decreases from 1, and nodes of degree $k_i > \langle k \rangle$ are being damaged when $\langle k_{\text{dam}} \rangle / \langle k \rangle$ increases towards 1.

The inset of Fig. 3.6 shows the average FI vs age for F_{low} and F_{high}. We see $\langle F_{\text{low}} \rangle$ initially larger than $\langle F_{\text{high}} \rangle$. Eventually with age, $\langle F_{\text{high}} \rangle$ increases to match $\langle F_{\text{low}} \rangle$ and even slightly exceed at very old ages. Thus, low-k nodes behave similarly to lab deficits, and high-k nodes behave similarly to clinical deficits in observational health

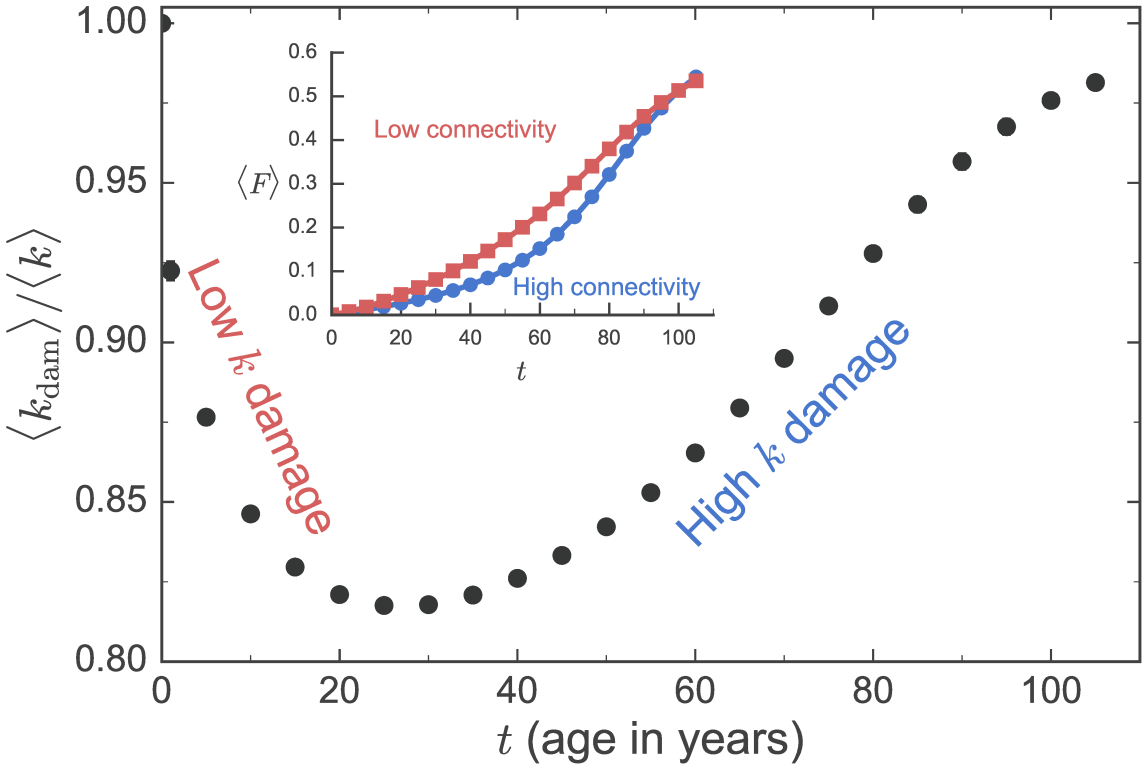


Figure 3.6: Average degree of damaged model deficits $\langle k_{\rm dam}(t) \rangle$, scaled by the average degree of the network $\langle k \rangle$, vs time t. Error bars, barely visible at low t, represent the standard deviation between randomly generated networks. As indicated, at earlier times low-connectivity nodes are preferentially damaged while at later times higher connectivity nodes are preferentially damaged. The inset shows the average damage of low-connectivity nodes $\langle F_{\rm low} \rangle$ (red squares) and of high-connectivity nodes $\langle F_{\rm high} \rangle$ (blue circles) vs age.

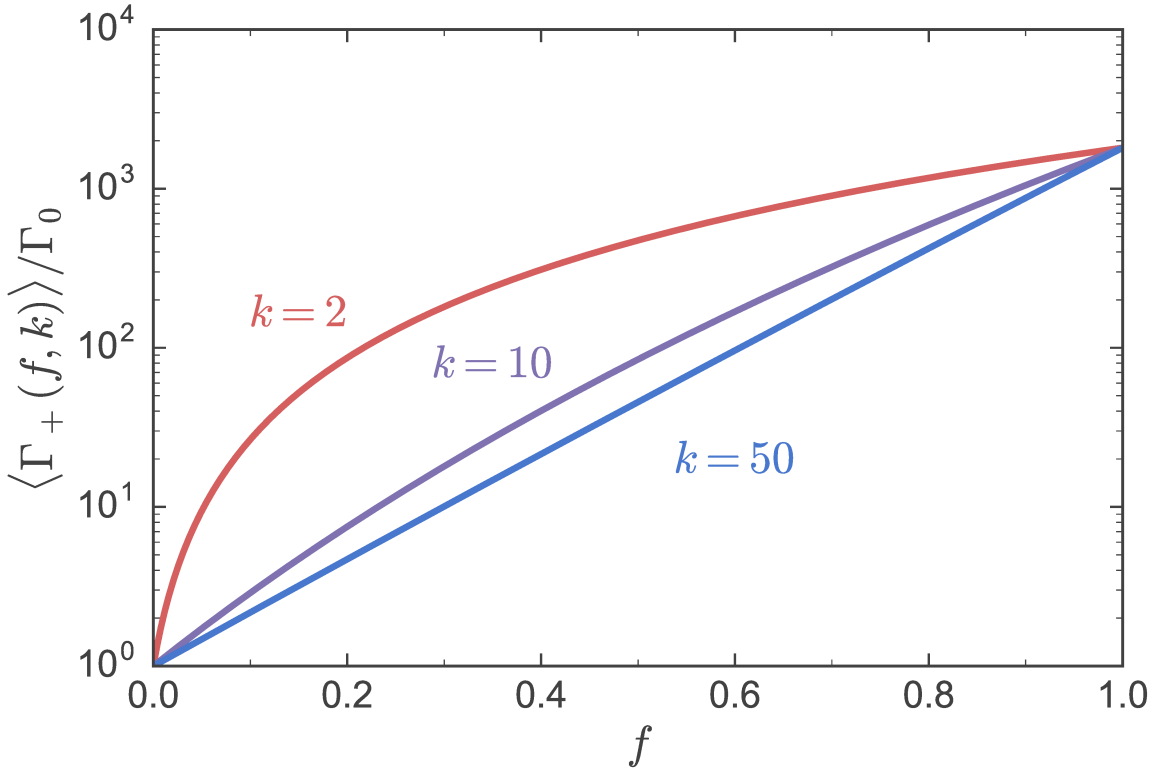


Figure 3.7: Average mean-field damage rates $\langle\Gamma_+\rangle/\Gamma_0$ for nodes of a given degree k (as indicated) vs the local frailty of these nodes f, as given by Eq. 3.5. Low-connectivity nodes exhibit significantly higher damage rates at intermediate values of f.

data. Low-k nodes and lab measures both damage early and high-k nodes and clinical measures both damage late.

We have not tuned our model parameterization to obtain this age-structure of damage in the network model. Indeed, for other parameter choices we see qualitatively similar behavior (data not shown) for the scale-free networks that we have been using. To better understand this age-structure we consider the effects of network connectivity within our mean-field theory.

In our mean-field theory, we find our averaged damage rates explicitly depend on k in Eq. 3.5. This is shown in Fig. 3.7, these mean-field damage rates increase with smaller k at a given f. This results from Jensen's inequality, since the damage rate is convex in the local frailty f and the lower degree nodes will have a broader distribution of local frailty for the same global frailty. This implies that low-k nodes should damage more frequently until they are exhausted and F_{low} saturates.

We can confirm this with the full MFT results. We can determine the FI from

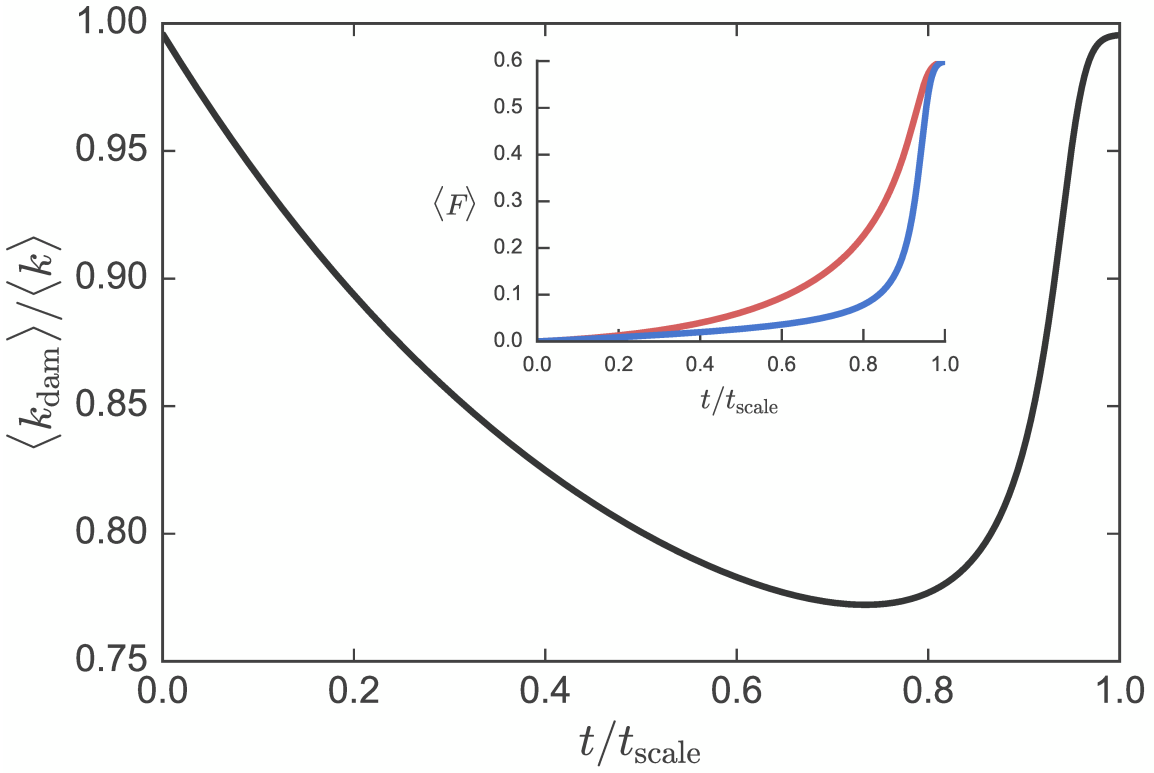


Figure 3.8: From our mean-field calculation in Sec. 3.3.4, we show the average degree of damaged deficits $\langle k_{\rm dam}\rangle$ scaled by the average network degree $\langle k\rangle$ vs time scaled by the time when the network becomes fully damaged, $t/t_{\rm scale}$. The inset shows the average damage of high connectivity nodes $\langle F_{\rm high}\rangle$ in blue and low connectivity nodes $\langle F_{\rm low}\rangle$ vs the scaled time.

Eq. 3.14 and calculate both $F_{\rm high}$ and $F_{\rm low}$ by choosing which degrees to include. The $k_{\rm low}$ and $k_{\rm high}$ determine the nodes included in the FI. For $F_{\rm high}$, we choose $k_{\rm high} = k_{m_2} - 1$ and $k_{\rm low}$ so that $N\sum_{k=k_{\rm low}}^{k_{\rm high}} P(k) = n \simeq 32$ for the smallest possible $k_{\rm low}$ (32 is the number of FI nodes typically used in our model and observational studies). For $F_{\rm low}$, we choose $k_{\rm low} = k_{\rm min}$ and choose the smallest $k_{\rm high}$ so that $n \simeq 32$.

We also calculate

$$\langle k_{\rm dam}(t)\rangle = \frac{\sum_{k=k_{\rm low}}^{k_{\rm high}} kP(k)p_{k|\rm alive}}{\sum_{k=k_{\rm low}}^{k_{\rm high}} P(k)p_{k|\rm alive}}, \tag{3.15}$$

which is the cumulative average degree of damaged nodes as was done for our computational results.

In Fig. 3.8 the age-structure from the mean-field calculation shows the same early

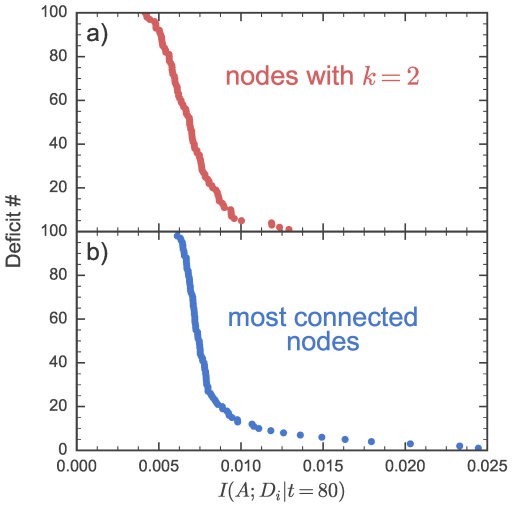


Figure 3.9: Mutual-information of selected model deficits $I(A; D_i|t = 80)$ at age 80 years, averaged over 10 randomly generated networks with 10^7 individuals each and rank-ordered. Red points are low-k deficits, blue points are high-k deficits.

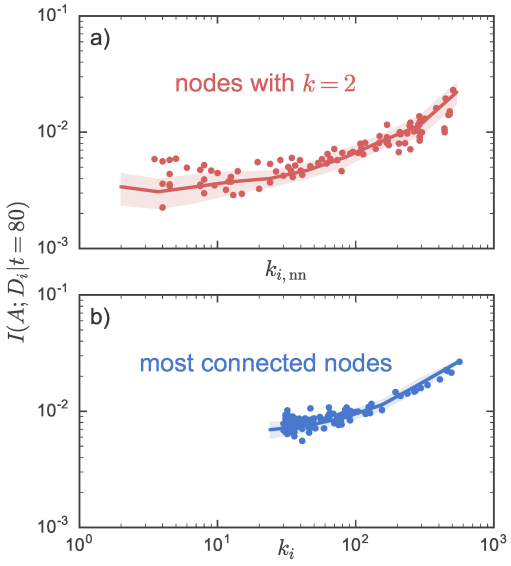


Figure 3.10: Model information spectra $I(A; D_i|t = 80)$ vs degree k_i for the top 100 most connected nodes in blue, or vs $k_{i,\mathrm{nn}}$ for a random selection of 100 peripheral nodes all with $k = k_{\min} = 2$. Points show a sample of a single network, line shows an average over 10 randomly generated networks and the random choice of 100 nodes with $k = 2$, the shaded error region shows the standard deviation over the random networks.

damage of low-k nodes shown in Fig. 3.6 and (in inset) the more-rapid growth of F_{low} compared to F_{high} at earlier times. Our mean-field calculation also shows a more-rapid growth of F_{high} compared to F_{low} at later times, as shown in the inset of Fig. 3.8. This largely is explained by the saturation of F_{low}.

We conclude that the age-structure seen observationally and in our network model, can be explained by the degree distribution and neighbor-degree correlations of our MFT. This motivates us to investigate how node degree and neighbor-degree affect mortality within the context of our network model.

3.4.3 Model Node Information

Fig. 3.9 shows the mutual information between death age and individual nodes $I(A; D_i)$ for our model. Red points are a random selection of 100 low-connectivity nodes all with $k = k_{\min} = 2$, the blue points are the top 100 most connected nodes (excluding the 2 mortality nodes). For each selection, we have rank-ordered the nodes in terms of mutual-information. The mutual-information for both high and low connectivity nodes are comparable. This is surprising since previous work showed a monotonic increase of the average information with connectivity [44]. However that work used a different network for each individual, so that network properties other than the average degree were lost by pooling nodes of the same degree.

Without parameter tuning, we obtain striking qualitative agreement of the magnitude of the mutual-information with mortality for both model and observational data (see Figs. 3.2 and 3.3). We also obtain an overlap of magnitudes of the mutual-information of low-degree and high-degree nodes that is similar to that seen between pre-clinical and clinical deficits. Since we know the model network connectivity, we can now examine what network properties cause this behavior for our model.

In Fig. 3.10, we show the "spectrum" of mutual information between death age and individual nodes $I(A; D_i|t = 80)$. We use individuals at age $t = 80$ years, where the mutual information is close to maximal [44]. We use the same network for every individual, so that we do not lose the properties of the network between individuals. For the most connected nodes, in blue, we plot mutual information vs. the connectivity of the nodes. Here we see the monotonic trend of mutual information vs connectivity, though there is significant variation for individual nodes. For the least connected nodes, in red, all of the nodes have $k = 2$. Instead of connectivity, we considered the nearest neighbor degree $k_{i,\mathrm{nn}} = \sum_{j \in \mathcal{N}(i)} k_j/k_i$ — i.e. the connectivity of the neighbors of a node. With respect to k_{nn}, we see a similar monotonic increase of the mutual information for $k = 2$ nodes.

Neighbor-connectivity k_{nn} is predictive of mortality for minimally connected nodes. We hypothesize that this is because the neighbor-connectivity affects when peripheral ($k = 2$) nodes are damaged, i.e. that peripheral nodes with low-k_{nn} are damaged earlier than those with large k_{nn}.

In the inset of Fig. 3.11 we confirm that high-k_{nn} $k = 2$ nodes damage later. This

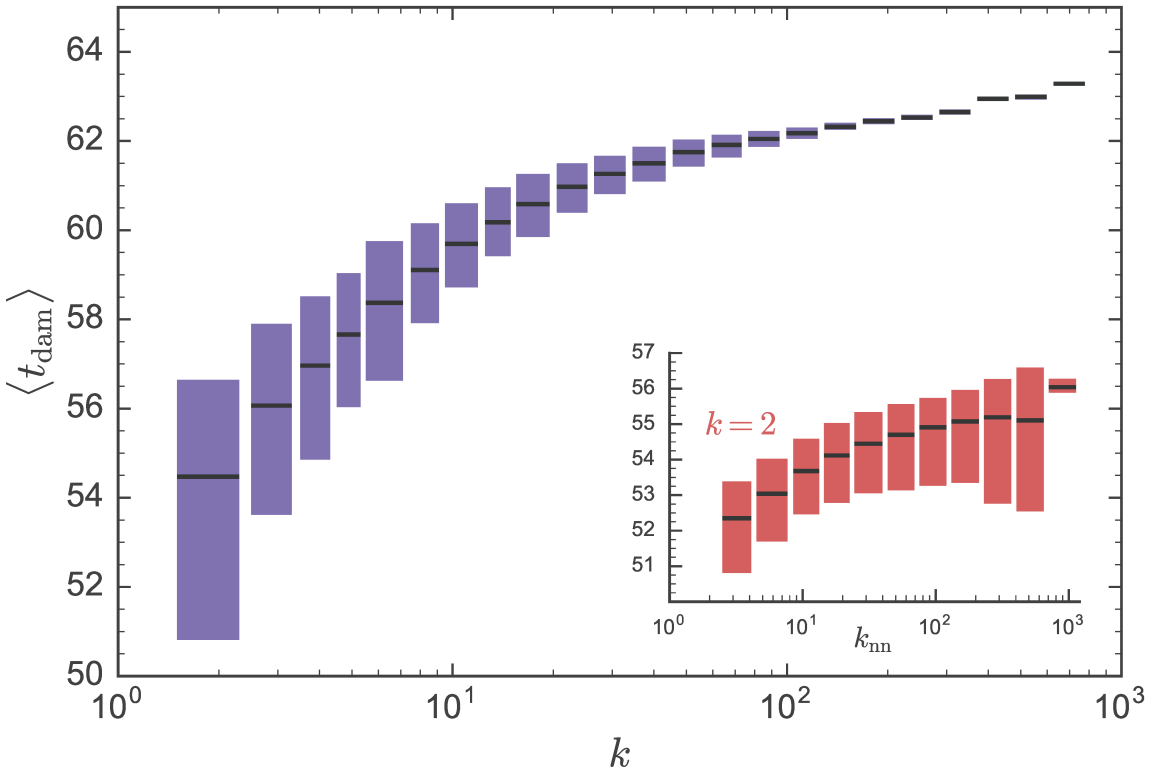


Figure 3.11: Average time of damage $\langle t_{\rm dam} \rangle$ vs degree k for all non-mortality nodes in the network. Inset shows $\langle t_{\rm dam} \rangle$ for $k = 2$ nodes vs nn-degree $k_{\rm nn}$. Nodes are binned based on k. The solid colored bars represent the entire range of average damage times observed for individual nodes within a bin, while the horizontal black lines indicate the average over the bin. All results are averaged for 10 randomly generated networks.

allows high-$k_{\rm nn}$ nodes to be informative of mortality because they are diagnostic of a more highly damaged network. From Fig. 3.11 we see that there is a large range of times for which lower-k nodes damage. Nevertheless, on average the high-$k_{\rm nn}$ nodes at $k = 2$ damage before high-k nodes even though (see Fig. 3.10) they can be similarly informative.

3.4.4 Model Network Structure

We have seen that our network model of aging is able to capture detailed behavior of lab and clinical FIs such as the the larger damage rates for low-k nodes at the same time as the surprising informativeness of some low-k nodes. The network is an important aspect of our model, and so far we have assumed that it is a preferential attachment scale-free network [98–100]. In this section, we explore the qualitative behavior of different network topologies.

Our network model has predominantly disassortative correlations (due to the scale-free exponent $\alpha < 3$ [101]) — meaning that low-k nodes tend to connect to high-k nodes, and that the average nn-degree decreases with degree [143]. We see this in Figure 3.12, where we plot the average nn-degree $\langle k_{\rm nn}(k)\rangle$ as a function of degree for our network. The purple points indicate our preferential attachment model network, and we see that the average nn-degree is inversely related to the degree.

The green curve shows a rewired assortative network [143] made by preserving the degrees of the original network but swapping links. To do this we use the method of Brunet *et al*, using N^2 rewiring iterations with a parameter $p = 0.99$ [102]. By modifying the nn-degrees of low degree nodes, we can investigate whether $k_{\rm nn}$ causes or is just correlated with informative low-k nodes. Note that we use only the largest connected component of the rewired network, with $\langle N\rangle = 9989$ nodes over 10 network realizations.

The yellow triangles in Figure 3.12 show an Erdős-Rènyi random network (ER). A random network is created by starting with N nodes, and randomly connecting each pair of nodes with probability $p_{\rm attach} = \langle k\rangle/(N-1)$ [142]. This results in a (peaked) binomial degree distribution, and completely uncorrelated connections where $k_{\rm nn} = \langle k^2\rangle/\langle k\rangle$ which is independent of individual node degree. As before, we only use the largest connected component, with $\langle N\rangle = 9805$ nodes over 10 network

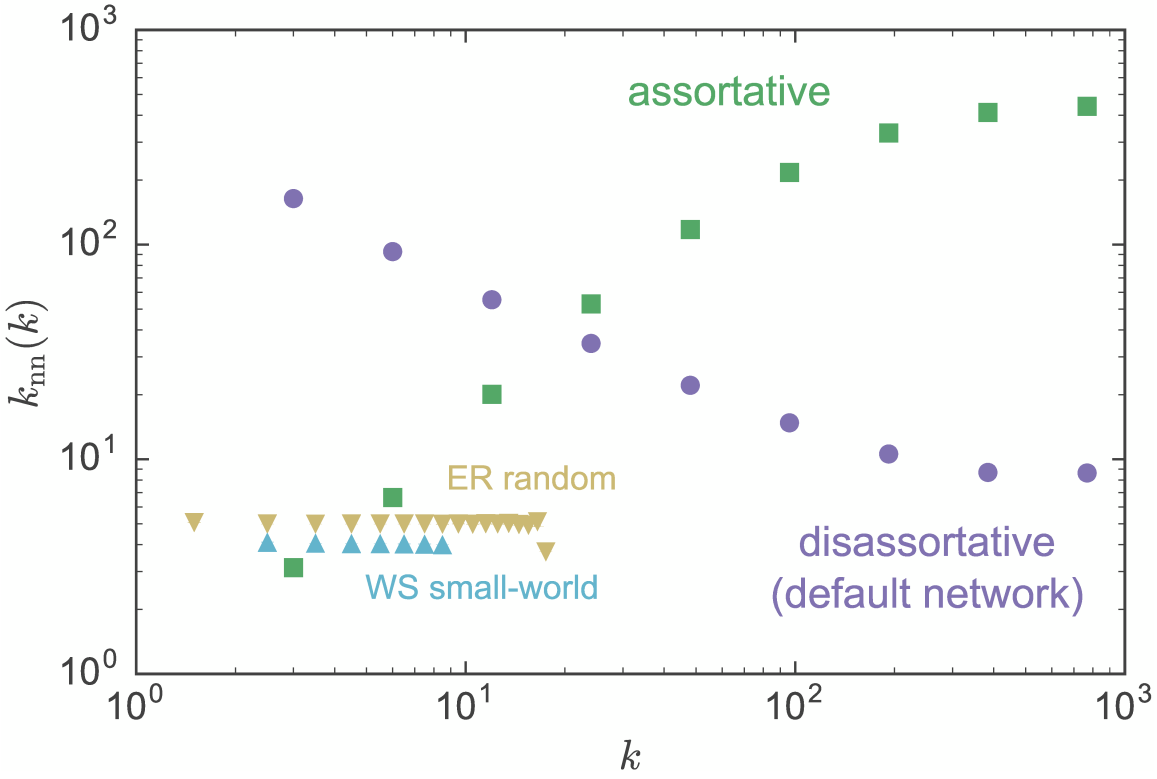


Figure 3.12: Average nn-degree $\langle k_{\rm nn}(k)\rangle$ vs degree k for a disassortative network (default network, $\alpha = 2.27$) (purple circles), an assortative network created by reshuffling the links (green squares) [102], a Erdős-Rènyi (ER) random network (yellow triangles), and a Watts-Strogatz (WS) small-world network (blue triangles). Note that $\langle k_{\rm nn}(k)\rangle$ is grouped into bins of powers of 2 and averaged within the bins for the scale-free networks. A bin for each degree is used for the ER random and WS small-world networks.

realizations. The ER network also allows us to explore whether the heavy tail of the scale-free degree distribution is required to recover our observational results.

The light blue triangles in Figure 3.12 show a Watts-Strogatz (WS) small-world network [163]. This network starts with a uniform ring network with $k_i = \langle k \rangle$ for all nodes, and randomly rewires each link with probability p_{rewire} to another randomly selected node. We use $p_{\text{rewire}} = 0.05$ to get the effects of both high clustering (i.e. links between neighbors of nodes) and short average path-lengths between arbitrary pairs of nodes [142]. This network has a narrowly peaked degree distribution, with a rapidly decaying exponential tail. ER and WS networks are similar, as both have short average path lengths between arbitrary nodes and non-heavy-tailed degree distributions, but the WS small-world network also has high clustering for small p_{rewire}.

To examine network effects on our network aging model, we have kept the same model parameters for the (default) preferential attachment disassortative network, the assortative network, the ER random network, and the WS small-world network. (The scale-free exponent α is only used in the disassortative and assortative networks.) We examine 10 random realizations of each network. We have also varied model parameters independently for each of these networks (data not shown) and obtain the same qualitative results.

In Fig. 3.13 we show rank ordered information fingerprints for individual deficits $I(A; D_i|t)$, for the different network topologies as indicated. We observe striking differences in the scale and range of the mutual information with respect to mortality, and in the differences between the most and least connected nodes. The random and small-world network both have a significantly smaller scale of mutual information, together with a much smaller range of variation.

The scale-free disassortative (default) and assortative networks both have significantly higher scale of information for the most connected nodes, as well as considerable variation (approximately 10-fold) among them. However, while the disassortative network exhibits similar scales of information between the most and least connected nodes the assortative network does not. Furthermore, the assortative network shows only minimal variation of information among its least connected nodes.

Only the disassortative (default) network exhibits the fingerprint of mutual information of the NHANES and CSHA observational studies, in Figs. 3.2 and 3.3

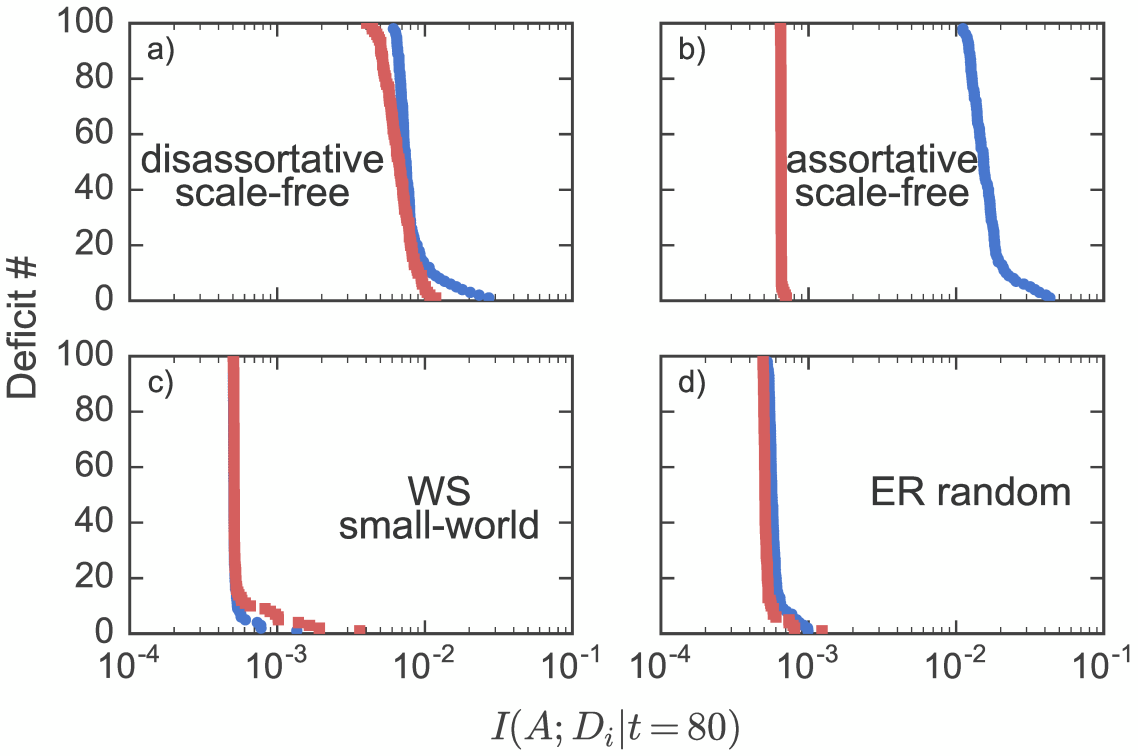


Figure 3.13: Rank ordered information $I(A; D_i|t = 80)$ for the different networks, as indicated. The top 100 most connected nodes in the network are in blue circles, and 100 randomly selected nodes of the lowest degrees are in red squares. Results for each different network topology are averaged over 10 randomly generated network realizations.

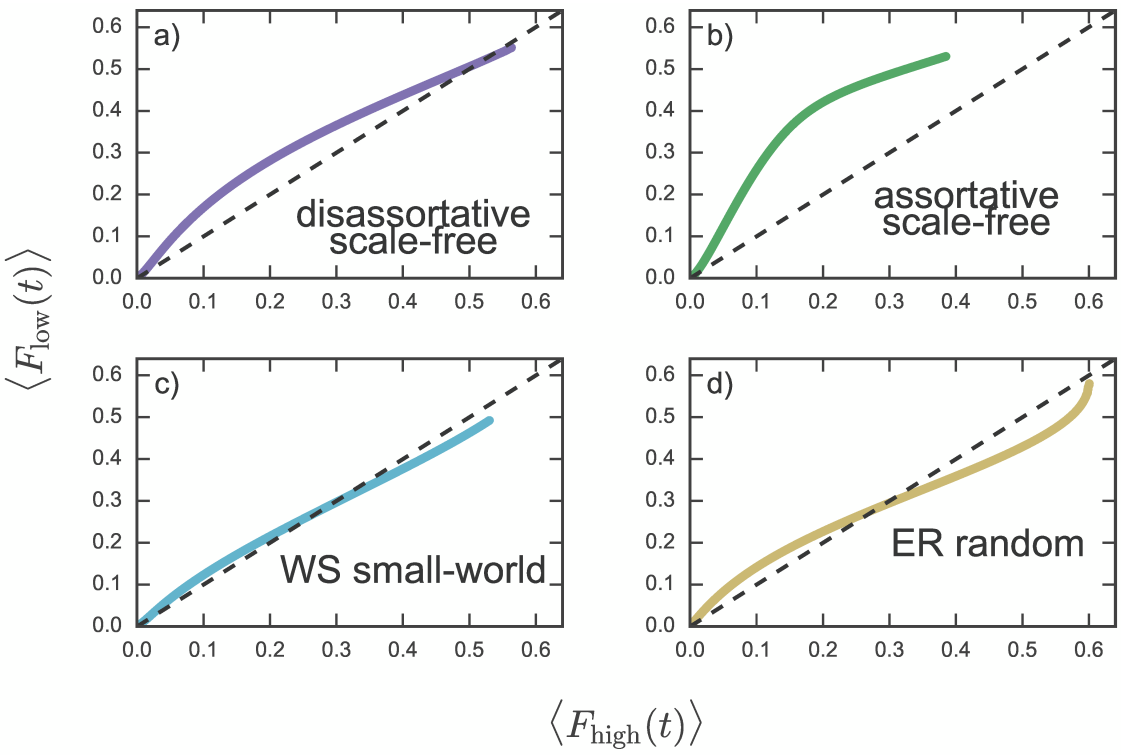


Figure 3.14: Average low-k $\langle F_{\text{low}}(t) \rangle$ vs average high-k $\langle F_{\text{high}}(t) \rangle$ plotted for $t = 0$ to $t = 110$ for our default network parameters (purple), the shuffled assortative network (green), the Erdős-Rènyi random network (yellow), and the Watts-Strogatz small world network (light blue). The dashed black line shows the line $\langle F_{\text{low}}(t) \rangle = \langle F_{\text{high}}(t) \rangle$. Results are averaged over 10 randomly generated networks and the standard deviations are smaller than the line width.

respectively: with considerable variation of mutual information between deficits, overlapping ranges between lab (low) and clinical (high) connectivity deficits, and mutual information on the order of 10^{-2} for individual deficits.

In Fig. 3.14, we investigate the age-structure of the FIs generated by the low and high connectivity nodes. We plot $\langle F_{\text{low}}(t) \rangle$ vs $\langle F_{\text{high}}(t) \rangle$ for the different network topologies. We see that the assortative network shows a rapid increase in F_{low}, followed by growth of F_{high}. In contrast, for the disassortative, random, and small-world networks there is comparable growth of both F_{low} and F_{high}, though with higher F_{low} and a later cross-over for the disassortative network.

In Fig. 3.15 we plot the average mortality rates vs age for different network topologies, with colored circles showing the computational model results and colored lines

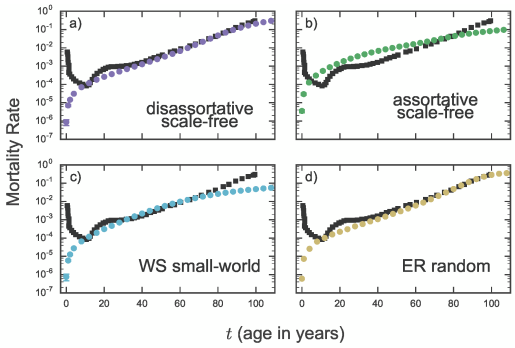


Figure 3.15: Mortality rate vs age for each of the networks. a) Disassortative scale-free network (purple circles), b) assortative scale-free network (green), c) WS small-world network (light blue), and d) ER random network (yellow). Computational results (circles) are averaged over 10 randomly generated networks and error bars show the standard deviations. Black squares are observed human mortality rates [164].

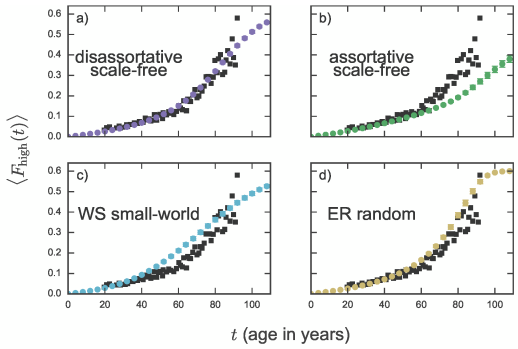


Figure 3.16: $\langle F_{\text{high}} \rangle$ vs age for each of the networks, as indicated. a) Disassortative scale-free network (purple circles), b) assortative scale-free network (green), c) WS small-world network (light blue), and d) ER random network (yellow). Computational results (circles) are averaged over 10 randomly generated networks and error bars show the standard deviations. Black squares are observed human clinical frailty [66].

for the corresponding mean-field model results. Black squares indicate observed mortality rates [164]. Similarly, in Fig. 3.16 we plot $\langle F_{\text{high}} \rangle$ vs age t for both observational data (black squares) and model data for different networks (coloured points).

Even without parameter adjustment, most of the network topologies approximately capture the observational data after $t = 20$ years. Some differences are seen, particularly for the assortative scale-free network in the mortality rate. This agreement indicates that mortality and frailty data alone do not strongly constrain the network topology.

From Fig. 3.13, we observed early damage of F_{low} in the assortative network. Our MFT allows us to narrow down what aspects of the network are leading to this behavior, since the only aspects of the network structure included are the degree distribution $P(k)$ and nn-degree correlations $P(k'|k)$.

Different network topologies are easily introduced provided $P(k)$ and $P(k'|k)$ are known. The exact $P(k)$ for our default shifted-linear preferential attachment networks [100], ER random networks, and WS small-world networks [165] are known. (We

remove zero degree nodes from the ER random degree distribution, so that $P_{k\neq 0}(k) = P(k)/\sum_{l\neq 0} P(l)$.) Using various $P(k'|k)$ we can then put different degree correlations into our MFT network. We include three types of degree correlations, uncorrelated (neutral), assortative, and disassortative [142].

For a network with uncorrelated (neutral) connections, $P(k'|k) = k'P(k')/\langle k\rangle$. We then have $k_{\rm nn}(k) = \sum_{k'} k'P(k'|k) = \langle k^2\rangle/\langle k\rangle$, so that all nodes have the same nn-degree. These correlations are used for ER random and WS small-world networks, and recover the approximately constant $k_{\rm nn}$ that we observed in Fig. 3.12.

In a network with assortative correlations, nodes tend to be connected to other nodes of similar degree. Assortative correlations that approximate those used in our computational model in Sec. 3.4.4 are [166] $P(k'|k) = \alpha\delta_{k'k} + (1-\alpha)k'P(k')/\langle k\rangle$. These lead to, $k_{\rm nn}(k) = \sum_{k'} k'P(k'|k) = \alpha k + (1-\alpha)\langle k^2\rangle/\langle k\rangle$, which increases linearly with k (see Fig. 3.12). Changing α modifies the amount of assortative correlation; we use $\alpha = 0.8$.

In a network with disassortative connections, nodes tend to be connected to other nodes of differing degree. The (disassortative) correlations for our default shifted-linear preferential attachment network are [100],

$$\begin{aligned} P(k'|k) = {} & \frac{\Gamma(k+\lambda+\alpha)\Gamma(k'+\lambda)}{k\Gamma(m+\lambda)\Gamma(k+k'+2\lambda+\alpha)} \\ & \times\Bigg[\sum_{i=m+1}^{k} \frac{\Gamma(i+m+2\lambda+\alpha-1)}{\Gamma(i+\lambda+\alpha-1)}\binom{k+k'-m-i}{k'-m} \\ & + \sum_{i=m+1}^{k'} \frac{\Gamma(i+m+2\lambda+\alpha-1)}{\Gamma(i+\lambda+\alpha-1)}\binom{k+k'-m-i}{k-m}\Bigg], \end{aligned} \tag{3.16}$$

where $m = \langle k\rangle/2 = k_{\rm min}$ and $\lambda = m(\alpha-3)$. This is exact in the limit $N\to\infty$ [100], and gives disassortative correlations where $k_{\rm nn}(k)$ decreases with k.

In Fig. 3.17 we show the average low-k FI vs the average high-k FI, $\langle F_{\rm low}(t)\rangle$ vs $\langle F_{\rm high}(t)\rangle$ from our MFT. In purple we use the (default) preferential attachment disassortative correlations, in green we use assortative correlations, and in light blue we use a WS small-world network. We see qualitative agreement with the age-structure shown in Fig. 3.14 – confirming that nn-degree correlations (included in our MFT) are important for the observed age-structure. [We have not shown MFT results for the ER random network since $\langle F_{\rm low}\rangle$ behaves poorly when it includes nodes with $k \leq 2$,

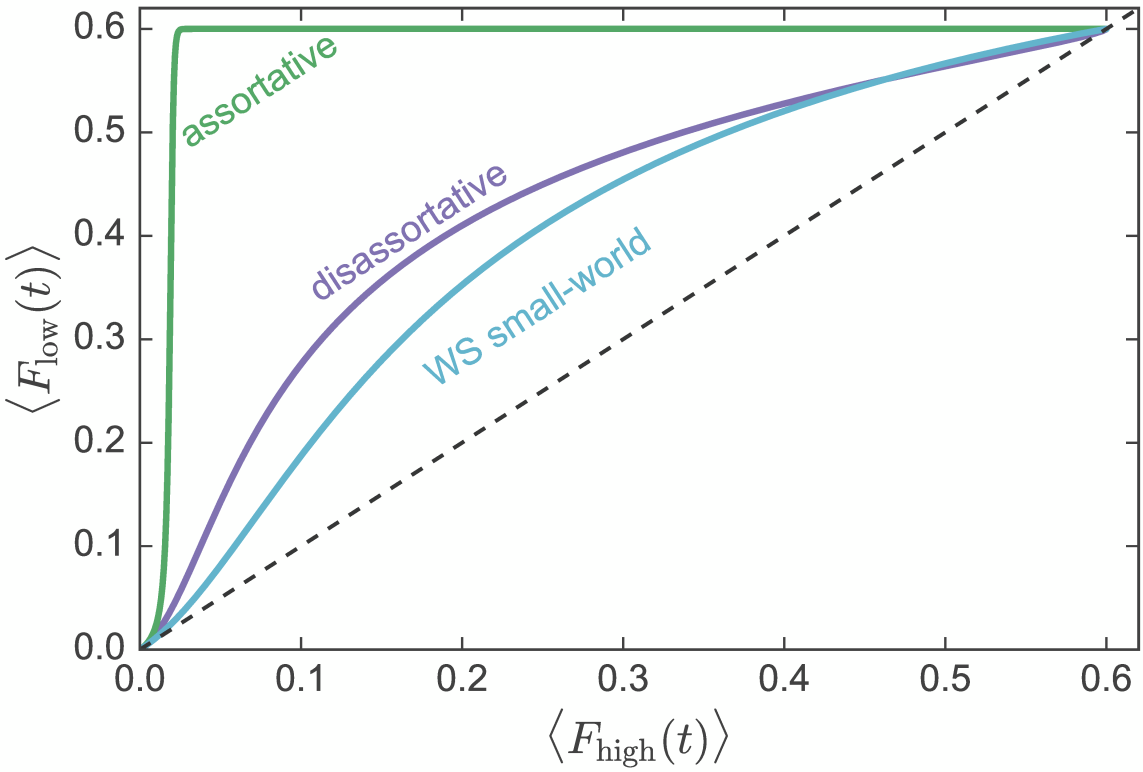


Figure 3.17: Average low-k $\langle F_{\text{low}}(t) \rangle$ vs average high-k $\langle F_{\text{high}}(t) \rangle$ from our mean-field model in Sec. 3.3.4. The dashed black line shows the line $\langle F_{\text{low}}(t) \rangle = \langle F_{\text{high}}(t) \rangle$. A scale-free network with preferential attachment disassortative correlations (default network) in purple, scale-free network with assortative correlations in green, and a WS small-world network with neutral correlations in light blue.

due to their great variability of local frailty f_i.]

3.4.5 Mutual information of FI with mortality

We have seen that F_{low} damages earlier than F_{high} (Fig. 3.6) and that the mutual information of poorly connected ($k = 2$) nodes with large nearest-neighbor degree significantly overlaps with the informativeness of the most connected nodes (Fig. 3.10) in our (disassortative) scale free network model. Because of these informative earlier damaged nodes, we were interested in whether F_{low} could be more informative of mortality than F_{high}, particularly at younger ages. In Fig. 3.18 we show the difference in information for F_{low} and F_{high} for different mortality outcomes vs age. We find that F_{low} is slightly more informative at ages less than ≈ 65 and is increasingly more informative than F_{high} at these younger ages for longer mortality outcomes. This is the result of F_{low} nodes damaging early but having a delayed effect on mortality, so that they are an early predictor of later mortality, but not so much immediate mortality. The relatively large standard deviations for different randomly generated networks shows that this result is affected by the particular randomly generated network.

While the observational NHANES and CSHA sample-sizes are much smaller, a similar calculation shows a slightly lower F_{lab} information -0.002 ± 0.013 compared to F_{clin} in the NHANES data for younger people ($65 - 75$ years) and a slightly higher mutual F_{lab} information $+0.033 \pm 0.027$ compared to F_{clin} in the CSHA data. While we do not have sufficient data to vary our mortality outcome to determine if F_{lab} is more predictive of later mortality outcomes as we did in the model, we can see in the CSHA data that F_{lab} is more informative for younger people.

Since we found that the most informative low-connectivity nodes were those with large k_{nn}, we also considered an FI constructed from $n = 32$ randomly chosen nodes of lowest degree ($k = 2$) from those that have above-average k_{nn}. The information advantage of $F_{\text{low}}^{\text{high-}k_{\text{nn}}}$ is indicated in Fig.3.18 with down and up triangles for 10 and 5 year mortality, as indicated. The advantage over F_{high} is large and significant for ages below $t = 80$ years, with a stronger advantage at earlier ages for later mortality. This will be an attractive avenue to pursue.

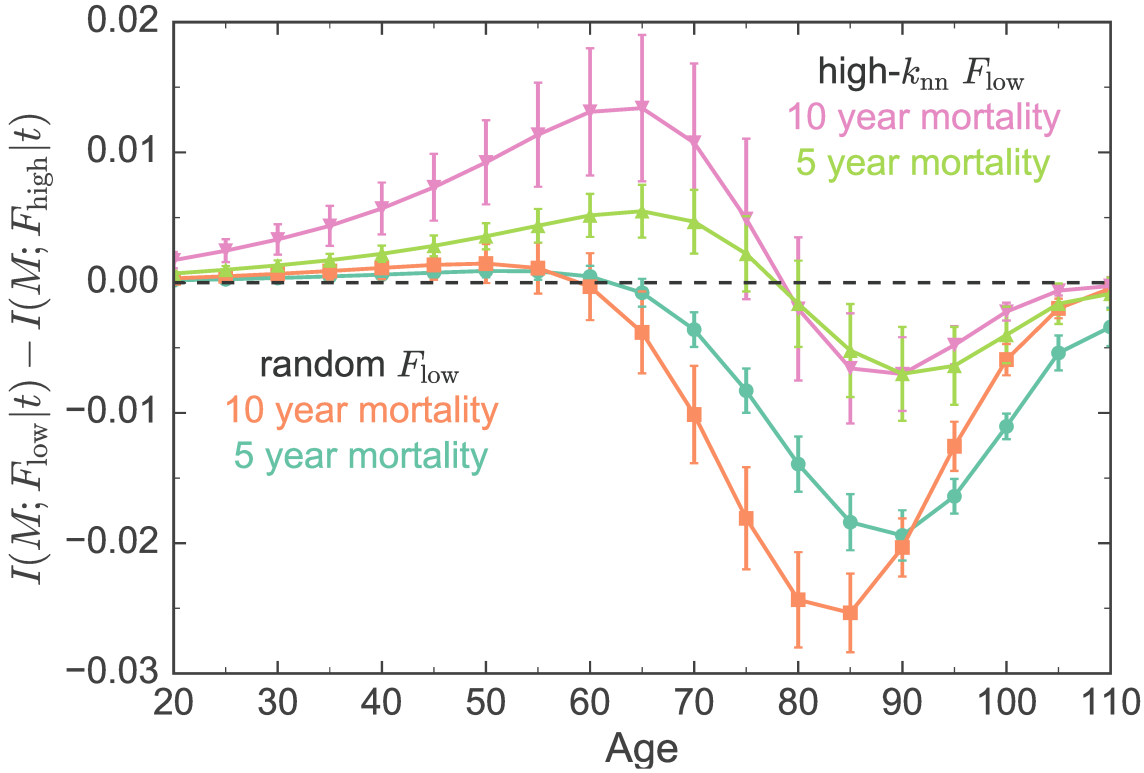


Figure 3.18: The difference in mutual information of F_{low} and F_{high} ($I(M; F_{\text{low}}|t) - I(M; F_{\text{high}}|t)$) vs age t for different binary mortality outcomes. 5 year mortality outcomes are shown as turquoise circles and 10 year as orange squares. The dashed line shows when the information of both FIs are equal. Error bars represent the standard deviation between randomly generated networks. The purple down and green up triangles indicate the information difference for 10 and 5 year mortality, respectively, of $F_{\text{low}}^{\text{high-}k_{\text{nn}}}$ which is constructed with $n = 32$ nodes with $k = 2$ that are randomly chosen from those with above-average k_{nn}.

3.5 Summary and Discussion

The observational F_{clin} or F_{lab} respectively measure clinically observable damage that tends to occur late in life or pre-clinical damage that is typically observable in lab tests or biomarkers before clinical damage is seen. However, they are similarly informative of human mortality [70, 72, 140]. Our analysis indicates that individual laboratory and clinical deficits have broad and overlapping ranges of mutual information.

Our working hypothesis is that clinical deficits correspond to high connectivity nodes of a complex network, while laboratory deficits correspond to lower connectivity nodes. With our network model of individual aging and mortality, we have confirmed that F_{high} and F_{low}, formed from high and low connectivity nodes respectively, behave similarly to the observational F_{clin} and F_{lab}.

Within the context of our aging model, we uncover the mechanisms of this observed behavior. In our model low-k nodes tend to damage before high-k nodes. This is because of the larger average damage rates of low-k nodes compared to high-k nodes (as calculated with our network mean field theory, and illustrated in Fig. 3.7). At the same time, our information spectrum shows that information $I(A; D_i|t)$ increases with k. Roughly speaking, high-k nodes need a larger local frailty f to have comparable damage rates as low-k nodes. Thus, damage of high-k nodes is informative of high network damage, which also leads to mortality. This is why high-k nodes both damage later and are informative of mortality (Fig. 3.10b).

However, some low-k nodes also damage later and are highly informative of mortality. Information $I(A; D_i|t)$ increases with k_{nn} for the low-k nodes, and low-k high-k_{nn} nodes damage later. This can also be explained using the network structure. Low-k nodes are protected from damage when they are connected to high-k nodes. Rapidly damaging low-k nodes without this protection tend to damage early for most individuals, giving these nodes a low information value of mortality. Conversely, protected nodes tend to damage only when their high degree neighbors start to damage, which only occurs when the network is heavily damaged and close to mortality. As a result, only the low-k nodes with high-k_{nn} are highly informative (Fig. 3.10a). Interestingly these nodes still tend to damage before high-k nodes, leading to an early predictor of mortality.

Degree correlations control the average degree of neighboring nodes and hence

control the amount of protection in low-k nodes. By modifying the degree correlations in the network in our computational model we have shown that this protection can be caused by disassortative correlations — where low-k nodes tend to attach to high-k nodes. Conversely, eliminating low-k high-$k_{\rm nn}$ nodes by modifying the network to introduce assortative correlations removes this protection, and we then find all low-k nodes have low information (Fig. 3.13b).

Our mean-field model allows us to explicitly modify the degree distribution and the degree correlations with the nearest-neighbor degree distribution $P(k'|k)$, and to include no other network features. In our mean-field model we see similar results to our computational model where, e.g., adding assortative correlations increases the rate at which $F_{\rm low}$ increases with respect to $F_{\rm high}$. This confirms that degree distribution and degree correlations largely determine the early damage of low-k nodes that we observe in scale-free networks.

Degree distributions and correlations only weakly control the behavior of ER random and WS-small world networks. The low variation in k and $k_{\rm nn}$ in those networks results in a lack of contrast between the damage rates of nodes. This leads to node information that is nearly constant throughout the network and to only small differences in the damage structure of low-k and high-k nodes (Fig. 3.13c and d). This also leads to low magnitude of the mutual information per node, since nodes behave much more uniformly and "randomly" than in a scale-free network. However, we can still see some protection in low-k nodes. This is particularly apparent in the ER random network when $F_{\rm high}$ surpasses $F_{\rm low}$ (Fig. 3.14d).

The behavior of observational deficits seems to best resemble the behavior of the computational model with a scale-free network and disassortative correlations. Node information seen in the (default) scale-free disassortative network is a much better qualitative match of observational data, as compared with scale-free assortative, WS small-world, or ER random networks.

Our analogy between observational deficits and model nodes allows us to make predictions about the underlying network structure of observational health deficits, even though we cannot directly measure this network. Assuming the mechanisms studied here (k and $k_{\rm nn}$ based) are dominant, the observational network should have a heavy-tail degree distribution, so that a large range of possible information values

can be obtained. The network should also include disassortative correlations so that there are connections between high-k and low-k nodes, allowing low-k nodes to be informative of mortality.

We remain open to the possibility that other network topologies not considered here, perhaps with different dynamical models, would provide a better phenomenological description of the observational data. Nevertheless, we have demonstrated that the phenomenology we have studied does discriminate among the network topologies we have considered. In particular, we have found that considering both the age structure and the information fingerprint of both low and high-connectivity deficits is useful in probing the network structure within human aging.

From observational data we find that clinical deficits that integrate many systems into their performance (e.g. functional disabilities, or social engagement) are very informative (Figs. 3.2 and 3.3). In contrast, single diagnoses, even ones strongly associated with age such as osteoporosis, on their own offer less value. The model interpretation of this is that these high information disability deficits have a higher connectivity than lower information clinical deficits. It intuitively makes sense for deficits that integrate many systems to have a large connectivity. In support of this, our partial network reconstruction (Fig. 3.5) shows that high information clinical deficits in both the NHANES and CSHA correspond to nodes with a high reconstructed degree.

We have shown that the age-structure of network damage is related to the network structure. Highly informative low-degree nodes (pre-clinical deficits) damaged early in life promote the damage of their high-degree neighbors, but the damage to their high-degree neighbors takes time and is not seen in the high-degree (clinical) FI until later ages. Indeed, we have shown that a $F_{\rm low}$ is slightly more informative at earlier ages, and is increasingly informative for longer mortality outcomes (5 year vs 10 year) (see Fig. 3.18). Choosing more high-$k_{\rm nn}$ nodes in $F_{\rm lab}$ significantly enhances this effect. Low-k nodes are informative of long-term mortality rather than short-term. Similar results are seen in the observational CSHA data, which indicates that $F_{\rm lab}$ could be used as an early measure of risk of future poor health.

Our network model is generic, without a specific mapping between model nodes and observed human deficits. This is because we have no reliable way of extracting

a specific network from observational data, though we have shown that a partial reconstruction of the rank-ordering of high-connectivity nodes can be done using a method of reconstruction similar to that of WGCNA [154], see also [167]. Distinct parameterization of every node of a network model would require enormous amounts of observational data, if it could be done at all. For example, we have fewer than 70 observational nodes in NHANES data, but our network model uses $N = 10000$ nodes, so observational data under-samples a large network. We similarly expect many more unobserved than observed deficits in any observational study. Instead, we rely on signatures of the network structure that we can calculate from our partial reconstruction. We can then use our generic model to identify robust qualitative phenotypes — to uncover generic mechanisms, to predict behavior, and to improve the utility of the Frailty Index in human aging and mortality.

In this paper we have kept our model parameterization unchanged from the default parameters, though we have checked (data not shown) that our results are qualitatively robust to parameter variation. This has allowed us to explore the impact of network topology on mortality statistics (a small effect) and on mutual information between health deficits (a strong and distinctive effect). The F_{high} and F_{low} model phenomenology are also affected by changes in network topology. This indicates that both F_{high} and F_{low} are usefully distinct characteristics of health in our network model. Our results provide insight into the mechanisms of the similarly useful and distinct observational F_{clin} and F_{lab} [70, 71, 140].

Chapter 4

Generating synthetic aging trajectories with a weighted network model using cross-sectional data

4.1 Background

In Chapter 3 we indirectly probed the network by varying the structure and studying "fingerprints" of dynamical behaviour. We observed the increase in Frailty Index and mortality rate with age and the structure of information content within the deficits to determine that networks within the generic network model require a structure with few high degree "hub" nodes, and many low degree nodes. This is a dissasortative scale-free structure, which also offered an explanation for the differences between FI-clin and FI-lab. This was done by comparing pre-defined network structures to the observed data, and did not involve inferring the network from data. Additionally, this also did not involve fitting damage rates or other model parameters. In this chapter, we build a model where the network and damage rates are parameterized and fit from data by maximum likelihood. This allows the model to be able to represent a network of specific health variables, and generate realistic simulated populations with individual binary health variables and mortality that correspond to the observed data.

To do this, we build from the generic network model by representing a subset of the nodes ($N = 10$ compared to $N = 10^4$) with specific health variables from observational aging data. In the generic network model, each node had the same rate parameters, and damage rates only differed due to the impact of neighbours. In this new approach, the rate parameters for each node are separately fit. We parameterize the rates in terms of a network of connection weights $\mathbf{W}$, and other rate parameters. We also simplify mortality by using a single mortality node.

Since only a small subset of the nodes are represented by specific health deficits from observational data (10 deficits), we perform a mean-field approximation for the unobserved nodes by representing the contribution to the local frailty of each node

by a deterministic function. Starting from the GNM with $N = 10^4$ we can see this in the following way,

$$f_i(t) = \sum_{j=1}^{N} a_{ij} d_j / k_i \tag{4.1}$$

$$= \sum_{j=N-10+1}^{N} a_{ij} d_j / k_i + \sum_{j=1}^{N-10} a_{ij} d_j / k_i \tag{4.2}$$

$$\approx \sum_{j=N-10+1}^{N} a_{ij} d_j / k_i + \mu_i(t), \tag{4.3}$$

where $\mu_i(t)$ is the average contribution of the nodes outside of the top 10 most connected nodes to the local frailty. Performing this approximation significantly speeds up the model, since only 10 nodes need to be simulated stochastically, and the average contribution of the unobserved nodes are computed deterministically with $\mu_i(t)$. In Figure 4.1 we measure this average contribution for each of the top 10 nodes in the GNM. We see the a cubic fit for this contribution captures the behaviour, and so in the weighted network model we fit functions $\mu_i(t) = \sum_{n=0}^{3} \mu_{in} t^n$ for each node i.

We further modify the GNM by generalizing the rates from exponentials with constant rate parameters for each node, to quartic power series with separate parameters for each node $\Gamma_i^+(t) = \phi(\sum_{n=0}^{4} \gamma_{in}^+ f_i(t)^n)$, and replace the binary adjacency matrix a_{ij} with a continuous-valued weight matrix w_{ij} allowing more flexible connections of varying strengths. The function $\phi(x) = \max(0, x)$ allows rates to instantly "turn on" at $f > 0$ when $\gamma_{i0}^+ < 0$. This is a key component of the model, since many of the deficits only occur late in life.

We use cross-sectional aging data to infer the parameters, which includes the network weight parameters. Cross-sectional data only involves one measurement of the health state per individual as well as follow-up mortality information, and does not include time-course health states. We use small networks of 10 binary health variables as nodes, representing potential health deficits of a Frailty Index.

This chapter presents the paper "Generating synthetic aging trajectories with a weighted network model using cross-sectional data" published in 2020 [48].

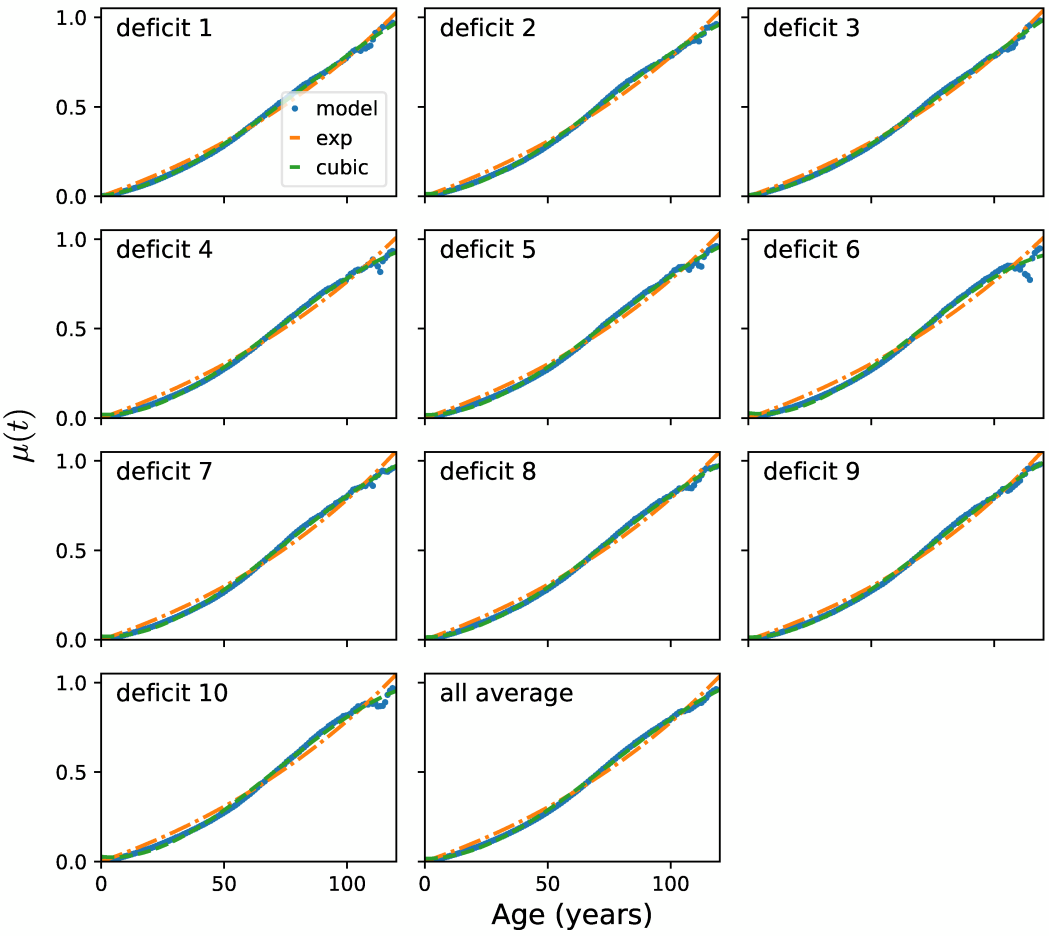


Figure 4.1: **GNM Top 10 nodes $\boldsymbol{\mu}(t)$.** The average contribution from the nodes outside the top 10 to the local frailty of the top 10 most connected nodes in the GNM. Exponential and cubic fits to each deficit are shown along with the average from the GNM.

4.2 Introduction

Human aging is a complex process of stochastic accumulation of damage [52] that occurs at many organismal scales ranging from the cellular [4] to the functional. Individual health trajectories are heterogeneous, but typically worsen with age as damage accumulates. Heterogeneity of aging trajectories arises even in studies of clonal organisms in controlled laboratory conditions [53, 129], and is an intrinsic part of aging. Heterogeneity in health as individuals age has been measured with a variety of methods, although here we focus on binary "health deficits" determined from routine clinical assessment and self-reported surveys [14–16, 69, 168]. Health deficits are indicators of an aging phenotype, indicating disease, laboratory abnormalities, cognitive impairment, disability, or difficulty performing everyday tasks.

While any single deficit may not be a good measure of overall health, or a very informative predictor of mortality, averaging many binary deficits to evaluate overall health provides measures that are strongly associated with both adverse health outcomes and mortality [14–16, 46, 69, 168]. Furthermore, such frailty indices (FIs) are robust to missing or heterogeneous data [15]. Using "high-level" health-deficits provides a measure of health that is both conveniently assessed and reflecting the functional aspects of healthy living that are important to the individual [169]. Such an FI also contains information about health that is not found in recent epigenetic measures based on DNA-methylation [170].

While the FI has been shown to be broadly predictive of both mortality [171] and of the accumulation of individual deficits [168, 172], it does not distinguish the health trajectories of two individuals with the same FI even if they have distinct sets of accumulated deficits. Capturing the heterogeneity in health trajectories requires modelling the full high-dimensional set of health variables. We develop a model to generate populations of synthetic health trajectories, which capture this heterogeneity.

However, the development of models of aging is complicated by the data currently available. Large observational studies (10^4+ individuals) with linked mortality are often cross-sectional (measuring most variables only when individuals enter a study), have short censored survival outcomes, and have a lot of missing data. This has made developing realistic models of human aging difficult. Nevertheless, such models

would be useful to generate model individual health trajectories during aging from birth, or from baseline data of actual individuals. Encouragingly, a model capable of generating general health trajectories during aging using cross-sectional data has recently been developed [173] – though it did not consider individual survival.

Here, we develop an intuitive model that can be fit with cross-sectional data with censored survival information to generate individual aging trajectories that include both health *and* survival. Our model is adapted from previous work modeling human aging with stochastic dynamics on a complex network [11, 44, 45], which was shown to capture population level aging phenomena, such as Gompertz' law of mortality [135]. This model was based on the intuitive assumption that having one health deficit increases the risk of acquiring another one, and so deficits can be thought of as interacting in a network, in which connections establish pairwise associations [47]. Nodes in this older model represented generic/abstract deficits and corresponded to no specific physiological systems in particular. Nevertheless, their collective behaviour captured key aspects of aging. This "generic" network model (GNM) used many nodes ($N = 10^4$) to abstractly represent the many interacting physiological systems in the human body, had simple interactions between nodes, and needed no age-dependent programming of damage.

Our new "weighted" network model (WNM) is parameterized so that each node represents an actual health attribute (potential health deficit) corresponding to observational aging data. We recognize that we will never be able to incorporate all possible health attributes as nodes in our network or to describe exact biological mechanisms. For this reason, we use more complex weighted interactions between observed nodes that can capture the effective behaviour of underlying and/or unobserved biological mechanisms. This new WNM can be considered as a coarse-grained adaptation of our previous GNM, with far fewer nodes.

We separately fit our WNM with cross-sectional observational data from the Canadian Study of Health and Aging (CSHA) [160] and the National Health and Nutrition Examination Survey (NHANES) [159]. These human aging studies consist of 8547 and 9504 individuals, age ranges of $65 - 99$ and $20 - 85$ years, in which mortality data are available for at most 6 or 10 years past study entry, respectively. Deficits in these datasets are binary indicators of health issues and integrate information across

physiological systems, such as difficulty performing activities of daily living (ADLs) or more complex instrumental activities of daily living (IADLs). We estimate parameters for each study by maximizing the log-likelihood for our model to recover the observations, where the likelihood is estimated from simulations of the stochastic model. By validating our model on a separate test set, we demonstrate that our model represents real aspects of aging, and is not overfitting to the training data.

We find that our synthetic individuals generated with our model capture the health outcomes and survival of observed data with a number of different measures. Indeed, rather than focusing on achieving optimal predictive performance on any one particular task, our goal was to obtain a robust model that can generate realistic trajectories for multiple health attributes at once for many individuals from either actual or synthetic baseline health status. Nevertheless, our model cannot overcome the intrinsic limitations of cross-sectional data — for example, the accuracy of health trajectories will only be assessed by comparing simulated cohorts of individuals to longitudinal data of the observed cohort.

4.3 Results

4.3.1 Health trajectories

Starting from an individual with a set of N potential binary deficits at a baseline age t_0, $\{d_i(t_0)\}_{i=1}^{N}$, our model (see Model section below, trained on the CSHA dataset) generates deficit trajectories $\{d_i(t)\}$ describing health for synthetic individuals for each age $t > t_0$ until mortality. (Here $d_i = 1$ indicates a deficit for the ith aspect of individual health, while $d_i = 0$ indicates no deficit. We generally refer to a set of potential health deficits $\{d_i\}$ as health attributes). We want to test whether these synthetic individuals age with the same properties as do real individuals in the observed data. Without longitudinal data, we cannot test these individual trajectories directly. However, we can use the population average of the observed cross-sectional data and compare with the average population trajectory predicted from our model. If the study population was randomly sampled with no biases, we expect these average trajectories to agree.

Given baseline age and $N = 10$ selected deficits for individuals from the test data

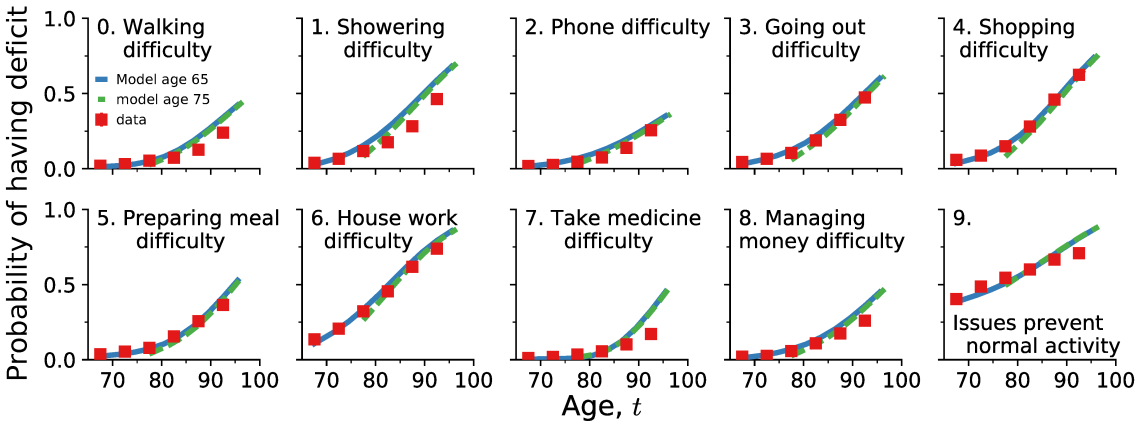


Figure 4.2: Average predicted trajectories of ten deficit prevalences (as indicated by the subplot titles, numbered 0 to 9) vs. age for individuals from the test data aged $65-70$ surviving past 70 (solid blue lines) and aged $75-80$ surviving past 80 (dashed green lines). Observed CSHA prevalence is shown in red squares; standard errors are smaller than the point size.

aged 65 - 70 that survive past age 70, and individuals aged 75 - 80 that survive past age 80, we use our model to forecast their health trajectories. These simulated trajectories allow us to the compute the full joint distribution of health states vs age for these simulated individuals, $\hat{p}(\{d_i\}|t)$. Since this distribution is 10-dimensional, with $2^{10} = 1024$ distinct age-dependent probabilities, we plot only the prevalences of each deficit (by marginalizing the distribution). We compare the average of these individual trajectories until death to the deficit prevalence from the observed cross-sectional CSHA data for ages 70+ and 80+. In Figure 4.2 we show health trajectories using prevalence at age t of the ith deficit, $\hat{p}(d_i = 1|t)$, for the model (blue solid lines for 65 - 70 and green dashed lines for 75 - 80) together with the observed CSHA prevalence (red squares). We see excellent agreement for nearly 30 years for most deficits. This shows that the model is able to project a population forward in time while correctly identifying changes in deficit prevalence. We also show the average predicted trajectories with an alternative set of 10 deficits in Appendix Figure A.3 and using the NHANES data in A.4.

A similar prediction is done for the prevalence of pair combinations of deficits, i.e. comorbidities. We predict the probability of having two specific deficits $\hat{p}(d_i = 1, d_j = 1|t)$ vs age in our model, and compare with the observed data. This is

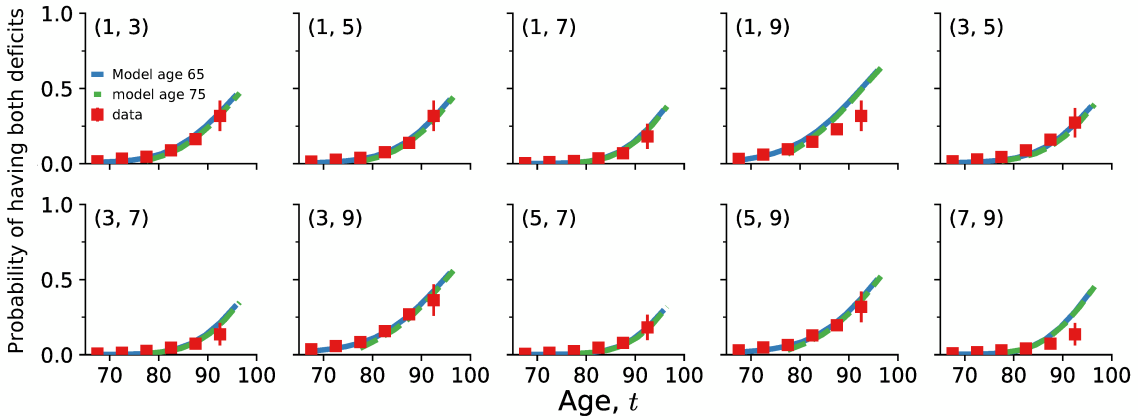


Figure 4.3: Average predicted trajectories of pairwise deficit prevalence vs. age for individuals from the test data aged $65-70$ surviving past 70 (blue solid lines) and aged $75-80$ surviving past 80 (dashed green lines). Subplot titles indicate the two deficits included in the pairwise prevalence, with numbers corresponding to titles in Figure 4.2. Only odd numbered pairs are shown here; other pairs are shown in Appendix Figure A.5. CSHA data is shown by red squares; errorbars represent standard errors of the pairwise prevalence.

shown in Figure 4.3 and in Appendix Figure A.5, and also shows excellent agreement between the model and the CSHA data for over 30 years. This indicates that our model is accurately capturing the association between pairs of deficits through network interactions. Notably, pairwise combinations in the model often perform better than corresponding prevalences – see for example $(1,7)$ in Figure 4.3 – confirming that pairwise combinations (together with higher order combinations, not shown here) are non-trivial predictions of the model.

We can represent overall individual health with the well-established "Frailty Index" (FI), an index that uses the proportion of deficits, $F = \sum_{i=1}^{N} d_i/N$, as a predictor of health and mortality [14,15]. In Appendix Figure A.6 we show that the heterogeneity in health as individuals age, as characterized by distributions of FI at different ages, is similar to the observed CSHA data.

We are not limited to modelling known individuals. In Appendix Figure A.7 we show that the population prevalences of synthetic individuals starting from birth with zero damage also agree with the observed CSHA prevalences. Indeed, we can generate trajectories and survival curves for any baseline age and individual set of

deficits. Using synthetic populations, we can also generate trajectories starting from any age with partially observed sets of deficits with missing values.

Figure 4.4 shows FI trajectories starting from the known baseline data (red circle) for 6 synthetic individuals with specific deficits. Horizontally, we vary baseline age with 65, 75, and 85 along the columns. Vertically, we vary baseline deficits, with bottom individuals having a higher initial FI by having two additional deficits. Individual trajectories are conditioned on dying at their median survival probability (dashed black lines), seen from the individual black survival curves. Shaded regions show a distribution of FI trajectories. The trajectories behave reasonably. Individuals with more baseline deficits accumulate additional deficits faster and die sooner. Individuals starting at older ages also have a more rapid increase in number of deficits and have a shorter time to death. Note that these trajectories exhibit FIs larger than the typically observed maximum of 0.7 [68,168,174], which is due to the small number of potential deficits used here (only 10) compared to typical studies (with 30 - 40 potential deficits).

4.3.2 Individual deficit predictions

Since our training and test data sets have similar distribution of health states (e.g. deficit prevalence), the good test performances in Figure 4.2 and Figure 4.3 do not rule out overfitting to the training set, because the model was only assessed against the distribution of health states for the training and test data. To assess overfitting, we need to consider predictions of individual health states – i.e. observed deficits for specific individuals. We first verify that there is only a $\sim 20\%$ overlap between the observed health states of individuals in the training and test data sets, see Appendix Figure A.8. (We also confirm that using only one or two attributes leads to a $\gtrsim 90\%$ overlap. Note that this overlap is not due to the same individuals being present in the training and test sets, but due to the discrete nature of the binary deficits.)

We test the model's ability to capture the age-dependent joint distribution of deficits $p(\{d_i\}|t)$ by evaluating its performance in predicting "left out" deficits from individuals in the test set at the same age, i.e. performing missing data imputation by estimating $p(\{d_j\}_{\text{missing}}|\{d_i\}_{\text{observed}}, t)$. Given a known age $t^{(m)}$ and known health

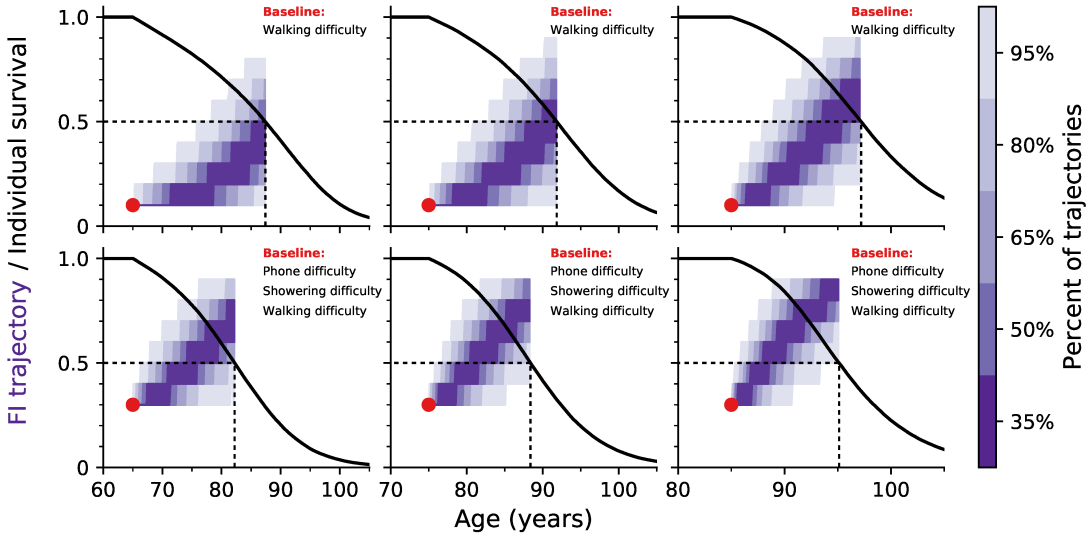


Figure 4.4: The purple shading indicates the distribution of Frailty Index trajectories vs age for simulated individuals starting from the red circle at a specific age, with the indicated starting set of deficits. The scale-bar at right indicates the probability that a simulated trajectory falls within a particular colour. In the top row individuals only start with one deficit. In the bottom row individuals start with three deficits. In the three columns, individuals start at age 65, 75, or 85. Individual survival curves are shown as black lines, and trajectories are conditioned on dying at the median death age (indicated with dashed lines).

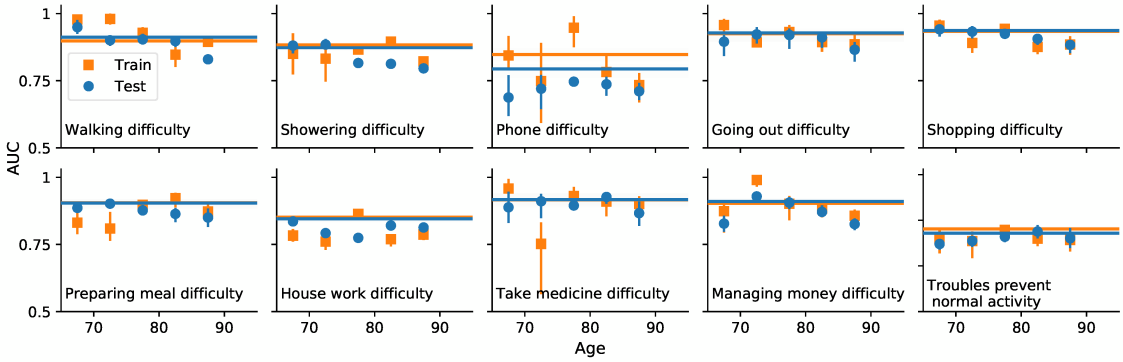


Figure 4.5: AUC for predicting the value of individual left-out deficits for test (blue circles and lines) and training (orange squares and lines) data. The points show the AUC when stratifying by age. Errorbars indicate 95% confidence intervals. The solid line shows the result for all ages. The baseline AUC is 0.5 for random (uninformative) predictions.

attributes $\{d_i\}^{(m)}$ for an individual m from the test set, we simulate from zero damage at birth and sample from the simulated individuals at age $t^{(m)}$ that have the health attributes $\{d_{i\neq j}\}_{i=1}^{N}$ to estimate the probability of having the left out deficit, $\hat{p}(d_j = 1|\{d_{i\neq j}\}^{(m)}, t^{(m)})$. We compare this probability with the actual value of the left out health attributes $d_j^{(m)}$ for this individual at the same age. This is a binary classification, and can be quantified by the area under a ROC curve – the AUC.

Figure 4.5 shows the AUC for each left-out deficit. This is equal to the probability that given two individuals with and without the deficit, the model correctly predicts a higher probability of having the deficit for the individual with the deficit than without. The full test set is shown as a solid blue line, while blue circles show the AUC stratified by age. For all deficits, we see AUC values well above 0.5 for the test data, meaning that our model is making informative predictions and not just overfitting the training data. Orange squares and lines show similar values for the AUC of the training data, which is further confirmation that any overfitting is minimal.

Obtaining essentially the same AUC when stratifying by age in Figure 4.5 demonstrates that the model is predicting as well even when we eliminate age as a factor. This indicates that the model is utilizing the network interactions between observed deficits to make non-trivial individual predictions, and not just using e.g. increasing prevalence with age.

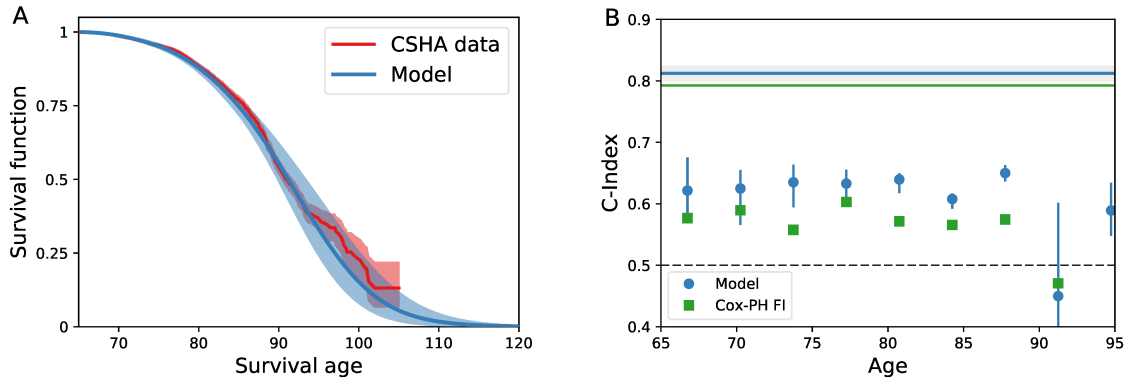


Figure 4.6: A) Model population average survival function (blue) and CSHA Kaplan-Meier population survival function (red). Error in the model and data are shown as shaded regions for a 95% confidence interval. B) Survival C-Index stratified by age (blue circles) and unstratified (solid blue line). Green squares and lines show the age-dependent C^{td} for a Cox-proportional hazards model with age and FI. Errorbars on points and grey region around the unstratified line show 95% confidence intervals.

4.3.3 Survival

We can also model individual mortality. We take baseline data from the test set for an individual m at age $t^{(m)}$ and health attributes $\{d_i\}^{(m)}$, and simulate from this age until mortality. This allows us to estimate their survival function $\hat{S}(a|t^{(m)}, \{d_i\}^{(m)})$, i.e. the individual's probability of surviving to an age $a > t^{(m)}$. We average these to get a population average survival function over M individuals, $\langle \hat{S}(a|t^{(m)}, \{d_i\}^{(m)}) \rangle_M = \frac{1}{M} \sum_m \hat{S}(a|t^{(m)}, \{d_i\}^{(m)})$. We show the comparison of this to a Kaplan-Meier estimate [175] of the population survival function from the observed CSHA test data in Figure 4.6A, with our model shown in blue and the observational test data in red. We observe good agreement, and also find that model predictions correctly drop to zero survival by age 120.

Since the training and test distributions are similar, a population measure of mortality does not tell us whether the model is overfitting – only whether the model is able to capture the population trends in mortality. Accordingly, we validate individual survival on the test set with a C-index [176] to measure how well the model discriminates individuals in terms of risk of mortality. The C-index is the probability that the model correctly predicts which of a pair of individuals lives longer, so a C-index above 0.5 is making predictions better than random.

Since our model includes potentially complex time-dependent effects, where survival curves can potentially cross, we use a more general age-dependent C-index. [177] We obtain this by comparing the rank ordering between survival probability and known survival age while including censoring, so $C^{\text{td}} = \Pr(\hat{S}(a^{(m_1)}|t^{(m_1)}, \{d_i\}^{(m_1)}) < \hat{S}(a^{(m_1)}|t^{(m_2)}, \{d_i\}^{(m_2)})|a^{(m_1)} < a^{(m_2)}, c^{(m_1)} = 0)$ [177].

Figure 4.6B shows this age-dependent C-index for both on the full test set (solid blue line) and stratified by age (blue circles). The C-index shows that the model discriminates well on the full test when the difference in ages between individuals can be used in the discrimination. When we stratify by age to eliminate this effect, we nevertheless see that the model still discriminates well based on just these 10 deficits alone, indicating that the model captures an increased risk of mortality from specific deficits. In particular, our model performs better than a standard Cox-proportional hazards [119] model using the Frailty Index and age (green squares and line). We note that stratified values are noisy due to the small number of individuals per age bin, especially at higher ages.

In Appendix Figure A.9A we show similar results for the C-Index for both training and test sets, which also indicates a lack of overfitting. Similarly, Appendix Figure A.9B shows an R^2 measure constructed from Brier Scores [178]. a measure of how well predicted and observed survival curves match, that behaves similarly for training and test data. Furthermore, Appendix Figure A.9C shows the ROC AUC for predicting binary dead/alive on the train/test sets within a specific window of time, finding a similar AUC of approximately for 1-5 year mortality windows. For all of these, we find similar behavior between the training and test sets, indicating a lack of overfitting. This is also seen for survival predictions for the alternative set of deficits used in Appendix Figure A.3, as shown in Appendix Figure A.10 and for survival predictions for the NHANES dataset, as shown in Appendix Figure A.11.

4.3.4 Inferred network structure

In the Appendix A section on Parameter Robustness, we explore the "robustness" of both our network parameters and predictions by sampling an ensemble of parameters around the maximum likelihood estimate [179]. Appendix Figure A.1 shows that significant deviations from the maximum likelihood parameters still leads to relatively

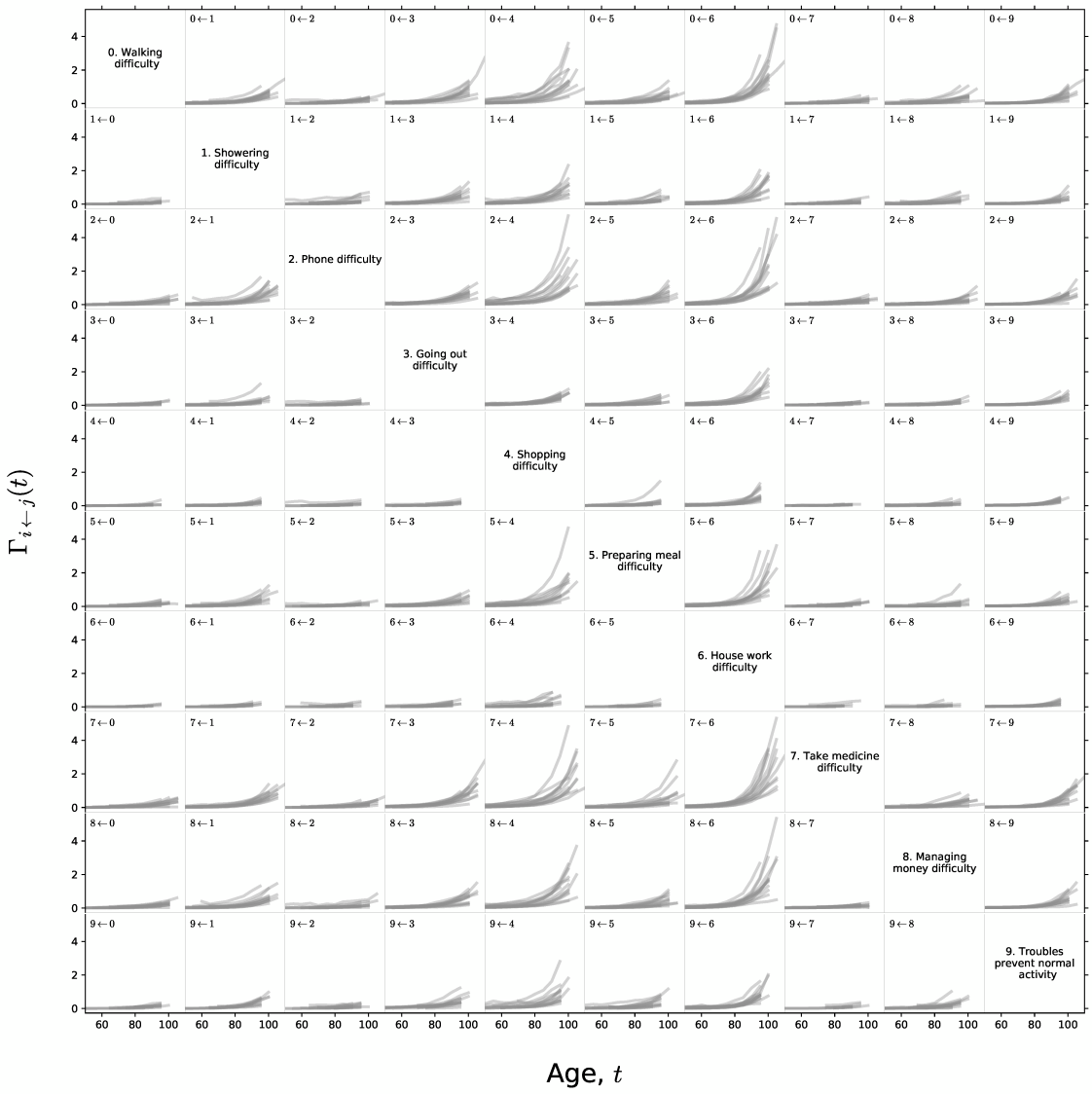


Figure 4.7: **Damage rate robustness.** Average pairwise damage rate $\Gamma_{i \leftarrow j}(t) = \langle \Gamma_i^+(t, d_i = 0, d_j = 1, \{d_l\}) \rangle_{p(\{d_l\}|t, d_i=0, d_j=1)}$ of node i given that node j is damaged, vs age. Each curve is averaged over combinations of the other node states, weighted by their probability of occurring in the simulation. Each intersecting row and column represents the indicated node labelled on the diagonal, and the direction of the links is indicated by labels on each subplot. Each subplot has 13 rate curves, each with different estimated parameters for a different starting seed. The qualitative agreement of these curves for particular pairs of nodes indicates that these detailed damage rates are qualitatively robust (note the linear scale), even though the effective degrees of the nodes are not (see Appendix Appendix Figure A.2).

accurate fits of the data – i.e. we obtain robust predictions. However, when we optimized the model several times with different random seeds we show in Appendix Figure A.2 that we find significantly different network parameterizations each time. Appendix Figure A.12D shows that even the sign of individual connections is not robust. We conclude that while the model behavior is robust, the network structures themselves are not robustly predicted by the available data.

Nevertheless, we show in Figure 4.7 that how damage propagates from node to node of the network does have some degree of robustness. We show average pairwise damage rates $\Gamma_{i\leftarrow j}(t) = \langle \Gamma_i^+(t, d_i = 0, d_j = 1, \{d_l\})\rangle_{p(\{d_l\}|t,d_i=0,d_j=1)}$ of the ith node conditional on prior damage of the jth node. This robustness shows that the behavior of our weighted network model (WNM) is robust despite some sloppiness of individual parameters.

4.4 Discussion

Our weighted network model (WNM) is trained with cross-sectional data, generates cohorts of synthetic individuals that resemble the observational data, and can forecast the future survival of real individuals from their baseline health and age. We have validated the WNM model through a variety of measures. Synthetic individuals age with trajectories that have approximately the same prevalence of deficits and comorbidities as in the observed data. The average trajectories predicted by the model agree very well for nearly 30 years. Given a set of known deficits, the model can predict the probability of having a missing or unknown deficit at the same age, demonstrating the models ability to capture the age-dependent joint distribution of the deficits. Estimated survival curves also agree with observed population survival, are predictive of mortality, and discriminate between individuals.

Our model has a large number (188) parameters while our health data has only $N = 10$ binary attributes. One concern with having at most $2^N = 1024$ discrete health states is that there could be significant overlap between the test and training sets. Nevertheless, we showed in Appendix Figure A.8, that a significant fraction of the health states from the test set are not in the training set. Using this (non-overlapping) test set we have shown that our model does not substantially overfit and can make predictions on unseen test individuals.

We emphasize that our model is not just fitting the prevalence of the $N = 10$ deficits, but is fit to the full 10-dimensional age-dependent joint distribution $p(\{d_i\}|t)$ which has 1024 distinct states. A full model of this distribution would require many more than 1024 parameters. Our network model substantially reduces the number of parameters required, by only directly fitting pairwise interactions and getting effective higher order interactions through the local damage term in our model (f_i). While a minimal model using our network approach using linear or simple exponential forms for the functions describing rates in our model would have 144 parameters, we have increased this to 188 by using 3rd order power series. This has offered increased flexibility to the model and also improves performance.

Our model generates accurate projections of the average health trajectories of groups of individuals. Taking a group of individuals and simulating them to their deaths, we find the average trajectory generally agrees with the average population data, which shows that model averaged trajectories are quite accurate and are consistent with the assumption that the study population is a random (representative) sample. Note that this does not mean that we can overcome the intrinsic limitations of cross-sectional data, and to validate the accuracy of individual predicted trajectories we would need longitudinal data.

Our model works when separately trained and tested on CSHA and NHANES cross-sectional datasets. This success indicates that our approach should work more broadly with comparable cross-sectional data from other datasets. Nevertheless, we find that performance is somewhat worse on the NHANES dataset. This may be due to the presence of many missing values in the NHANES dataset, the increased age range of the NHANES dataset ($20 - 85$ years for NHANES vs $65 - 99$ for CSHA), or perhaps differing biases in the cohorts studied.

When we predict health trajectories until very old ages (80 - 90 years old), our model tends to estimate slightly higher prevalence than are observed in cross-sectional data (see Figure 4.2, and Appendix Figures A.3 and A.4), particularly in the NHANES data where baseline measurements are taken further away from the actual age of death than in the CSHA data. One possible explanation for this is that there is a compromise in the model between fitting these trajectories and fitting survival, and the model is attempting to amend this compromise by fitting trajectories well for

early ages, then rapidly damaging deficits to induce mortality in individuals to obtain the correct survival predictions. Another possibility is that the damage rates need to rapidly increase from ages 60 to 90, but then slowly taper off of this increase for these older individuals. Fixing this would require more flexible damage rate functions capable of tapering off for very old individuals.

Alternatively, our model could be describing a real acceleration of health decline before death that is not captured in the observational cross-sectional data due the lack of health measurements near death. In other words, these cross-sectional studies could be biased by excluding subjects near death, and our model is correctly including a large increase in the rate of damage near death. Indeed, in longitudinal studies a rapidly rising FI has been shown to identify individuals with a high risk of death within 1 year [180] – this is called "terminal decline" [181]. Using such longitudinal data (see below) would allow us to better predict and to better test generated health trajectories for specific individuals, including health near death.

Individual survival is assessed with the C-index [176]. The C-index evaluates the model's ability to predict the relative risk of death for pairs of individuals. A C-index of 1 represents perfect predictions, but in practice intrinsic variability of individual mortality will limit the C-index below that in an age and health-dependent way. The availability of individual data will further limit the C-index below that – but generally above 0.5, for an uninformed random guess. We observe C-index values of around 0.6 when stratifying by age, which means that just by using 10 binary deficits, we can predict which of two individuals of the same age lives longer with 60% accuracy. Our model achieves better results than a simple Cox-proportional hazards model [119] using the Frailty Index, but a larger number of health deficits and more data would further improve our model's predictions.

Our previous generic network model (GNM) captured population level aging behaviour like Gompertz' law of mortality and the average increase in the Frailty Index (health decline) vs age [44, 45], however the nodes did not correspond to any particular health attribute (i.e. they were generic). Adding more complexity to damage/mortality rates with more flexible functional forms, node-dependent fitting parameters, and a weighted interaction network, we have here been able to represent individual health attributes from observational aging data with specific nodes

in our WNM. This has allowed us to model individual health trajectories, including individual survival.

The choice of which deficits to use with our model is arbitrary, the only assumptions we require are that they are binary and not reversible. We do not need the deficits to have strong correlations between them or be good predictors of mortality to capture the sample trends in health trajectories (Figure 4.2 and Figure 4.3) or overall survival function (Figure 4.6A), since these do not show individual predictions but instead captures how well the model overall captures the trends seen in the data. However, the quality of individual predictions for left-out deficits and survival (Figure 4.5 and Figure 4.6B) do depend on the deficits chosen, because these are predictions for specific individuals. For independent deficits that are absolutely uninformative of mortality, we would expect AUCs of 0.5 for left-out deficit prediction and a C-index of 0.5 for the mortality prediction.

The only health variables we have included are health deficits that accumulate through damage. Other static or non damage-accumulation variables are often considered in aging studies as well, such as sex [182–184], and environmental variables like socioeconomic status [185], or lifestyle. Variables that don't change by damage accumulation but still interact with health deficits can be easily added to the model as network nodes with static values. In this way they could naturally interact with the damage-accumulation health attributes. Similarly, individual non-damage variables that could be deliberately modified – such as physical activity levels [186, 187] – could be added as nodes with explicit time-dependent values that depend on the individual.

There has been significant work inferring biological networks, using a variety of approaches [151, 152, 154, 188] and at different scales [189]. In the context of human frailty, previous work has created a network representation of health attributes with measures of association or correlation [45, 190, 191]. Different methods result in different networks, and thus it has not been clear what underlying association between the deficits a given network represents. Equivalently, it has not been clear how to test a given network representation. In this work we conclude that while the model behavior is robust, the network structures themselves inferred by the model are not robustly predicted by the available data.

This lack of robustness of the network is not surprising. Due to the complex interactions between many parameters in our model, we expect that many network parameters are "sloppy" [192]. This results in robust collective behavior of the system for many different combinations of parameters – i.e. many different networks that are consistent with the observed data. Indeed, this robust behavior seems necessary to perform well in predicting deficit prevalence and mortality.

Our network model imposes casual mechanisms within the simulation – it assumes that there is a direction to the network weights, and attempts to infer those weights within the assumptions of the model. Since we do not have adequate data to be able to infer true casual relations, these directed links are simply chosen for accurate prediction. A directed link in our model is just defined in terms of prediction: a particular directed connection between two variables is included if it improves prediction accuracy. Similarly, our network connections are not simply correlations between the variables (for those, see Appendix Figure A.12), but are chosen to improve prediction.

In recent models of disease progression or aging either the mortality is not considered [173], or the models require longitudinal data [193], or both [194–196]. Structurally, our model differs from others by using an explicit network describing pairwise interactions, and it uses this network to generate stochastic changes to their health state as they age until death – rather than capturing the dynamics with unobserved latent variables that are harder to interpret. Using discrete health states within our model allows us to simply compare with observed health states using maximum likelihood methods, and allows our success while using only cross-sectional data. That said, our approach could be extended to use longitudinal data for training – and we would expect this to further improve model behavior.

The interpretability of our model structure makes it straightforward to adapt our model to new applications. We can easily generate synthetic tracked health trajectories, or forecast the future trajectories of individuals from specified health states. This means that our model can generate many different stochastic realizations for the same individual after baseline, and can show how differences in possible health trajectories lead to differences in mortality. Another application of our computational approach that would be facilitated by our model structure is to manipulate model individuals to perform "health interventions" on specific observed nodes or sets of

nodes. We could then observe the affect of general interventions on health trajectories and mortality. These predictions could then be tested with longitudinal data. This is left for future work.

4.5 Methods

4.5.1 Model structure

In previous work, we developed a generic network model (GNM) [44, 45] to study how damage propagation in a network can lead to similar behaviour as observed in aging, in terms of population health and mortality. In this work we expand upon and generalize the GNM to be able to fit the model to individuals with specific observed deficits with a maximum likelihood approach. This allows us to generate synthetic individuals from the model, which age with similar properties as the cohort used to fit the model.

We consider a network of N nodes representing binary health attributes. Each node $i = 1, ..., N$ can be in state $d_i = 0$ for undamaged (healthy) or $d_i = 1$ damaged (deficit). Nodes in the network undergo stochastic damage transitions ($0 \to 1$) as an individual ages. These transitions occur with rates that depend on the local damage of neighbouring nodes. We call this local damage the "local frailty", f_i.

In the GNM, we measured the local damage around a specific node as the proportion of damaged neighbours, $f_i = \frac{1}{k_i}\sum_{j=1}^{N} a_{ij}d_j$, where a_{ij} is the binary-valued adjacency matrix of an undirected network and $k_i = \sum_j a_{ij}$ is the node degree. Damage transitions ($0 \to 1$) between states occurred with rates that depend exponentially on the proportion of damaged neighbours, described by a function $\Gamma^+(f_i) = \Gamma_0^+ \exp(\gamma^+ f_i)$ with tunable parameters γ^+ and Γ_0^+. Here we use superscript "+" to denote that this is a damage rate, and correspondingly use "-" to denote repair rates. The baseline rate Γ_0^+ controls the damage rate when $f_i = 0$, and γ^+ controls how strongly the rate increases with increasing f_i. Similar repair transitions were also included (with separate parameters γ^- and Γ_0^-), but were found to be negligible. The parameters γ^+ and Γ_0^+ were identical for each node and chosen to fit population mortality rates (Gompertz' law) and overall health decline (average Frailty Index). For the GNM studies we used $N = 10^4$ nodes [44, 45].

In this work, we generalize the GNM to allow the model to represent specific health attributes measured in observed health data as nodes in the network. We generalize the original binary and undirected network to a weighted and directed network, described by a continuous-valued adjacency matrix of weights, w_{ij}. These weights represent the strength of connections between pairs of nodes. We call this a weighted network model (WNM). We use far fewer than 10^4 nodes in this WNM network, but account for the contribution of these missing nodes by introducing a time-dependent function $\mu_i(t)$ to the local damage of each node, f_i. This function $\mu_i(t)$ represents the average contribution to the local damage by the dynamics of the unobserved nodes. This average local damage contribution $\mu_i(t)$ for each node i is implemented as a power series in terms of t with coefficients $\{\mu_{i\,n}\}_{i=1,n=0}^{i=N,n=n_f}$, where n_f is a hyperparameter for the number of terms used in the power series. This means our new measure of local damage for the ith node is a weighted sum over all of the nodes of the network, with the additional contribution from the $\mu_i(t)$ term,

$$\begin{aligned} f_i(t,\{d_j\}) &= \phi\Big(\sum_{j=1}^{N} w_{ij}d_j + \mu_i(t)\Big), \qquad (4.4)\\ \text{where} \quad \mu_{\mathrm{i}}(\mathrm{t}) &= \sum_{n=0}^{n_f} \mu_{i\,n}t^n. \end{aligned}$$

Powers in power series are indexed by n, individual deficits or rates are indexed by i, and sums over nodes in the network are indexed by j. We use this convention of indexing throughout the methods. The function $\phi(x) = \max(x, 0)$ is a "rectifier" or "hinge" function [197] that clips negative values to zero, resulting in a continuous non-negative function. This can allow strong non-linear behaviour by allowing the function to be able to effectively "turn on" at older ages. The network weights $\{w_{ij}\}_{i,j=1}^{N}$ and power series coefficients $\{\mu_{i\,n}\}_{i=1,n=0}^{i=N,n=n_f}$ are included as fitting parameters of the model. The coefficients $\{\mu_{i\,n}\}$ are constrained so that $\mu_i(t)$ increases monotonically with age, details are in Appendix A.

The exponential damage rates of the GNM have been replaced by more general power-series in terms of f_i, with node-dependent coefficients to allow each node to have a different damage rate. This way, specific nodes in the network are able to represent the distinct behaviour for specific deficits in the observed data (in contrast

to the generic network model). This more general damage rate for node i is given by,

$$\Gamma_i^+(t,\{d_j\}) \;=\; \phi\Big(\sum_{n=0}^{n_+}\gamma_{in}^+ f_i(t,\{d_j\})^n\Big). \tag{4.5}$$

This function describes the damage transition rate from 0 to 1 for node i. The power series coefficients $\{\gamma_{in}^+\}_{i=1,n=0}^{i=N,n=n_+}$ are fitting parameters of the model and n_+ is a hyperparameter that determines the highest-order in the power series. The coefficients $\{\gamma_{in}^+\}$ are constrained so that the rate increases monotonically with f_i.

Mortality occurs as a separate process with a rate of death that follows the same form with a power series,

$$\Gamma_D(t,\{d_j\}) \;=\; \phi\Big(\sum_{n=0}^{n_{D_1}}\alpha_n x(t,\{d_j\})^n\Big), \tag{4.6}$$

$$x(t,\{d_j\}) \;=\; \phi\Big(\sum_{j=1}^{N}\beta_j d_j + \sum_{n=0}^{n_{D_2}}\eta_n t^n\Big). \tag{4.7}$$

This mortality rate is equivalent to having a single node that corresponds to mortality, and death occurs when it damages (in contrast to the two nodes that were used in the GNM [11, 44]). The measure of local damage x that controls mortality is analogous to the local damage f_i for damage rates, and depends on each deficit linearly in a weighted sum. Additionally, it includes an age-dependent deficit-independent function represented as a power series (analogous to $\mu_i(t)$). This mortality rate uses fitting parameters $\{\alpha_n\}_{n=0}^{n_{D_1}}$, $\{\beta_j\}_{j=1}^{N}$, and $\{\eta_n\}_{n=0}^{n_{D_2}}$ as well as n_{D_1}, n_{D_2} as hyperparameters determining the number of terms in the power series. The coefficients in both Γ_D and x are constrained so that they increase monotonically in x and t, respectively.

In total the model has $N_{\text{tot}} = N(N + n_f + n_+ + 1) + n_{D_1} + n_{D_2} + 2$ fitting parameters. We restrict parameter values to ensure that f_i, Γ_i^+, Γ_D, and x are all monotonically increasing functions of age. Details of the requisite parameter bounds are in Appendix A. Despite the large number of parameters, we have many more individual observations. We also carefully test predictions for a test population that has small overlap of observed states with our training population (see Appendix Figure A.8). We find no evidence of overfitting.

The model is stochastically simulated by assuming the transition rates describe exponentially distributed waiting times between transitions, and then using an exact

event-driven stochastic simulation algorithm (SSA/Kinetic Monte Carlo) [144]. Details of the stochastic simulation are in the SI. For one run of the model until death, i.e. for each synthetic individual, the model outputs death age t_D and all node trajectories, $\{d_i(t)\}_{i=1}^{i=N}$ for all t between the initial age and mortality. Fully synthetic individuals are started at $t = 0$ with all $d_i(t = 0) = 0$, while predicted trajectories for observed individuals are initialized at some t_0 with the completely observed health state at t_0. Due to the exact nature of the SSA, all transition times are precisely resolved in our model data.

4.5.2 Likelihood

We calculate our likelihood using cross-sectional data. For the mth of M individuals, we have measurements of health attributes $\{d_i\}^{(m)}$ at age $t^{(m)}$. Instead of death age, we have an observed survival age $a^{(m)}$ due to right censoring. This is the oldest age that an individual is known to be alive, which can be written $a^{(m)} \equiv \min(t_D^{(m)}, t_c^{(m)})$, where $t_D^{(m)}$ is actual death age and $t_c^{(m)}$ is the censoring age i.e., the age of the individual when they are known to be still alive due to observed health state(s) but after which their mortality is not recorded. We indicate censoring with a binary variable $c^{(m)} = 1$, and uncensored with $c^{(m)} = 0$. In summary, we consider observed cross-sectional data of the form $\left\{t^{(m)}, \{d_i\}^{(m)}, a^{(m)}, c^{(m)}\right\}_{m=1}^{M}$.

By simulating synthetic individuals from the model, we sample and estimate the probability $p(\{d_i\}^{(m)}, t_D^{(m)}|t^{(m)}, \boldsymbol{\theta})$ for each individual m in the data. We denote all parameters by the vector $\boldsymbol{\theta}$. For simplicity we split this probability into two separate parts, representing mortality and health respectively:

$$\log p(\{d_i\}^{(m)}, t_D^{(m)}|t^{(m)}, \boldsymbol{\theta}) = \log p(t_D^{(m)}|\{d_i\}^{(m)}, t^{(m)}, \boldsymbol{\theta}) + \log p(\{d_i\}^{(m)}|t^{(m)}, \boldsymbol{\theta}). \quad (4.8)$$

For uncensored individuals, we can calculate their likelihood by using their known death age using Equation 4.8. For censored individuals, we also need to integrate the mortality term over all possible death ages above the censoring age,

$$S(a^{(m)}|t^{(m)}, \{d_i\}^{(m)}, \boldsymbol{\theta}) = \int_{a^{(m)}}^{\infty} p(a'|t^{(m)}, \{d_i\}^{(m)}, \boldsymbol{\theta}) da', \quad (4.9)$$

which is the probability of surviving to at least age a. Then we can calculate the full

log-likelihood,

$$\begin{aligned} L(\boldsymbol{\theta}) &= L^{\text{uncensored}}_{\text{mortality}} + L_{\text{health}} + L^{\text{censored}}_{\text{mortality}} \\ &= \sum_{m|c^{(m)}=0} \log p(a^{(m)}|\{d_i\}^{(m)}, t^{(m)}, \boldsymbol{\theta}) \\ &+ \sum_{m} \log p(\{d_i\}^{(m)}|t^{(m)}, \boldsymbol{\theta}) \\ &+ \sum_{m|c^{(m)}=1} \log S(a^{(m)}|t^{(m)}, \{d_i\}^{(m)}, \boldsymbol{\theta}), \end{aligned} \tag{4.10}$$

where the last term is added for censored individuals.

For an individual with missing data that does not have the full N health attributes measured, we marginalize over the missing values implicitly by sampling all possible combinations of the missing (binary) values. This is done using a synthetic population that has been initialized at $t = 0$ with no damage. Additional details of the likelihood estimation from simulations are in the SI.

4.5.3 Observed data

We use data from the Canadian Study of Health and Aging (CSHA) [160] to develop and test our model. The CSHA study used stratified sampling to be a representative sample of the older Canadian population. We use the first wave of the sample with 8547 individuals that range from ages $65-99$ and death ages that are available within a 6 year censoring window. The mean age is 76 years with a standard deviation of 7 years, the individuals are 60% female, and 78% of individuals have a censored death age. The 10 binary deficits used in the main plots are "Walking difficulty", "Showering difficulty", "Phone difficulty", "Going out difficulty", "Shopping difficulty", "Preparing meal difficulty", "House work difficulty", "Take medicine difficulty", "Managing money difficulty", and "Issues prevent normal activity". These were chosen by selecting deficits that had large hazard ratios in a Cox proportional hazards analysis [119], although any alternate sets of deficits can work, and an alternative set are shown in Appendix A.

We split the data into a training set of 1020 individuals and a test set of 7527 individuals. We do this such that dividing the training set into 5 year age bins has an approximately uniform age distribution, and the remaining individuals are put into

the test set. This balances the training set and ensures no age is "prioritized" in the model training by having a much larger number of individuals.

We validate our conclusions on the National Health And Nutrition Examination Survey (NHANES) [159]. The NHANES dataset used stratified sampling to be a representative sample of the US population. We use a combined sample from the 2003-2004 and 2005-2006 cohorts. The sample has 9504 individuals that range from ages $20 - 85$ with death ages that are available within a 10 year censoring window. The mean age is 51 years, with a standard deviation of 20 years, the individuals are 52% female, and 88% of individuals have a censored death age. In the same way as the CSHA data, this data is split into 2352 training individuals and 7151 test individuals.

4.5.4 Parameter optimization

For each data-set (and choice of health attributes), we maximize the log-likelihood in Equation 4.10 using particle swarm optimization [198] in order to train the model and estimate the parameters $\hat{\theta}$. Details of the parameter optimization procedure are in Appendix A. We use parameter bounds shown in the SI to impose monotonic dependence of damage rates on existing damage. We regularize the fitting as detailed in Appendix A. We choose hyperparameters $n_+ = 4$ and $n_f = n_{D_1} = n_{D_2} = 3$. These are the number of terms in our power-series expansions used in damage and mortality functions. The hyperparameters are hand chosen for simplicity. These hyperparameters result in a model with $N(N+8)+8$ parameters, where N is the number of binary health attributed modelled for each individual. Due to computational demands, this practically limits the size of N – here we take $N = 10$ and so have 188 parameters.

Chapter 5

Interpretable machine learning for high-dimensional trajectories of aging health

5.1 Background

Our previous work on the Weighted Network Model in Chapter 4 used an interaction network with 10 binary health variables. The model was fit with cross-sectional data that only had one measurement for each individual, along with mortality information. This approach was limited by computational power available, since the model involved discrete transitions we had to use slower non-gradient based optimization methods to fit the model. Additionally, the model required many simulations to approximate the likelihood function (known as simulation based inference [199, 200]).

In this chapter, I develop the Dynamic, Joint, Interpretable Network (DJIN) model of aging. Instead of building from the Generic Network Model as was done with the Weighted Network Model, we take a new approach. The main distinctions between the WNM and the DJIN model is the use of continuous-valued health variables as nodes in the network, and the use of a large longitudinal dataset to train the model.

In this model, the network consists of 29 continuous-valued health variables from the English Longitudinal Study of Aging (ELSA) [201], a large longitudinal study of aging with up to 8 follow-up measurements per individual. Additionally, the model is much more efficient allowing us to fit this model with more variables and many more individuals when compared to the WNM. The continuous dynamics used in the DJIN model let us use the stochastic gradient optimization techniques used in machine learning to greatly increase the speed of training the model, allowing us to train with large networks for a large number of individuals.

This model is similar to the Stochastic Process Model of Aging previously developed by Yashin *et al.* [33–36]. In this model, the observed health state $\mathbf{y}(t)$ is

described by a stochastic differential equation,

$$d\mathbf{y}(t) = \mathbf{W}(t)\big(\mathbf{y}(t) - \mathbf{f}(t)\big)dt + \boldsymbol{\sigma}(t) \cdot d\mathbf{B}(t), \tag{5.1}$$

where $\mathbf{f}(t)$ describes the baseline evolution of the health state, $\mathbf{W}(t)$ is a matrix describing the interactions between health variables, and $\boldsymbol{\sigma}(t)$ is the strength of the diffusive noise in the evolution of the health state.

Additionally, the hazard rate of mortality is described by "quadratic hazards",

$$h(t) = h_0(t) + (\mathbf{y}(t) - \mathbf{g}(t))^T \mathbf{Q}(t)(\mathbf{y}(t) - \mathbf{g}(t)), \tag{5.2}$$

where $h_0(t)$ is the baseline hazard, $\mathbf{g}(t)$ is the baseline evolution, and $\mathbf{Q}(t)$ is a matrix describing the interactions for mortality.

This *joint* model of both health and mortality can theoretically be used to model any number of health variables, however it has only used to model 1 or 2 variables. Additionally, only simple forms of the functions $\mathbf{W}(t), \mathbf{f}(t), \boldsymbol{\sigma}(t), \mathbf{g}(t), \mathbf{Q}(t)$ were considered, being either set constant or as linear functions of time.

The DJIN model represents an extension and advancement of this approach to modelling aging, using techniques from modern machine learning. This is done by scaling the approach up to incorporating many interacting health variables and large datasets, Bayesian inference to include uncertainty in parameters and predictions, and using flexible neural networks to learn the unknown functions involved in the model, except the interaction network $\mathbf{W}$. Instead we use a constant interaction network $\mathbf{W}$, which allows us to interpret the interactions between the health variables in terms of this network. Note, we do not attempt to interpret all aspects of the model, but rather use "black-box" neural networks for all components of the model except the interaction network.

This chapter presents the paper "Interpretable machine learning for high-dimensional trajectories of aging health" submitted for publication in 2021.

5.2 Introduction

Aging is a high-dimensional process due to the enormous number of aspects of healthy functioning that can change with age across a multitude of physical scales [4,52]. This complexity is compounded by the heterogeneity and stochasticity of individual aging

outcomes [53, 129]. Strategies to simplify the complexity of aging include identifying key biomarkers that quantitatively assess the aging process [57, 61] or integrating many variables into simple and interpretable one-dimensional summary measures of the progression of aging, as with "Biological Age" [18,153,202], clinical measures such as frailty [14,16], or recent machine learning models of aging [173,203]. Nevertheless, one-dimensional measures only summarize the progression of aging, and so can miss significant aspects of high-dimensional aging trajectories and of heterogeneous aging outcomes. We introduce a machine learning approach to model high-dimensional trajectories directly, while still learning interpretable aspects of our model through an explicit network of interactions between variables. We also compare this model with lower-dimensional modelling approaches, exploring the dimensionality required to model aging health outcomes.

The increasing availability of large longitudinal aging studies is beginning to provide the rich data-sets necessary for the development of flexible machine learning models of aging [23]. Methods for predictive modelling of individual health trajectories of disease progression have already been developed [193–196, 204, 205], but they generally are not joint models that include both mortality and the progression of aging [193]. There has also been progress on learning interpretable summaries of aging progression [173, 203], generalizing biological-age approaches but still producing low-dimensional summaries of aging.

Less progress has been made on the more general problem of modeling high-dimensional aging trajectories. Stochastic-process joint models that simultaneously model longitudinal and survival data have been proposed [33, 35, 36], but have only been implemented for one or two health variables at a time. Farrell *et al.* [48] used cross-sectional data to build a network model that generated trajectories of 10 health variables and predicted survival, but it was limited to binary health measures.

In this work we use the English Longitudinal Study of Aging (ELSA, [201]), which is a large observational population study including a wide variety of variables with follow-up measurements for up to 20 years including mortality. Like other large observational studies, for most individuals it has many missing measurements, few irregularly-timed follow-ups, and censored mortality. Any practical approach to model such data must confront the challenges provided by missing and irregularly

timed data and by mortality censoring.

While machine learning (ML) approaches can help us navigate these challenges with available data, they face additional challenges of interpretability [23, 206]. "Scientific Machine Learning" [24] or "Theory guided data science" [25] suggests that domain knowledge be used to constrain and add interpretability to ML models. For example, we can require that aging is modelled as a network of interacting health components [46, 47], and that stochastic differential equations (SDEs) model the dynamical evolution of high-dimensional health states [33]. On the other hand we can use general ML approaches to model survival or to impute missing data for baseline (initial) health states, where we may not be interested in interpretation.

The result (see Fig. 5.1) is a powerful and flexible, but interpretable, approach to modelling aging and mortality from high-dimensional longitudinal data – one that preserves but is not crippled by the complexity of aging. We evaluate the resulting model with test data and compare with simpler linear modelling approaches. We use a Bayesian approach to infer the posterior distribution of the both interaction network and individual health trajectories to estimate confidence bounds. We demonstrate our model's ability to robustly predict health trajectories using an interpretable network of interactions. Additionally, we demonstrate that low-dimensional latent variable models of a similar structure cannot predict aging health outcomes as well as this high-dimensional network model.

5.3 Results

5.3.1 ELSA dataset

We combine waves 0 to 8 in the English Longitudinal Study of Aging (ELSA, [201]) to build a dataset of $M = 25290$ individuals, with longitudinal follow-up of up to 20 years. In this study, self-reported health information is obtained approximately every 2 years and nurse-evaluated health with physical assessment and blood tests approximately every 4 years. Considering all waves together with 2 year increments, 27% of values are missing for self-reported variables, 78% of values are missing for nurse-evaluated variables, and 96% of individual mortality is censored. Training and test trajectories (see below) are sampled starting with baseline times starting at each

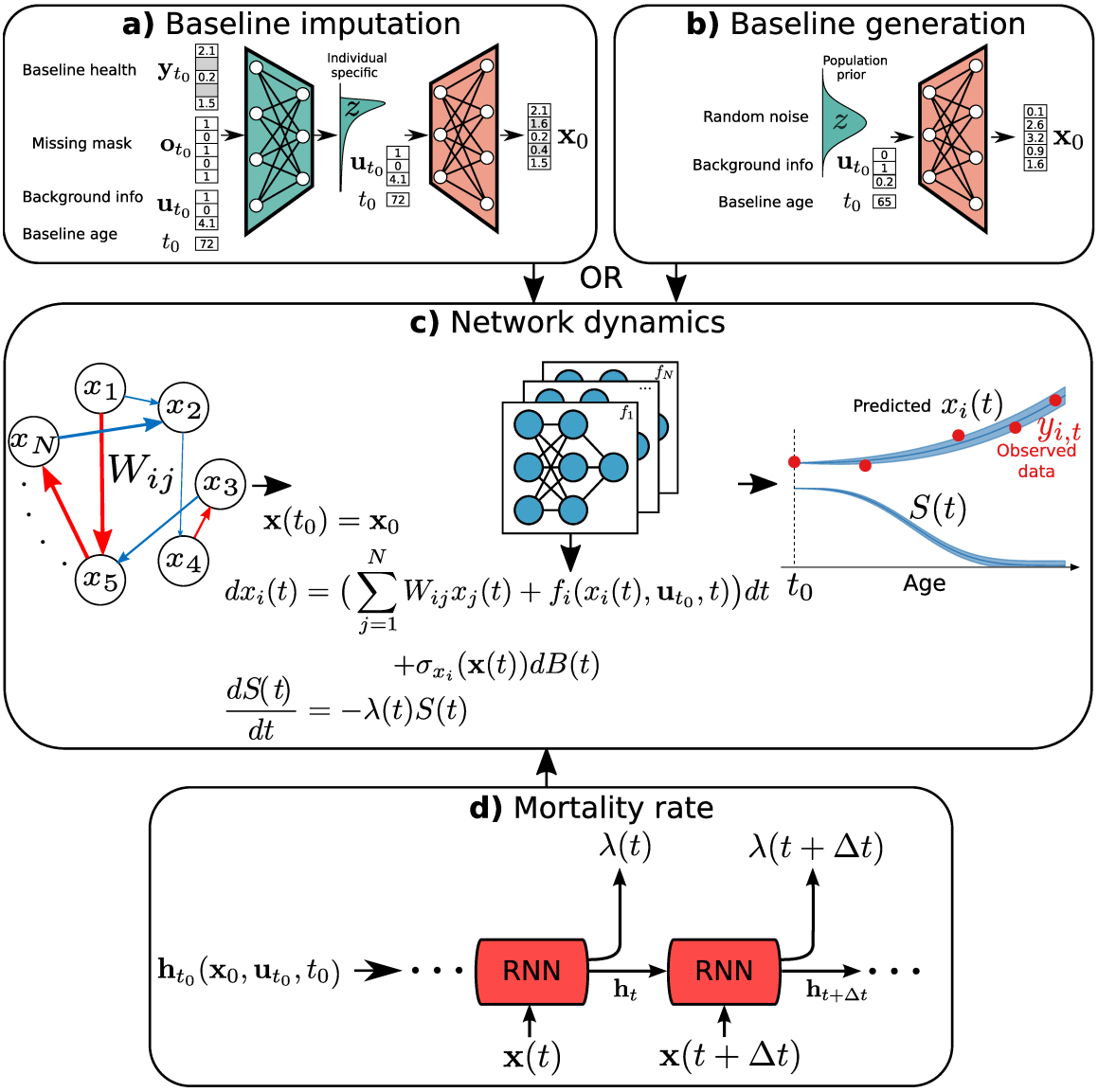


Figure 5.1: **DJIN model of aging.** (Caption on next page)

Figure 5.1: **DJIN model of aging. a)** Baseline imputation is performed using the baseline health measurement $\mathbf{y}_{t_0}$, missing mask $\mathbf{o}_{t_0}$, background health information $\mathbf{u}_{t_0}$, and baseline age t_0 as input to an encoder neural network (green) that parameterizes a latent distribution. Sampling from this latent distribution and using a decoder neural network (orange) gives an imputed complete baseline health-state $\mathbf{x}_0$. **b)** Baseline generation conditional on background health information $\mathbf{u}_{t_0}$, and baseline age t_0 can be used instead of imputation. The population latent distribution is sampled and used with the same decoder neural network (orange) to produce a synthetic baseline health state $\mathbf{x}_0$. **c)** Network dynamics stochastically evolve the health state $\mathbf{x}(t)$ in time starting from the baseline state $\mathbf{x}_0$. The stochastic dynamics are modeled with a stochastic differential equation which includes the pairwise network interactions with connection weight matrix $\mathbf{W}$, general diagonal terms $f_i(x_i(t), \mathbf{u}_{t_0}, t)$ parameterized as neural networks, and a diagonal covariance matrix for the noise $\sigma_x(\mathbf{x})$ also parameterized with a neural network. **d)** The survival function evolves in time based on the state and history of the health state $\mathbf{x}$ using a recurrent neural network (RNN). The initial state of the RNN, $\mathbf{h}_{t_0}$, is set using the background health information $\mathbf{u}_{t_0}$, baseline age t_0, and $\mathbf{x}_0$. Details are provided in the Methods. The code for our model is available at `https://github.com/Spencerfar/djin-aging`.

of the waves; though at least one followup wave is required for test trajectories.

For a given starting wave, an individual's health state is observed at $K+1$ times $\{t_k\}_{k=0}^{K}$ with a set of health variables $\{\mathbf{y}_{t_k}\}_{k=0}^{K}$. The vectors $\mathbf{y}_{t_k}$ describe the N-dimensional health state of an individual, where each of the N dimensions represents a separate health measurement. We select $N = 29$ continuous-valued or discrete ordinal variables that were measured for at least two of the waves. Individuals also have background (demographic, diagnostic, or lifestyle) information observed at baseline, which is described by a B-dimensional vector $\mathbf{u}_{t_0}$. We select $B = 19$ of these continuous or discrete valued background variables. These are used as auxiliary variables at baseline; they aide the subsequent prediction of the health variables $\mathbf{y}_t$ vs time.

Variables used from the data-set were selected only by availability, not by predictive quality. All chosen variables and the number of observed individuals for each is shown in Appendix Figure B.1, the details of the variables are given in Appendix Table B.1 and B.2.

5.3.2 DJIN model of aging

We build a model to forecast an individual's future health $\{\mathbf{y}_{t_k}\}_{k>0}$ and survival probability $\{S(t_k)\}_{k>0}$ given their baseline age t_0, baseline health $\mathbf{y}_{t_0}$ and background health variables $\mathbf{u}_{t_0}$. It is a dynamic, joint, interpretable network (DJIN) model of aging. A schematic of our model is shown in Fig. 5.1, while mathematical details are provided in the Methods.

Effective imputation is essential because none of the 25290 individuals in the dataset have a fully observed baseline health state. Fig. 5.1a illustrates our method of imputation for the baseline health state. Variational auto-encoders have shown promising results for imputation [207, 208]. We impute with a normalizing-flow variational auto-encoder [209], where a neural network (green trapezoid) encodes the known information about the individual into an individual-specific latent distribution, and a second neural network (orange trapezoid) is used to decode states sampled from the latent distribution into imputed values. This is a multiple imputation process that outputs samples from a distribution of imputed values rather than a single estimate.

We have chosen this imputation approach because we can also use it to generate totally synthetic baseline health states given background/demographic health information and baseline age. Fig. 5.1b illustrates this method. We randomly sample the prior population distribution of the same latent space used in imputation, and then combine this with arbitrary background information and use the same decoder as in imputation to transform the latent state into a synthetic baseline health state. With repeated random samples of the latent space, we generate a distribution of synthetic baseline health states.

Fig. 5.1c illustrates the temporal dynamics of the health state in the model. Dynamics start with the imputed or synthetic baseline state $\mathbf{x}_0$. The health state is then evolved in time with a set of stochastic differential equations, similar to the Stochastic Process Model of Yashin *et al.* [33–35, 210]. The stochastic dynamics capture the inherent stochasticity of the aging process. We assume linear interactions between the variables, with an interpretable interaction network $\mathbf{W}$. This interaction network describes the direction and strength of interactions between pairs of health variables.

Fig. 5.1d illustrates the mortality component of the model. The temporal dynamics of the health state is input into a recurrent neural network (RNN) to estimate the

individual hazard rate for mortality, which is used to compute an individual survival function. Recent work shows that this approach can work well in joint models [193]. The RNN architecture uses the history of previous health states in mortality, otherwise mortality could only depend on the current health state and could not capture the effects of a history of poor health. We have chosen this RNN approach to mortality because it performs better than either feed-forward (no history) or Cox proportional hazards models (as shown in Appendix Figure B.11).

We use a Bayesian approach to model uncertainty by estimating the posterior distribution of parameters, of health trajectories and of survival curves – as illustrated by the shaded blue confidence intervals in Fig. 5.1C. To handle our large and high-dimensional datasets, we use a variational approximation to the posterior [211] instead of slower MCMC methods. The variational approximation reduces the sampling problem to an optimization problem, which we can efficiently approach using stochastic gradient descent. Mathematical details are provided in the Methods. The code for our model is available at `https://github.com/Spencerfar/djin-aging`.

5.3.3 Validation of model survival trajectories

We evaluate our model with test individuals withheld from training. Given baseline age t_0, baseline health variables $\mathbf{y}_{t_0}$, and background information $\mathbf{u}_{t_0}$ for each of these test individuals, we impute missing baseline variables and predict future health trajectories and mortality with the model. These predictions are compared with their observed values.

The C-index measures the model's ability to discriminate between individuals at high or low risk of death. We use a time-dependent C-index [177], which is the proportion of distinct pairs of individuals where the model correctly predicts that individuals who died earlier had a lower survival probability. Higher scores are better; random predictions give 0.5. In Fig. 5.2a we see that our model (red circles) performs substantially better than a standard Cox proportional hazards model (green squares) with elastic net regularization and random forest MICE imputation [212, 213]. The horizontal lines show the C-index scores for the entire test set, and the points show predictions stratified by baseline age. Stratification allows us to remove age-effects

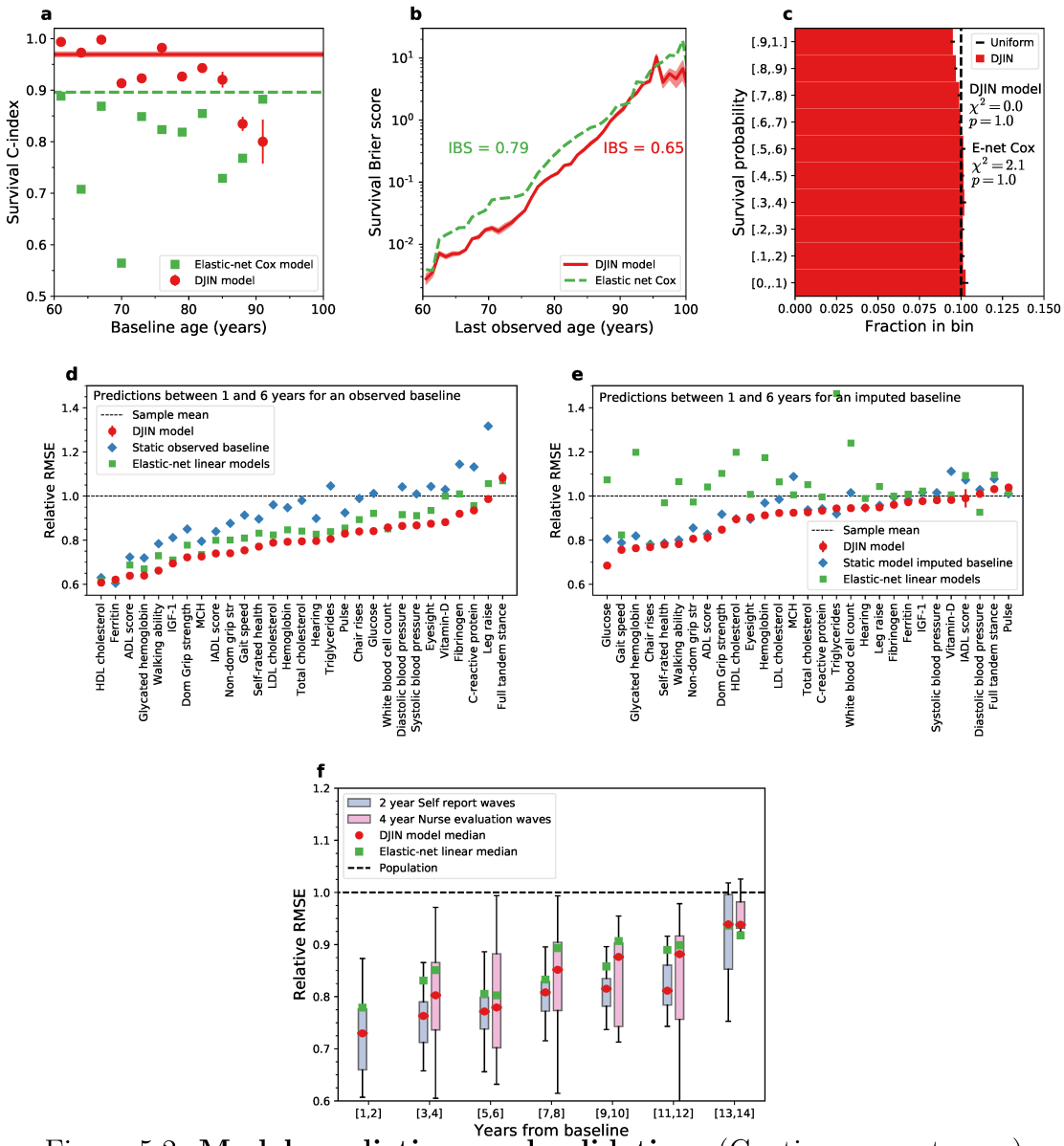


Figure 5.2: **Model predictions and validation.** (Caption on next page)

Figure 5.2: **Model predictions and validation.** Errorbars for all plots represent standard errors of the mean for 10 fits of the DJIN model (often smaller than point size). **a)** Time-dependent C-index stratified vs age (points) and for all ages (line). Results are shown for our model (red) and a Elastic net Cox model (green). (Higher scores are better). **b)** Brier scores for the survival function vs death age. Integrated Brier scores (IBS) over the full range of death ages are also indicated. The Breslow estimator for the baseline hazard is used for the Cox model. (Lower scores are better). **c)** D-calibration of survival predictions. Error bars show the standard deviation. Estimated survival probabilities are expected to be uniformly distributed (dashed black line). We use Pearson's χ^2 test to assess the distribution of survival probabilities for our network model ($\chi^2 = 0.0$, $p = 1.0$) and an elastic net Cox model ($\chi^2 = 2.1$, $p = 1.0$). (Higher p-values and smaller χ^2 statistics are better). **d)** RMSE scores when the baseline value is observed for each health variable for predictions between 1 and 6 years from baseline, scaled by the RMSE score from the age and sex dependent sample mean (relative RMSE scores). We show the predictions from our model starting the baseline value (red circles), predictions assuming a static baseline value (blue squares), and 29 distinct elastic-net linear models trained separately for each of the variables (green squares). The DJIN predictions here come from the same model as for mortality and the elastic net Cox model is also a distinct model. (Lower RMSE is better). **e)** Relative RMSE scores when the baseline value for each health variable is imputed for predictions between 1 and 6 years from baseline. We show the predictions from our model starting from the imputed baseline value (red circles), predictions assuming a static imputed value (blue squares), and predictions assuming an elastic-net linear model (green squares). (Lower RMSE is better). **f)** RMSE score distributions over all health variables for increasing years of prediction from baseline. The median RMSE score is shown as a black dotted line between the boxes showing upper and lower quartiles. Whiskers show 1.5x the interquartile range from the box. (Lower RMSE is better). Self-report and nurse-evaluated waves have distinct patterns of missing variables; nurse-evaluated waves have higher missingness overall.

in the predictions; we determine how well the model uses health variables to discriminate between pairs of individuals at the same age. Our model predictions do not substantially degrade when controlling for age, indicating that it is learning directly from health variables, rather than from age. Predictions degrade at older baseline ages due to the limited sample size.

We evaluate the detailed accuracy of survival curve predictions with the Brier score [178]. Individual Brier scores calculate squared error between the full predicted survival distribution $S(t)$ and the exact survival "distribution" for that individual, which is a step-function equal to 1 when the individual is alive and 0 when they are dead. Lower Brier scores are better, though the intrinsic variability of mortality will

provide some non-zero lower bound to the Brier scores. In Fig. 5.2b we show the average Brier score for different death ages for our model (blue) and a Cox model with a Breslow baseline hazard (green), indicating our model has a substantially lower error between the predicted and exact survival distributions for older ages (note the log-scale). The Integrated Brier Score (IBS) is computed by integrating these curves over the range of observed death ages, and highlights the improvement of predictive accuracy of our model as compared to Cox.

We evaluate the calibration of survival predictions with the D-calibration score [214]. For a well-calibrated survival curve prediction, half of the test individuals should die before their predicted median death age and half should live longer. Calibrated survival probabilities can be interpreted as estimates of absolute risk rather than just relative risk. The D-calibration score generalizes this to more quantiles of the survival curve, where the proportion observed in each predicted quantile should be uniformly distributed. In Fig. 5.2c, we show deciles of the survival probability for our model (red bins), compared with the expected uniform black straight line. We compute χ^2 statistics and p-values for the predictions of our model vs the uniform ideal, as well as for a Cox proportional hazards model (histogram in Appendix Figure B.12). Our model is consistent with a uniform distribution under this test ($p = 1.0$, $\chi^2 = 0.0$) as desired for calibrated probabilities. The Cox model is also calibrated ($p = 1.0$, $\chi^2 = 2.1$), but with a slightly worse χ^2 statistic.

These results demonstrate that our DJIN model accurately predicts the relative risk of mortality of individuals (assessed by the C-index), predicts accurate survival probabilities (assessed by the Brier score), and properly calibrates these survival probabilities so that they can be directly interpreted as an absolute risk of death.

5.3.4 Validation of model health trajectories

Model predictions of individual health trajectories are also evaluated on the test set. We compute the Root-Mean-Square Error (RMSE) for each health variable, and create a relative RMSE score by dividing by the RMSE obtained when using the age and sex matched training-set sample mean as the prediction. In Fig. 5.2d, we show that the model (red circles) performs better than the age and sex dependent

sample mean (black dashed line) when the baseline value of the particular variable is observed. The RMSE here is computed for all predictions between 1 and 6 years from baseline. In Fig. 5.2e we show that the model is predictive of future health values even when the initial value of the particular variable is imputed.

As measured by the relative RMSE, our model is better than a null model (blue squares) that carries forward the observed baseline (**d**) or imputed baseline value (**e**) for all ages. For comparison purposes, we also trained linear models with elastic net regularization and random-forest MICE imputation [212, 213] that have been trained separately to predict each health variable. We are therefore comparing our single DJIN model that predicts all 29 variables, to 29 independently-trained linear models. While the linear models perform better than the null model for observed baselines, our model performs better than both. For imputed baselines, the linear models with random-forest MICE imputation performs poorly even compared to the imputed null model, while our model continues to outperform both.

In Fig. 5.2f, we show boxplots of RMSE scores over the health variables for 1-14 years past baseline, when the variable was initially observed at baseline. The model is predictive for long term predictions, and remains better than linear elastic net predictions for at least 14 years past baseline for the self-report waves (blue) and 12 years past baseline for the nurse-evaluated waves including blood biomarkers (pink).

In Appendix Figure B.3 we show example DJIN trajectories for 3 individuals in the test set for the 6 best predicted health variables. We show both the mean predicted model trajectory and a visualization of the uncertainty in the trajectory. For comparison, the sample mean and elastic net linear model are shown. The predicted trajectories visually agree well with the data, and is often substantially better than either the elastic net linear predictions or the sample means for the corresponding variables.

These results demonstrate that our DJIN model predicts the values of future health variables from baseline better than standard linear models, and also better than sample-mean or constant baseline models.

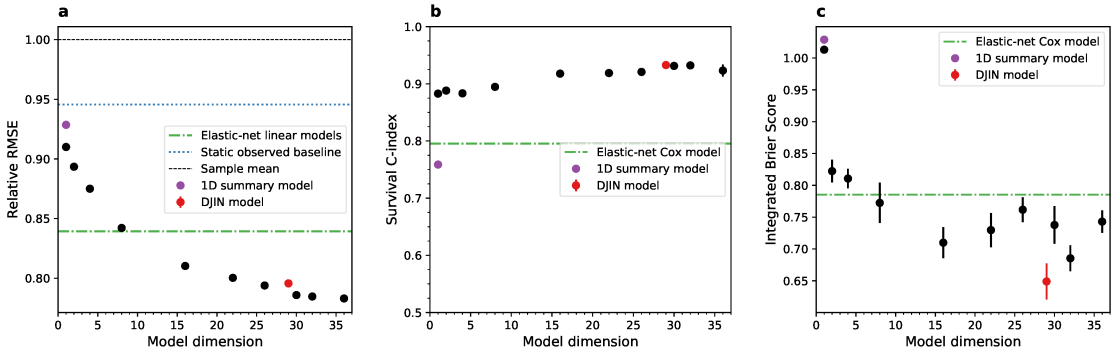


Figure 5.3: **Latent variable model performance vs dimension.** Black points so latent variable models of various dimension, red points show the DJIN model, green lines show the elastic net linear models. The purple point shows the 1D summary model described in Appendix B, the includes the information from the auxiliary background variables $\mathbf{u}_{t_0}$ within the latent state. Black points include $\mathbf{u}_{t_0}$ in addition to $\mathbf{z}$. Points indicate the mean of 10 independent fits of the model, and error bars represent standard error of mean (often smaller than point size). **a)** RMSE for health predictions, relative to predictions with the sample mean (Set to value 1, black dashed line). (Lower is better). **b)** Survival C-index. (Higher is better). **c)** Integrated brier score for survival. (Lower is better).

5.3.5 Comparison with latent variable models

In Figure 5.3, we compare the DJIN model to latent variable alternatives. While the DJIN model uses dynamics on the observed health variables, these alternatives use dynamics on latent variables. In the model diagram in Figure 5.1, this latent variable model uses dynamics on the latent state output by the encoder $\mathbf{z}$. The dynamics of the latent variable use a full feed-forward neural network for the drift of the SDE for the dynamics. These predicted latent trajectories $\mathbf{z}(t)$ are decoded into predicted observed health states $\mathbf{x}_t$ with the decoder. Details of this model are in the Methods section. These latent variables can be any dimension, and so can compress the information to a lower-dimensional state, but lack the network interpretation of the DJIN model.

Figure 5.3a shows that a large number of dimensions are required to predict health outcomes. A much lower dimension can be used to predict survival, show in Figures 5.3b and 5.3c). However, full predictions of the survival probability are poor for the latent variable models, seen in 5.3c. Note that the Brier score of the latent variable models was very noisy, varying strongly between fits. This may suggest that these

latent variable models are overfitting survival, and can improve with hyperparameter tuning.

These results suggests that work on low-dimensional summary measures of aging such as biological age can capture the progression of aging, but can not predict the specific heterogeneous health outcomes during aging [21,215,216]. In Appendix Figure B.9, we show in greater detail the one-dimensional summary model and in Appendix Figure B.10 the 30-dimensional latent variable model.

Our network model that only includes pair-wise linear interactions performs similarly to high-dimensional latent variable models that use non-linear dynamics. This suggests that our network approximation is sufficient for describing the dynamics of these variables. One reason this linear pair-wise network approximation may work well is because we are interested in long term predictions, rather than short-time scale dynamics where the variables are likely more strongly non-linearly coupled.

5.3.6 Validation of generated synthetic populations

Given baseline age t_0, and background information $\mathbf{u}_{t_0}$ for test individuals, we generate synthetic baseline health states and simulate a corresponding synthetic aging population. We evaluate these aging trajectories by comparing with the observed test sample. We train a logistic regression classifier to evaluate if the synthetic and observed populations can be distinguished [196, 205, 217, 218]. We find that this classifier has below a 57% accuracy for the first 14 years past baseline in Figure 5.4 – only slightly better than random. Additionally, we show that this approach is equivalent or better than the non-linear latent variable models.

In Appendix Figures B.5 and B.6 we show the population and synthetic baseline distributions and population summary statistics for the trajectories vs age for ages 65 to 90. We find that our model captures the mean of the population, but slightly underestimates the standard deviation of the population (as expected due to our variational approximation of the posterior [211]). In Appendix Figure B.4 we show the population synthetic survival function agrees with the observed population survival below age 90, where the majority of data lies.

The agreement of the synthetic and test populations demonstrates the DJIN

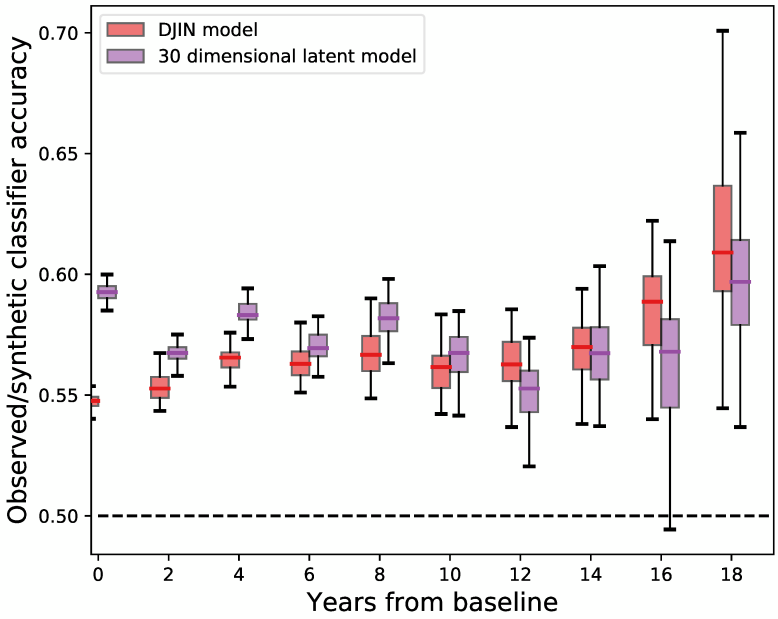


Figure 5.4: **DJIN generates synthetic populations.** We use a logistic regression classifier to evaluate the quality of our generated synthetic population by the classifier's ability to differentiate the synthetic population from the observed population. The boxplot shows the median with the horizontal lines, interquartile range with the box, and 1.5x from the interquartile range with the whiskers. Completely indistinguishable natural and synthetic populations would have a classification accuracy of 0.5. We show the classification accuracy vs years from baseline, showing low classification accuracies that increase slowly with time from baseline in the DJIN model, and the DJIN model is equivalent or better than non-linear latent variable models.

model's ability to generate a synthetic population of aging individuals that resemble the observed population, though with slightly less variation.

5.3.7 DJIN infers interpretable sparse interaction networks

Our Bayesian approach infers the approximate posterior distribution of the interaction network weights; Fig. 5.5 visualizes the network with the mean posterior weights. Weights with a 99% posterior credible interval including zero have been pruned (white) – all visible weights have posterior credible intervals either fully above or fully below zero. This cutoff is demonstrated in Appendix Figure B.8.

Connections are read as starting at the variable on the horizontal axis (j), and ending at the variable on the vertical axis (i), representing the connection weight

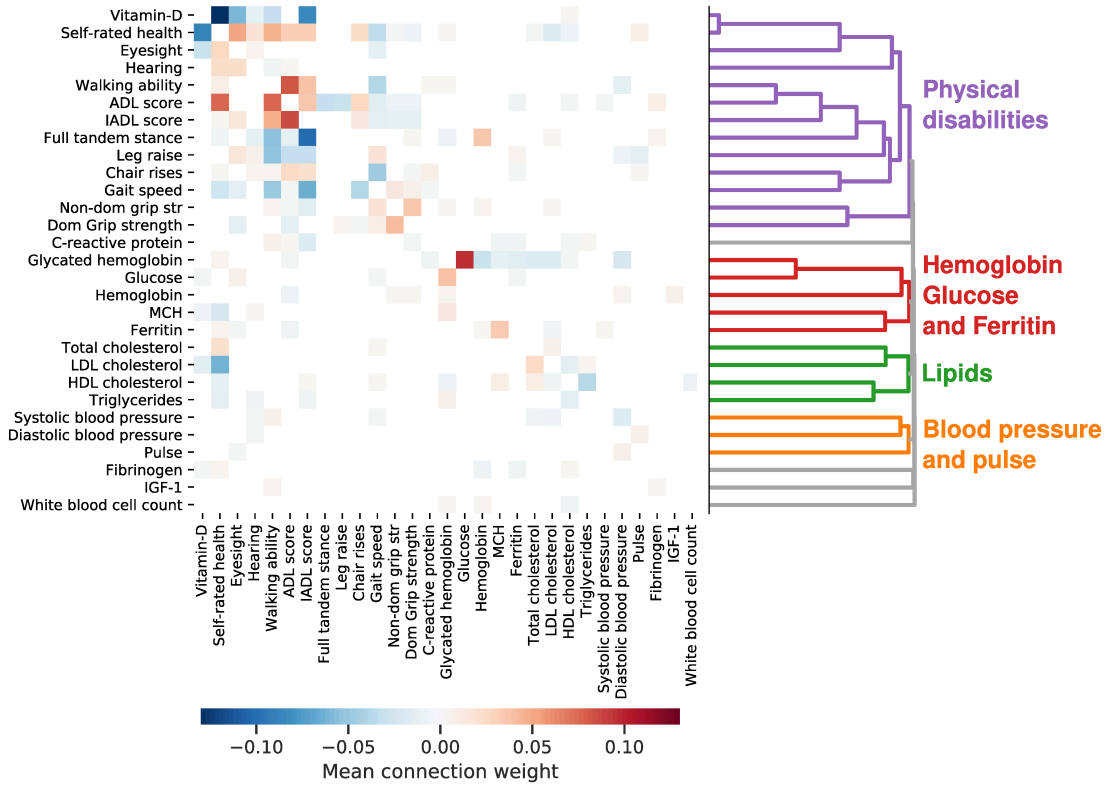


Figure 5.5: **Inferred interaction network.** Heatmap of the posterior mean value of the robust network weights. Weight directions are read from the horizontal axis (j) towards the vertical (i), $W_{i \leftarrow j}$. The sign and color of the weight signify the direction of effect – a positive weight implies that an increase in a variable along the horizontal axis influences the vertical axis variable to increase. A negative weight implies that an increase in a variable along the horizontal axis influences the vertical axis variable to decrease. Hierarchical clustering is applied to the absolute posterior mean of the robust weights to create a dendrogram (at right).

matrix $W_{i \leftarrow j}$. Positive connections indicate that an increasing variable j influences an increase in variable i. Negative connections mean an increasing variable j influences a decrease in variable i. The interaction network is sparse, with typically only a small number of inferred interactions for each health variable.

This inferred causal network can be readily and directly interpreted. For example, we see strong connections between Vitamin-D and self-rated health, between activities of daily living (ADL) score and walking ability, and between glucose and glycated hemoglobin. The sign of the connections indicates the direction of influence. For example, a decrease in gait speed influences an increase in self-reported health score (worse health), an increase in the time required to complete chair rises, and a decrease

in grip strength.

Hierarchical clustering on the connection weights is indicated in Fig. 5.5, and the ordering of the variables in the heatmap represents this hierarchy. Many of these inferred clusters of nodes plausibly fit with known physiology. For example, most blood biomarker measurements (bottom half) are separated from the physical/functional measurements (top half, purple cluster). Other inferred clusters include blood pressure and pulse (orange) and lipids (green).

5.4 Discussion

We have developed a machine learning aging model, DJIN, to predict both multidimensional health trajectories and survival given baseline information, and to generate realistic synthetic aging populations – while also learning interpretable network interactions that characterize the dynamics in terms of realistic physiological interactions. The DJIN model uses continuous-valued health variables from the ELSA dataset, including physical, functional, and molecular variables. We have shown that the comprehensive DJIN model performs better than 30 independent regularized linear models that were trained specifically for each separate health variable or survival prediction task.

Using a latent-variable model approach, we analyzed the multi-dimensionality of aging. While summary measures of aging such as the frailty index or biological age are easily interpretable, they are only one-dimensional. We find that the changes that occur due to aging are multidimensional. This is due to the heterogeneity in individual aging trajectories – individuals with the same frailty index or biological age can have distinct health states.

Previously, we had built a weighted network model (WNM) using cross-sectional data with only binary health deficits [48]. That WNM did not incorporate continuous health variables and could not be efficiently trained with longitudinal data. As a result, the networks inferred by that model were not robust – and resulted in many qualitatively distinct networks that were all consistent with the observed cross-sectional binary data. In contrast, the DJIN model uses many continuous valued health variables and can be efficiently trained with large longitudinal datasets. As a result, the DJIN model produces a robust and interpretable interaction network of

multidimensional aging (Appendix Figure B.13 shows robustness between separate fits of the model).

Recently, other machine learning models of aging or aging-related disease progression have been emerging [173, 193, 196, 205, 218]. Since they each differ significantly in terms of both the datasets, types of data used, and scientific goals, it is still too early to see which approaches are best – and for which data and what goals. We use ELSA data since it is longitudinal (to facilitate modelling trajectories), has many continuous variables (to allow modelling of continuous trajectories and constructing an interaction network that is at the core of our model), and includes mortality (to develop our joint mortality model). The ELSA data is representative of many large-scale aging data sets.

Our scientific goals were to obtain good predictive accuracy from baseline for both health trajectories and mortality, while at the same time obtaining an interpretable network of interactions between health variables [23]. To achieve these goals with the ELSA data, we had to do significant imputation to complete the baseline states. We include stochastic dynamics within a Bayesian model framework to obtain uncertainties for both our predictions and the interaction network. The Bayesian approach is computationally intensive and necessitated a variational approximation to the posterior that tends to underestimate uncertainty [211]. From the analysis of synthetic populations (see Appendix Figures B.4, B.5, and B.6), this underestimate appears to be modest.

The DJIN model is not computationally demanding, needing only approximately 12 hours to run with 1 GPU for $M = 25290$ individuals, $B = 19$ background variables, $N = 29$ health-variables, and up to $K = 20$ years of longitudinal data. We expect better performance and generalizability with more individuals M. Because of the interactions between health variables we also expect better performance with more health variables N. We note that binary health variables, or mixtures of binary and continuous variables, could be used with only small adjustments. Since computational demands for a forward pass of the model scale approximately linearly with M and K, and quadratically with $B + N$, our existing DJIN model is already practical for significantly larger datasets.

In this work we only consider predictions from the baseline state at a single age.

We anticipate that individual prediction could be significantly improved by utilizing more than one input time to impute the baseline health state $\mathbf{x}_0$ or by conditioning the predictions on multiple input ages. This conditioning can be done using a recurrent neural network [219,220]. Observed time-points after baseline can also be used to update the dynamics [221] for predictions of continually observed individuals in personalized medicine applications. However, both of these developments would either require data with more follow-up times than we had available, or limiting predictions to very short time intervals. For these reasons, we chose to model trajectories using only a single baseline health state.

We developed an imputation method that is trained along with the longitudinal dynamics to impute missing baseline data. This imputation method can also be used to generate synthetic individuals conditioned on baseline demographic information. Large synthetic datasets can facilitate the development of future models and techniques by providing high-quality training data [222], and are especially needed given the lack of large longitudinal studies of aging [23]. In Figure 5.4 and Appendix Figures B.4, B.5, and B.6 we show that our synthetic population is comparable to the available individuals in the ELSA dataset. We have also provided a synthetic population of nearly 10^7 individuals with annually sampled trajectories from baseline for 20 years [223].

At the heart of our dynamical model is a directed and signed network that is directly interpretable. The DJIN model does not just make "black-box" predictions, but is learning a plausible physiological model for the dynamics of the health variables. The network is not a correlation/association network (see comparison in Appendix Figure B.7) [153, 154, 191], but instead determines how the current value of the health variables influence future changes to connected health variables, leading to coupled dynamics of the health variables. This establishes a predictive link between variables [224]. Similar directed linear networks are inferred in neuroscience with Dynamic Causal Modelling [225,226]. While previous work on learning networks for discrete stochastic dynamics has been done in the past [227–229], we have used continuous dynamics here. When interpreting the magnitude of weights, links function as in standard regression models: weight magnitudes will be dependent on the variables included in the model, and can decrease if stronger predictor variables are

added. Given the efficiency of our computational approach, including more health or background variables is recommended if they are available.

The directed nature of the network connections lend themselves to clinical interpretation – for example that ADL impairment has an effect on instrumental ADL (IADL) impairment and not vice versa, and that both have an effect on general function score and vice versa. The directed network of interactions suggests avenues to explore for interventions. For a given intervention (for example drug, exercise, or diet) we can ascertain which effects of the intervention are beneficial and which are deleterious. In principle, we could also predict the outcome of multiple interventions such as in polypharmacy [230]. A similar approach could be taken for chronic diseases or disorders. While static interventions could simply be included as background variables, our DJIN model could also easily be adapted to allow for time-dependent interventions. These avenues will be increasingly feasible and desirable with longitudinal 'omics data-sets, where the interactions are not already largely determined by previous work.

We caution that our model does not currently take into account how interventions may change relationships over time, or any higher order interactions than the pairwise interactions considered here. For example, the interaction between sodium levels, mobility, and diuretics appears to be strong [231], but would not be captured in our current model. Extending our approach to include such effects in the interactions would be desirable.

The accuracy of our model can be slightly improved if a network interpretation of the dynamics is not desired – for instance if the goal is only prediction. High-dimensional non-linear latent variable models that using a neural network for the drift function instead of pair-wise network interactions somewhat improve health variable prediction accuracy. However, our goal was to demonstrate both good predictions *and* interpretability. The network form of our dynamics is not the only type of dynamics that can be used, and can be replaced with any alternative model of the aging dynamics for bespoke interpretability.

The advantage of more interpretable models will be more clearly seen when multiple data-sets are compared – since interpretability facilitates comparisons between cohorts, groups, or even between model organisms. Every aspect of our DIJN model

can be made more structured, explicit and "interpretable". For example, proportional hazards [119] or quadratic hazards [33] could be used for mortality. While this would reduce performance compared to our more general neural network, it would add interpretability to the survival predictions.

Our work opens the door to many possible follow-up studies. Our DJIN model can be applied to any organism or set of variables that has enough individual longitudinal measurements. With genomic characterization of populations, the background health information $\mathbf{u}_{t_0}$ can be greatly expanded to examine how the intrinsic variability of aging [53,129] and mortality are affected by genetic variation. By including genomic, lab-test, and functional data we could use the interpretable interactions to determine how different organismal scales of health data interact in determining human aging trajectories. By including drug and behavioral (exercise, diet) interventions as background variables, we can better determine how they affect health during aging. Finally, large longitudinal multi-omics datasets [59,60] could be used to build integrative models of human health.

We have demonstrated a viable interpretable machine learning (ML) approach to build a model of human aging with a large longitudinal study that can predict health trajectories, generate synthetic individual trajectories, and learn a network of interactions describing the dynamics. The future of these approaches is bright [23], since we are only starting to embrace the complexity of aging with large longitudinal datasets. While ML models can find immediate application in understanding patterns of aging health in populations, we foresee that similar techniques will eventually reach into clinical practice to guide personalized medicine of aging health.

5.5 Methods

5.5.1 ELSA dataset

We use waves 0-8 of the English Longitudinal study of Aging (ELSA) dataset [201], with 25290 total individuals. We include both original and refreshment samples that joined the study after the start at wave 0. In Appendix Table B.1 we list all variables used. In Appendix Figure B.1, we show the number of individuals for which the variable is available at different times from their entrance wave. Each available wave

is used as a baseline state for each individual, see section 5.5.2 for details.

We extract 29 longitudinally observed continuous or discrete ordinal health variables (treated as continuous) and 19 background health variables (taken as constant with age). We set the gait speed of individuals with values above 4 meters per second to missing, due to a likely data error. Sporadic missing ages are imputed by assuming the age difference between waves is 2 years – the time difference in the design of the study.

Individuals above age 90 in the ELSA dataset have their age privatized. By assuming the time difference between waves is 2 years, we "deprivatize" these ages within our analysis pipeline. For example, an individual may have recorded ages $87, 89, \langle privatized \rangle, \langle privatized \rangle$, which we deprivatize as $87, 89, 91, 93$. When individuals are known to die at an age above 90 at a specific wave, the same approach is done to deprivatize the death age. We have examined the accuracy of reported ages compared to this fixed two-year wave interval deprivatization method, and we find that the majority of deviations range from 0-1 years (with 78% at 0 years, and an average deviation of 0.23 years) – we expect similar variability for deprivatized ages above 90.

Height is imputed with the last observation carried forward (if it is missing, the first value is carried backwards from the first available measurement). Individuals with no recorded death age are considered censored at their last observed age.

The data is randomly split into separate train (16689 individuals), validation (4173 individuals), and test sets (5216 individuals). The training set is used to train the model, the validation set is used for control of the optimization procedure during training (through a learning rate schedule, see Section 5.5.6 below), and the test set is used to evaluate the model after training. Individuals with fewer than 6 variables measured at the baseline age t_0 are dropped from the training and validation data. Individuals with fewer than 10 variables measured at the baseline age t_0 are dropped from the test data for predictions, while all individuals in the test data are used for population comparisons.

All variables are standardized to mean 0 and standard deviation 1 (computed from the training set); however variables with a skewed age-aggregated distribution $p(\mathbf{y})$ covering multiple orders of magnitude are first log-transformed. Log-scaled variables

are indicated in Appendix Table B.1.

5.5.2 Data augmentation

Since some health variables are measured only at specific visits, using the entrance wave as the only baseline of every individual forces some variables to be rarely observed at baseline, hindering imputation of variables that are only observed at later waves. To mitigate this and to augment the dataset, we replicate individuals to use all possible starting points, $t_k^{(m)}, k \in \{0, ..., \text{argmax}_k(t_k^{(m)})\}$. Since individuals have different numbers of observed times we weight individuals in the likelihood who have multiple times available by $s^{(m)} = 1/(\text{argmax}_k(t_k^{(m)}) + 1)$. This helps to prevent the over-weighting of individuals with many possible starting times. Nevertheless, we still assume that replicated individuals are independent in the likelihood.

To further augment the available data, we artificially corrupt the input data for training by masking each observed health variable at baseline with probability 0.9. This allows more distinct "individuals" for imputation of the baseline state, and allows us to use self-supervision for these artificially missing values by training to reconstruct the artificially corrupted states.

5.5.3 DJIN model

We model the temporal dynamics of an individual's health state with continuous-time stochastic dynamics described with stochastic differential equations (SDEs). These SDEs include linear pair-wise interactions between the variables to form a network with a weight matrix $\mathbf{W}$. We assume the observed health variables $\mathbf{y}_t$ are noisy observations of the underlying latent state variables $\mathbf{x}(t)$, which evolves according to these network SDEs. This allows us to separate measurement noise from the noise intrinsic to the stochastic dynamics of these variables.

These SDEs for $\mathbf{x}(t)$ start from each baseline state $\mathbf{x}_0$, which is imputed from the available observed health state $\mathbf{y}_t$. This imputation process is done using a normalizing-flow variational auto-encoder (VAE) [209]. In this approach, we encode the available baseline information into a latent distribution for each individual, and decode samples from this distribution to perform multiple imputation. The normalizing-flow VAE allows us to flexibly model this latent distribution. The details

are described in Section 5.5.4 below.

Our model is described by the following equations:

$$\boldsymbol{\theta} = \big\{\mathbf{W}, \sigma_{\mathbf{y}}, \sigma_{\mathbf{x}}, \boldsymbol{\theta}_{\lambda}, \boldsymbol{\theta}_{p}, \boldsymbol{\theta}_{f}\big\}, \tag{5.3}$$

$$\mathbf{z}, \boldsymbol{\theta} \sim p(\mathbf{z})p(\boldsymbol{\theta}), \tag{5.4}$$

$$\mathbf{x}_0 = \mathbf{o}_{t_0} \odot \mathbf{y}_{t_0} + (1 - \mathbf{o}_{t_0}) \odot \tilde{\mathbf{x}}_0, \;\; \tilde{\mathbf{x}}_0 \sim \mathcal{N}(\mathbf{x}_0 | \boldsymbol{\mu}_{\mathbf{x}}(\mathbf{z}, \mathbf{u}_{t_0}, t_0, \boldsymbol{\theta}_p), \boldsymbol{\sigma}_{\mathbf{y}}^2), \tag{5.5}$$

$$dx_i(t) = \Big(\sum_{j=1}^{N} W_{ij} x_j(t) + f_i(x_i(t), \mathbf{u}_{t_0}, t; \boldsymbol{\theta}_{f_i}) \Big) dt + \sigma_{x_i}(\mathbf{x}(t)) dB(t), \;\; \mathbf{x}(t_0) = \mathbf{x}_0, \tag{5.6}$$

$$\mathbf{y}_t \sim \mathcal{N}(\boldsymbol{\psi}^{-1}(\mathbf{x}(t)), \mathrm{diag}(\sigma_{\mathbf{y}}{}^2)), \tag{5.7}$$

$$S(t) = \exp\big(- \int_{t_0}^{t} \lambda(\{\mathbf{x}(\tau)\}_{\tau \le t'}, \mathbf{u}_{t_0}, t'; \boldsymbol{\theta}_{\lambda}) dt' \big), \tag{5.8}$$

$$a \sim \lambda(\{\mathbf{x}(\tau)\}_{\tau \le a}, \mathbf{u}_{t_0}, a; \boldsymbol{\theta}_{\lambda}) S(a), \tag{5.9}$$

$$p(\mathbf{z}, \{\mathbf{x}(t)\}_t, \boldsymbol{\theta} | \{\mathbf{y}_{t_k}\}_k, \mathbf{u}_{t_0}, \{\mathbf{o}_{t_k}\}_k, t_0, a, c) \propto p(\boldsymbol{\theta})p(\mathbf{z})p(\mathbf{x}_0|\mathbf{z}, \mathbf{u}_{t_0}) \times \tag{5.10}$$
$$p(\{\mathbf{x}(t)\}_t | \mathbf{x}_0, \mathbf{u}_{t_0}, t_0, \boldsymbol{\theta}) p(a, c | \{\mathbf{x}(t)\}_t, \mathbf{u}_{t_0}, t_0, \boldsymbol{\theta}) \prod_{k=0}^{K} p(\mathbf{y}_{t_k} | \{\mathbf{x}(t_k)\}_k, \mathbf{o}_{t_k}, \boldsymbol{\theta}),$$

Model parameters (θ) include the explicit network of interactions between health variables (W), measurement noise (σ_y) and dynamical SDE noise (σ_x), and network weights for mortality RNN (θ_λ), imputation VAE decoder ($\boldsymbol{\theta}_p$), and dynamical SDE ($\boldsymbol{\theta}_f$). Equation (5.4) represents priors on the model parameters and latent state $\mathbf{z}$. We use Laplace($\mathbf{w}|0, 0.05$) priors for the network weights and Gamma($\sigma_{\mathbf{y}}|1, 25000$) priors for the measurement noise scale parameters. We use a normal (Gaussian) prior distribution for the latent space $\mathbf{z}$. We assume uniform priors for all other parameters.

In Equation (5.5) we sample the baseline state. The distribution for $\mathbf{x}_0$ given $\mathbf{z}$ is modeled as Gaussian with mean computed from the decoder neural network and the same standard deviation as the measurement noise, $\mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}(\mathbf{z}, \mathbf{u}_{t_0}, t_0; \boldsymbol{\theta}_p), \boldsymbol{\sigma}_{\mathbf{y}}^2)$. The missing value imputation and the dynamics model are trained together simultaneously (see details below). This allows us to utilize the additional longitudinal information for training the imputation method, and helps to avoid an imputed baseline state that leads to poor trajectory or survival predictions.

Equation (5.6) describes the SDE network dynamics, starting from the imputed baseline state. We capture single-variable trends with the non-linear

$f_i(x_i(t), \mathbf{u}_{t_0}, t; \boldsymbol{\theta}_{f_i})$, and couple the components of $\mathbf{x}(t)$ linearly by the directed interaction matrix $\mathbf{W}$, which represents the strength of interactions between the health variables. In this way, f_i can be thought of as a non-linear function for the diagonal components of this matrix, while $\mathbf{W}$ gives linear pair-wise interactions for the off-diagonal components. The intrinsic diffusive noise in the health trajectories is modeled with Brownian motion with Gaussian increments $d\mathbf{B}(t)$ and strength $\boldsymbol{\sigma}_{\mathbf{x}}$. The functions f_i and $\boldsymbol{\sigma}_{\mathbf{x}}$ are parameterized with neural networks.

Equation (Health observation) describes the Gaussian observation model for the observed health state. Measurement noise here is separate from diffusive noise $d\mathbf{B}(t)$ in the SDE from Equation (5.6). The component-wise transformation $\boldsymbol{\psi}$ applies a log-scaling to skewed variables (indicated in Appendix Table B.1) and z-scores all variables.

Equation (5.8) describes the survival probability as computed with a recurrent neural network (RNN) for the mortality hazard rate λ. The RNN allows us to use the stochastic trajectory for the computation of the hazard rate (i.e. it has some memory of health at previous ages), rather than imposing a memory-free process where the hazard rate only depends on the health state at the current age. We use a 2-layer Gated Recurrent Unit (GRU [112]) for the RNN, with hidden state $\mathbf{h}_t$. The initial hidden state $\mathbf{h}_0$ is inferred from the initial health variables $\mathbf{x}(t_0)$, background health information $\mathbf{u}_{t_0}$, and baseline age t_0, with a neural network $\mathbf{h}_0 = H(\mathbf{x}(t_0), \mathbf{u}_{t_0}, t_0)$. Equation (5.9) describes the observation model for survival with age of death or last age known alive $a = \max(t_d, t_c)$, and censoring indicator c.

Instead of just a maximum likelihood point estimate of the network and other parameters of the model, we use a Bayesian approach. This is a natural approach for this model, since the stochastic dynamics of $\mathbf{x}(t)$ are separate from the noisy observations $\mathbf{y}_t$. This also allows us to infer the posterior distribution of the health trajectories and interaction network, and so lets us estimate the robustness of the inferred network and the distribution of possible predicted trajectories, given the observed data. In Equation (5.10) we show the form of the unnormalized posterior distribution.

5.5.4 Variational approximation for scalable Bayesian inference

While sampling based methods of inference for SDE models do exist [232, 233], these are generally not scalable to large datasets or to models with many parameters. Instead, we use an approximate variational inference approach [234, 235]. We assume a parametric form of the posterior that is optimized to be close to the true posterior. While variational approximations tend to underestimate the width of posterior distributions and simplify correlations, they typically capture the mean [211].

Our factorized variational approximation to the posterior in Equation (5.10) is

$$q(\mathbf{z}, \mathbf{x}(t), \boldsymbol{\theta} | \mathbf{y}_0, \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, t_0, \boldsymbol{\phi}) = q(\mathbf{z}|\mathbf{y}_0, \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, t_0, \boldsymbol{\phi}_z) \tag{5.11}$$
$$\times\, q(\mathbf{x}(t)|\mathbf{x}_0, \mathbf{u}_{t_0}, t, \boldsymbol{\phi}_x) q(\boldsymbol{\theta}|\boldsymbol{\phi}_\theta),$$
$$\{\mathbf{x}(t)\}_t \sim q(\mathbf{x}(t)|\mathbf{x}_0, \mathbf{u}_{t_0}, t, \boldsymbol{\phi}_x) \implies \tag{5.12}$$
$$d\mathbf{x}(t) = \Big(\bar{\mathbf{W}}\mathbf{x} + \mathbf{f}(\mathbf{x}, \mathbf{u}_{t_0}, t; \boldsymbol{\theta}_f) + \mathbf{g}(\mathbf{x}, \mathbf{u}_{t_0}, t; \phi)\Big) dt + \boldsymbol{\sigma}_{\boldsymbol{x}}(\mathbf{x}(t)) d\mathbf{B}(t),$$

with variational parameters $\phi = \{\phi_x, \phi_z, \phi_\theta\}$. Instead of assuming an explicit parametric form for $q(\mathbf{x}(t)|\boldsymbol{\phi}_x)$, we instead assume the trajectories $\{\mathbf{x}(t)\}_t$ are described by samples from a posterior SDE with drift modified by including a small fully connected neural network $\mathbf{g}$ [236]. This approach allows an efficient and flexible form of the variational posterior in Equation 5.12. $\bar{\mathbf{W}}$ is the posterior mean of the network weights. The functional form of the posterior drift is both more general and more easily trainable than the network SDE in Equation 5.6, but ultimately is forced to be close to the network dynamics in Equation (5.6) by the loss function. The loss function for this approach has been previously derived [234, 235]. The imputed baseline states $\mathbf{x}_0$ are averaged over.

For the latent state $\mathbf{z}$, the approximate posterior takes the form

$$\boldsymbol{\mu}_z,\ \boldsymbol{\sigma}_z,\ \boldsymbol{\gamma}_z = \text{Encoder}(\tilde{\mathbf{y}}_{t_0}, \mathbf{o}_{t_0}, \mathbf{u}_{t_0}, t_0, \phi_z), \tag{5.13}$$
$$\tilde{\mathbf{y}}_{t_0} = \mathbf{o}_{t_0} \odot \mathbf{y}_{t_0} + (1 - \mathbf{o}_{t_0}) \odot \epsilon_{\mathbf{y}_{s,t_0},\text{pop}}, \tag{5.14}$$
$$q(\mathbf{z}|\mathbf{y}_{t_0}, \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, t_0, \boldsymbol{\phi}_z) \equiv q(\mathbf{z}^{(L)}|\tilde{\mathbf{y}}_{t_0}, \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, t_0, \boldsymbol{\phi}_z), \tag{5.15}$$
$$= \mathcal{N}(\mathbf{z}^{(0)}|\boldsymbol{\mu}_z, \boldsymbol{\sigma}_z^2) \prod_{l=1}^{L} \left|\det \frac{\partial a^{(l)}(\mathbf{z}^{(l)}, \boldsymbol{\gamma}_z, \phi_z)}{\partial \mathbf{z}^{(l)}}\right|^{-1}, \tag{5.16}$$

where the functions $a^{(l)}$ are RealNVP normalizing flows [237] used to transform the

Gaussian base distribution for $\mathbf{z}^{(0)}$ to the non-Gaussian posterior approximation, conditioned on the specific individual with $\boldsymbol{\gamma}_z$. These are invertible neural networks that transform probability distributions while maintaining normalization. In Equation 5.14 we fill in missing values in the observed health state since $\mathbf{o}$ is a mask of observed variables and $\epsilon_{\mathbf{y}_{s,t_0},\text{pop}}$ is sampled from a Gaussian distribution with the sex and age-dependent sample mean and standard deviation. $\odot$ is element-wise multiplication. These filled in values are temporarily input to the encoder neural network, and replaced after imputation.

The variational parameters $\boldsymbol{\phi}$ of the approximate posterior are optimized to minimize the KL divergence between the approximation and the true posterior. This minimized KL divergence provides a lower bound to the model evidence that can be maximized,

$$\begin{aligned}\log p(\{\mathbf{y}_{t_k}\}_k|\mathbf{u}_{t_0},\mathbf{o}_{t_0},t_0) \geq \mathbb{E}_{\boldsymbol{\theta},\,\mathbf{z},\,\mathbf{x}_0|\mathbf{z},\,\{\mathbf{x}(t)\}_t|\mathbf{x}_0}\Bigg[\\ \log p(\boldsymbol{\theta})p(\mathbf{z})p(\{\mathbf{x}(t)\}_t|\mathbf{x}_0,\mathbf{u}_{t_0},\boldsymbol{\theta})p(a,c|\{\mathbf{x}(t)\}_t,\mathbf{u}_{t_0},t_0)\prod_k p(\mathbf{y}_{t_k}|\mathbf{x}(t_k),\mathbf{o}_{t_k},\boldsymbol{\theta}) \\ -\log q(\mathbf{z}|\mathbf{y}_0,\mathbf{u}_{t_0},\mathbf{o}_{t_0},t_0)q(\boldsymbol{\theta})q(\{\mathbf{x}(t)\}_t|\mathbf{x}_0,\mathbf{u}_{t_0})\Bigg],\end{aligned} \tag{5.17}$$

where in the expectation $\boldsymbol{\theta}$, $\mathbf{z}$, and $\{\mathbf{x}(t)\}_t$ are sampled from their respective posterior distributions. The imputed baseline state is sampled as,

$$\boldsymbol{\mu}_{\mathbf{x}} = \text{Decoder}(\mathbf{z},\mathbf{u}_{t_0},t_0) \tag{5.18}$$

$$\tilde{\mathbf{x}}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}},\boldsymbol{\sigma}^2_{\mathbf{y}}) \tag{5.19}$$

$$\mathbf{x}_0 = \mathbf{o}_{t_0}\odot\mathbf{y}_{t_0} + (1-\mathbf{o}_{t_0})\odot\tilde{\mathbf{x}}_0. \tag{5.20}$$

Note that we keep the observed value $\mathbf{y}_{t_0}$ when available.

The final objective function to be maximized is $\mathcal{L}$, where the derivation is provided

in Appendix 5. We obtain

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\phi}) &= \mathbb{E}\Bigg[\sum_{k=0}^{K} \mathbf{o}_{t_k} \odot \log \mathcal{N}(\mathbf{y}_{t_k}|\mathbf{x}(t_k), \boldsymbol{\sigma}_{\mathbf{y}}) \\
&+ (1-c)\big[\log \lambda(a|\mathbf{x}(t), \mathbf{u}_{t_0}, t_0) + \log S(a|\mathbf{x}(t), \mathbf{u}_{t_0}, t_0)\big] \\
&+ \int_{t_0}^{a} c \log S(t|\mathbf{x}(t), \mathbf{u}_{t_0}, t_0)dt + \int_{a}^{a_{\max}} (1-c)\log\left(1 - S(t|\mathbf{x}(t), \mathbf{u}_{t_0}, t_0)\right)dt \\
&- \frac{1}{2}\int_{t_0}^{a} \Big|\Big|\boldsymbol{\sigma}_x^{-1} \odot \big(\mathbf{W}\mathbf{x} - \bar{\mathbf{W}}\mathbf{x} - \mathbf{g}(\mathbf{x}, \mathbf{u}_{t_0}, t)\big)\Big|\Big|_2^2 dt\Bigg] \\
&- KL(q(\boldsymbol{\theta})||p(\boldsymbol{\theta})) - KL(q(\mathbf{z}^{(0)}|\mathbf{y}_0, \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, t_0)||p(\mathbf{z}^{(0)})) \\
&+ \sum_{l=1}^{L} \log\left|\det \frac{\partial a^{(l)}(\mathbf{z}^{(l)}, \boldsymbol{\gamma}_z, \phi_z)}{\partial \mathbf{z}^{(l)}}\right|, \qquad (5.21)
\end{aligned}$$

as the loss function for each individual. This is for all individuals in the data multiplied by the sample weights $s^{(m)}$ for each individual m.We penalize the survival probability by integrating the probability of being dead from the death age a to $a_{\max}$, which better estimates survival probabilities [238]. We set $a_{\max} = 5$ years. Otherwise, it is difficult for the model to learn $S \to 0$ for large t. The last 3 lines are the KL-divergence terms for variational inference. The very last line is for the normalizing flow portion of the variational auto-encoder.

To simplify the evaluation of $\mathcal{L}$ and decrease the number of parameters, we assume independent Gamma posteriors for each measurement error parameter σ_y with separate shape α_i and rate β_i. We also assume independent Laplace posteriors for each of the network weights W_{ij} with separate means $\bar{W}_{ij}$ and scales b_{ij}. For the approximate distribution of all other parameters we use delta functions, and together with uniform priors this leads to simplifying the approach to just optimizing these parameters instead of optimizing variational parameters of the posterior.

5.5.5 Summarized training procedure

1. Pre-process data. Assign N dynamical health variables and B static health variables. Reserve validation and test data from training data.

2. Sample batch and apply masking corruption and temporarily fill in missing

values with samples from the population distribution,

$$\tilde{\mathbf{y}}_{t_0} = \mathbf{c} \odot \mathbf{o}_{t_0} \odot \mathbf{y}_{t_0} + (1 - \mathbf{c} \odot \mathbf{o}_{t_0}) \odot + \epsilon_{\mathbf{y}_{s,t_0},\text{pop}}, \tag{5.22}$$

$$\mathbf{c} \sim \text{Bernoulli}(0.9). \tag{5.23}$$

3. Impute initial state $\mathbf{x}_0$ with the VAE and compute the initial memory state of the mortality rate GRU,

$$\begin{aligned} \mathbf{z} &\sim q(\mathbf{z}|\tilde{\mathbf{y}}_{t_0}, \mathbf{u}_{t_0}, \mathbf{c} \odot \mathbf{o}_{t_0}, t_0), && (5.24)\\ \tilde{\mathbf{x}}_0 &\sim \mathcal{N}(\mathbf{x}_0|\boldsymbol{\mu}_{\mathbf{x}}(\mathbf{z}, \mathbf{u}_{t_0}, t_0), \boldsymbol{\sigma}^2_{\mathbf{y}}) && (5.25)\\ \mathbf{x}_0 &= \mathbf{o}_{t_0} \odot \mathbf{y}_{t_0} + (1 - \mathbf{o}_{t_0}) \odot \tilde{\mathbf{x}}_0, && (5.26)\\ \mathbf{h}_{t_0} &= \mathbf{H}(\mathbf{x}_0, \mathbf{u}_{t_0}, t_0). && (5.27) \end{aligned}$$

4. Sample trajectory from the SDE solver for the posterior SDE and compute mortality rate from GRU,

$$\begin{aligned} \{\mathbf{x}(t)\}_t &= \text{SDESolver}(\mathbf{x}_0, \mathbf{u}_{t_0}, t_0), && (5.28)\\ \{S(t)\}_t &= \text{GRU}(\{\mathbf{x}(\mathbf{t})\}_t|\mathbf{h}_{t_0}). && (5.29) \end{aligned}$$

5. Compute the gradient of the objective function (Equation B.1) and update parameters, returning to step 2 until training is complete.

6. Evaluate model performance on test data.

5.5.6 Network architecture and Hyperparameters

The different neural networks used are summarized in Appendix Table B.4. We use ELU activation functions for most hidden layer non-linearities, unless specified otherwise. We have $N = 29$ dynamical health variables, and $B = 19$ static health variables. Additionally, we append a mask to the static health variables indicating which are missing, of size 17 (sex and ethnicity are never missing).

The functions f_i in Equation (5.6) are feed-forward neural networks with input size $2 + B + 17$, hidden layer size 12, and output size 1. Each $f_i, i \in \{1, ..., N\}$ has its own weights. The noise function $\boldsymbol{\sigma}_{\mathbf{x}}$ has input size N, hidden layer size N, and output size N. The posterior drift $\mathbf{g}$ is a fully-connected feed-forward neural network

with input size $N + B + 1 + 17$, hidden layer size 8, and output size N. The VAE encoder has input size $2N + B + 1 + 17$, hidden layer sizes 95 and 70, and output size 40, with batch normalization applied before the activation functions for each hidden layer. The VAE decoder has input size $20 + B + 17$, hidden layer size 65, and output size N with batch normalization applied before the activation for the hidden layer. The size of the latent state $\mathbf{z}$ is 20. The mortality rate λ is a 2-layer GRU [112] with a hidden layer sizes of 25 and 10.

We use 3 normalizing flow networks to transform the latent distribution from the Gaussian $\mathbf{z}^{(0)}$ to $\mathbf{z}$. We use RealNVP normalizing flow networks [237] with layer sizes 30, 24, and 10 with batch normalization before a Tanh activation function for the hidden layer. The size of $\boldsymbol{\gamma}_z$ is 10.

We use batchsize of 1000 and learning rate 10^{-2} with the ADAM optimizer [111]. We decay the learning rate by a factor of 0.5 at loss plateaus lasting for 40 or more epochs. We use KL-annealing with β increasing linearly from 0 to 1 during the first 300 epochs for the KL loss terms for $q(\mathbf{x}(t))$ and $q(\mathbf{z}(t))$, and increase linearly from 0 to 1 from 300 to 500 epochs for the KL terms for the prior on $\mathbf{W}$. SDEs are solved with the strong order 1.0 stochastic Runge-Kutta method [239] with a constant time-step of 0.5 years. Integrals in the likelihood are computed with the trapezoid method using the same discretization as the dynamics.

5.5.7 Evaluation metrics

RMSE scores

Longitudinal health trajectory predictions are assessed with the Root-Mean-Square Error (RMSE) of the predictions with respect to the observed values. The RMSE is evaluated for each health variable and is weighted by the sample weights $s^{(m)}$. We compute these RMSE values for predictions for a specific age t_k,

$$\text{RMSE}_i(t_k) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\sum_{k:t_k \geq t} s^{(m)}(\psi_i^{-1}(x_i^{(m)}(t_k)) - y_{i,t_k}^{(m)})^2}, \tag{5.30}$$

where the inverse transform ψ_i^{-1} reverse any log-scaling and the z-scoring performed on the variables. The index (m) indicates the individual, for M total individuals.

Time-dependent C-index

The C-index measures the probability that the model correctly identifies which of a pair of individuals live longer. Our model contains complex time-dependent effects where survival curves can potentially intersect, so we use a time-dependent C-index [177],

$$\begin{aligned} C_{\text{td}} &= \Pr(\hat{S}^{(m_1)}(a^{(m_1)}) < \hat{S}^{(m_2)}(a^{(m_1)}) | a^{(m_1)} < a^{(m_2)}, c^{(m_1)} = 0) \\ &= \frac{\sum_{m_1,m_2} s^{(m_1)} s^{(m_2)} \delta[\hat{S}^{(m_1)}(a^{(m_1)}) < \hat{S}^{(m_2)}(a^{(m_1)})] \delta[a^{(m_1)} < a^{(m_2)}] \delta[c^{(m_1)} = 0]}{\sum_{m_1,m_2} s^{(m_1)} s^{(m_2)} \delta[a^{(m_1)} < a^{(m_2)}] \delta[c^{(m_1)} = 0]}, \end{aligned} \tag{5.31}$$

where $s^{(m)}$ are individual sample weights. We denote death ages by t_d and censoring ages by t_c, and define $a^{(m)} = \min(t_c^{(m)}, t_d^{(m)})$ as the last observed age for censored individuals ($c^{(m)} = 1$) or the death age for uncensored individuals ($c^{(m)} = 0$). The indexes (m_1) and (m_2) indicate the pair of individuals that are being compared. Delta functions $\delta[]$ have value 1 if the argument is true, otherwise have value 0.

Brier score

The Brier score compares predicted individual survival probabilities to the exact survival curves, i.e. a step function where $S = 1$ while the individual is alive, and $S = 0$ when the individual is dead. The censoring survival function $G(t)$ is computed from the Kaplan-Meier estimate of the censoring distribution (using censoring as events rather than the death [178]), which is used to weight the individuals to account for censoring. Then the Brier score is computed for all possible death ages,

$$\text{BS}(t) = \frac{1}{M} \sum_m s^{(m)} \left[\frac{\delta(a^{(m)} \le t, c^{(m)} = 0) \big(S^{(m)}(t)\big)^2}{G(a^{(m)})} + \frac{\delta(a^{(m)} > t) \big(1 - S^{(m)}(t)\big)}{G(t)} \right]. \tag{5.32}$$

D-calibration

For well-calibrated survival probability predictions, we expect $p\%$ of individuals to have survived past the pth quantile of the survival distribution. This can be evaluated using D-calibration, and we follow the previously developed procedure [214] for computing the D-calibration statistic. The result is a discrete distribution that should match a uniform distribution if the calibration is perfect.

We use a χ^2 test to compare to the uniform distribution. Using 10 bins, we use a χ^2 test with 9 degrees of freedom. Larger p-values (and smaller χ scores) indicate that the survival probabilities are more uniformly distributed, as desired.

2-sample classification tests

To assess the quality of our synthetic population, we train a logistic regression classifier and evaluate its ability to differentiate between the observed and synthetic populations [196, 205, 217, 218]. Ideally, a synthetic population would be indistinguishable from the observed population, giving a classification accuracy of 50%.

Our classifier takes the current age t, the synthetic or observed health variables $\mathbf{y}_t$, and the background health information variables $\mathbf{u}_{t_0}$, and then outputs the probability of being a synthetic individual or a real observed individual from the data-set. Missing values in the observed population are imputed with the sex and age-dependent sample mean, and these same values are applied to the synthetic health trajectories by masking the predicted values.

Hierarchical clustering

We perform hierarchical clustering on the network weights $\mathbf{W}$. This is done by constructing a dissimilarity matrix,

$$\boldsymbol{\omega} = (\mathbf{W}^T + \mathbf{W})/2, \tag{5.33}$$

$$\mathbf{D} = \max(\boldsymbol{\omega}) - \boldsymbol{\omega}, \tag{5.34}$$

and then using this dissimilarity matrix $\mathbf{D}$ to perform agglomerative clustering with the average linkage [240]. We use the Scikit-learn [241] package with the "AgglomerativeClustering" routine.

5.5.8 Comparison with linear models

Imputation for comparison models

For the linear survival and longitudinal models, we use MICE for imputation [212] with a random forest model [213]. We impute with the mean of the estimated values. We use 40 trees and do a hyperparameter search over the maximum tree depth. We use the Scikit-learn [241] package.

Proportional hazards survival model

To compare with a suitable baseline model for survival predictions, we use a proportional hazards model [119] with the Breslow baseline hazard estimator [120]:

$$\lambda(t|t_0, \mathbf{y}_{t_0}, \mathbf{u}_{t_0}) = \exp\left(\beta_0 t_0 + \boldsymbol{\beta}_y \cdot \mathbf{y}_{t_0} + \boldsymbol{\beta}_u \cdot \mathbf{u}_{t_0}\right), \tag{5.35}$$

$$S(t|t_0, \mathbf{y}_{t_0}, \mathbf{u}_{t_0}) = \exp\left(-\hat{\Lambda}_0^{\mathrm{Br}}(t)\lambda(t|t_0, \mathbf{y}_{t_0}, \mathbf{u}_{t_0})\right). \tag{5.36}$$

We include elastic net regularization [242] for the coefficients of the covariates.

Linear trajectory model

We use a simple linear model for health trajectories given baseline data,

$$y_{t_k,i} = y_{t_0,i} + \beta(\mathbf{y}_{t_0}, \mathbf{u}_{t_0}, t_0)(t_k - t_0), \tag{5.37}$$

$$\beta_i(\mathbf{y}_{t_0}, \mathbf{u}_{t_0}, t_0) = \beta_{0,i} t_0 + \boldsymbol{\beta}_{1,i} \cdot \mathbf{y}_{t_0} + \boldsymbol{\beta}_{2,i} \cdot \mathbf{u}_{t_0}, \tag{5.38}$$

trained independently for each variable i. The parameters $\beta_{0,i}$, $\boldsymbol{\beta}_{1,i}$, and $\boldsymbol{\beta}_{2,i}$ are trained with elastic net regularization.

Linear models' hyperparameters

We perform a random search over the L_1 and L_2 elastic net regularization parameters and the MICE random forest maximum depth using the validation set. The regularization term in the elastic net models is $\alpha l_{1,ratio}||\beta||_1 + \frac{1}{2}\alpha(1 - l_{1,ratio})||\beta||_2^2$, the common form of elastic net regularization used in Scikit-learn [241], the package we use to implement the elastic net linear model. We do the random search over $\log_{10}\alpha \in [-4, 0]$, $\log_{10} l_{1,ratio} \in [-2, 0]$, and maximum tree depth in $[5, 10]$ for 25 iterations.

We find the parameters $\alpha = 0.40423$, $l_{1,ratio} = 0.55942$, and a maximum tree depth of 10 for the longitudinal model hyperparameters. We find the parameters $\alpha = 0.00016$, $l_{1,ratio} = 0.15613$, and a maximum tree depth of 10 for the survival model hyperparameters.

5.5.9 Latent variable models

We compare our pair-wise interactions network model with alternate latent-varaible models, where we directly incorporate dynamics for the latent state $\mathbf{z}(t)$ and apply

the decoder to estimate the health variables $\mathbf{x}(t)$ at specific ages. With this approach we do not need to impute the baseline state of health variables, or to directly include dynamics for the observed health state. Rather an encoder maps the baseline health state $\mathbf{y}_{t_0}$ to the baseline latent state $\mathbf{z}_0$, dynamics are run on this latent space for $\mathbf{z}(t)$, and a decoder directly maps the latent states $\mathbf{z}(t)$ to the predicted output of the health variables $\mathbf{y}_t$. In this model, we also can choose the size of the latent state $\mathbf{z}$, and so we use this approach to explore how many dimensions are required for good predictions of health outcomes and survival.

These models have the form,

$$\mathbf{z}_0, \boldsymbol{\theta} \sim p(\mathbf{z}_0)p(\boldsymbol{\theta}) \qquad \text{(Prior)}$$

$$d\mathbf{z}(t) = \mathbf{f}(\mathbf{z}(t), \mathbf{u}_{t_0}, t; \boldsymbol{\theta}_{\boldsymbol{f}}) + \boldsymbol{\sigma}_{\mathbf{z}}(\mathbf{z}(t))d\mathbf{B}(t),\ \mathbf{z}(t_0) = \mathbf{z}_0, \qquad \text{(Dynamics)}$$

$$S(t) = \exp\big(-\int_{t_0}^{t} \lambda(\{\mathbf{z}(\tau)\}_{\tau \le t'}, \mathbf{u}_{t_0}, t'; \boldsymbol{\theta}_\lambda)dt'\big), \qquad \text{(Survival)}$$

$$\mathbf{y}_t \sim \mathcal{N}\Big(\boldsymbol{\psi}^{-1}(\boldsymbol{\mu}(\mathbf{z}(t), \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, \boldsymbol{\theta}_p)), \mathrm{diag}({\sigma_{\mathbf{y}}}^2)\Big), \qquad \text{(Health observation)}$$

$$a \sim \lambda(\{\mathbf{z}(\tau)\}_{\tau \le a}, \mathbf{u}_{t_0}, a; \boldsymbol{\theta}_\lambda)S(a), \qquad \text{(Survival observation)}$$

$$p(\{\mathbf{z}(t)\}_t, \boldsymbol{\theta} | \{\mathbf{y}_{t_k}\}_k, \mathbf{u}_{t_0}, t_0, a, c) \propto p(\boldsymbol{\theta})p(\mathbf{z}_0)p(\{\mathbf{z}(t)\}_t | \mathbf{z}_0, \mathbf{u}_{t_0}, t, \boldsymbol{\theta}) \times \qquad \text{(Inference)}$$
$$p(a, c | \{\mathbf{z}(t)\}_t, \mathbf{u}_{t_0}, t, \boldsymbol{\theta}) \prod_k p(\mathbf{y}_{t_k} | \{\mathbf{z}(t_k)\}_k, \boldsymbol{\theta}),$$

$$\boldsymbol{\theta} = \big\{\mathbf{W}, \sigma_{\mathbf{y}}, \sigma_{\mathbf{x}}, \boldsymbol{\theta}_\lambda, \boldsymbol{\theta}_p, \boldsymbol{\theta}_f\big\}, \qquad \text{(Parameters)}$$

where instead of the variable-wise neural networks in the pair-wise network model, the function $\mathbf{f}$ is now a full feed-forward neural network including the interactions between all variables. The function $\boldsymbol{\mu}$ is a decoder neural network which outputs the mean of a Gaussian distribution for the health variables $\mathbf{y}_t$, from the latent state at that age. Other than the size of the latent state $\mathbf{z}$, all other hyperparameters and the training procedure remain the same.

5.5.10 Code and Data availability

The English Longitudinal Study of Aging waves 0-8, 1998-2017 with identifier UKDA-SN-5050-17 is available at `https://www.elsa-project.ac.uk/accessing-elsa-data`. This requires registering with the UK Data Service.

Our code is available at `https://github.com/Spencerfar/djin-aging`.

Chapter 6

Network inference discussion

6.1 Summary of results from chapters 3, 4, and 5

Three different models of aging were used or developed in Chapters 3, 4, and 5. The main connecting theme of this work is the exploration or inference of the network structure in these models. These networks represent pairwise interactions between the health variables, and allow interpretation of the models. These models serve different purposes, and do not necessarily represent a strict replacement of the previous model.

Table 6.1 summarizes the aspects of these models. Details of this table are elaborated on in the sub-sections below.

6.1.1 Generic network model

Chapter 3 used the Generic Network Model of aging (GNM) [125,243]. This model was developed to describe the process of damage accumulation in aging, within a complex system described by a network of interacting components. This model involves only a few parameters that are manually tuned to agree with population-level mortality and Frailty Index scores. Nodes in this model do not represent any particular health variable, but are generic, representing abstract binary damageable components of health. We probed the network by observing the population-level behaviour with different pre-defined network structures, rather than directly inferring the network structure or predicting individual health trajectories.

We identified that a scale-free and disassortative structure best captures the behaviour seen in the observational data. This was the only structure to capture both the mortality and Frailty Index behaviour, as well as the hierarchical information structure within the nodes. Additionally, we offered an explanation for the difference between FI-clin and FI-lab. Based on the observed data for FI-clin and FI-lab [71, 126, 127], FI-clin deficits best correspond to high-degree nodes and FI-lab

Table 6.1: **Summary of network models.** This table highlights the features or uses of the models discussed in the preceding chapters. * binary variables are possible in the DJIN model, but they haven't been used in this thesis.

	Model		
	GNM	WNM	DJIN
Fitting with population average data	✓	✗	✗
Fitting with longitudinal data	✗	✗	✓
Fitting with cross-sectional data	✗	✓	✗
Binary variables	✓	✓	*
Continuous variables	✗	✗	✓
Fast training	✗	✗	✓
Infers robust network	✗	✗	✓
High dimensional	✓	✗	✓
Network inferred from data	✗	✓	✓
Conceptual/theoretical	✓	✗	✗
Predictive	✗	✓	✓
Generates synthetic populations	✓	✓	✓

deficits best correspond to low-degree nodes in a disassortative scale-free network. The temporal order of damage in the network results in the corresponding behaviour of these types of nodes.

This model enables the simulation of millions of aging individuals, with mortality rates and Frailty Index scores corresponding to observed population means. These simulated individuals can be used to theoretically study aging, without required detailed observational aging data. Members of the Rutenberg group have used this model as a basis for several distinct research directions. A simple theoretical model of disease has been developed by inducing increased damage rates in a particular region of the network, and then studying the resulting aging trajectories and mortality (work by Rebecca Tobin, unpublished). Similarly a model of interventions have been developed by enhancing repair of specific nodes for specific time-periods (work by Esha Sawant, unpublished). Work has also been done on understanding the optimal network structures for healthy aging and lifespan (work by Garrett Stubbings [244]).

While this model lets us explore the associations between different types of nodes in the network and their relation to categories of health variables in observed data, we do not have a direct mapping between a node in the network and a specific health variable, or even to a specific physiological system. This prevents the model

from being able to make predictions for individuals. To do this mapping and make predictions, the Weighted Network Model (WNM) was developed in Chapter 4 [48].

6.1.2 Weighted network model

The WNM used a much smaller network of 10 nodes, and used cross-sectional observational aging data to fit the network weights and damage rate parameters. This data allows us to map the model nodes to specific health variables, moving away from the generic nature of the nodes in the GNM. The model is fit with binary health variables, allowing the model to describe the evolution of the deficits in a Frailty Index with age. However, this model had to drop the simplicity of the GNM by introducing variable-weight connections, node-specific rate parameters, and replaced simple exponential functions with power-series relationships between the amount of damage to neighbouring nodes and damage rate.

The trained model is able to generate synthetic aging populations with realistic deficit accumulation and mortality. Given the state of current health deficits, the model was able to simulate the individuals health state until death. While this model is able to generate realistic synthetic aging populations, no one network was consistent with the data, resulting in a network structure that was not robust – fitting the model with a different random seed resulted in different networks. However, average pairwise rates were robust, suggesting that while the model behaviour is robust the actual network parameters were not. This is not uncommon in complex models, and is known as parameter sloppiness [245].

This model has a similar use as the GNM, but can simulate populations with specific health variables. However, this requires data with these health variables to fit model parameters.

While this model is well-suited to model the binary deficits of a Frailty Index, the model was very slow to train and slow to simulate populations. Attempting to train it with longitudinal data or a larger number of health deficits would be severely limited by computational power. Additionally, only being able to use binary health deficits limits the type of data that can be modelled. In Chapter 5 we developed the DJIN model of aging which addresses these concerns.

6.1.3 Dynamic Joint Interpretable Network model

The DJIN model makes substantial changes from the WNM. The model uses a much larger network of continuous-valued health variables, and is computationally much faster so that it can be trained on large longitudinal datasets. This model also makes use of demographic or background health information used as auxiliary variables to further improve predictions.

The DJIN model is able to make detailed longitudinal predictions for individuals. We show that the single comprehensive model achieves better performance than 30 regularized linear models separately trained for each health variable and survival. Survival predictions are calibrated to predict both relative and absolute risk.

This model also included an imputation method to impute the baseline state, so that predictions can be made even when the baseline health state is not fully observed. This imputation method was trained together with the model, which gives the model additional information about the quality of trajectory predictions to train the imputation component. This is in contrast to other imputation approaches such as MICE [212]. Additionally, this imputation method can be used to generate a fully synthetic population conditional on the background health variables and baseline age.

In the DJIN model we found robust network connections, in contrast to the WNM. Examining these network connections, we found that these connections show realistic physiological interactions. Hierarchical clustering on these connections shows clusters that make sense physiologically.

We compared the DJIN model similar latent-variable models. These latent-variable models considered non-linear dynamics on a latent health state, rather than the observed variables. We found that the DJIN model performs similarly to latent variable models with the latent state as a similar size to the number of observed health variables. This indicates that the linear pairwise network interactions are a good assumption, and that good predictions of health outcomes in aging require a high-dimensional model, although relative mortality risk predictions or the "progression of aging" can be reduced to a lower dimensional state.

Chapter 7

Discovering latent aging phenotypes in *C. elegans* with machine learning

7.1 Background

The DJIN model developed in Chapter 5 involved dynamics on the specific health variables from observed data, and inferred a network of interactions between these variables. The network of interactions for the observed health variables allows us to interpret the model by examining the causal interactions used by the model to make predictions. However, more flexible models can be built by considering dynamics on unobserved latent variables instead of the observed variables [219, 220, 246, 247].

In the machine learning literature, this style of latent variable model is known as an auto-encoder [110], and was discussed in the Background section of Chapter 2. In this model, an *encoder* embeds the observed state into a latent space, and a *decoder* decodes this latent state back into the observed state. Flexible dynamics can then be included on the latent space. To perform predictions, the observed early time health state is embedded into the latent space, then the latent dynamics are used to estimate the latent state at a later time, which is then be decoded into predictions of the observed state for these later times. This process takes the form,

$$\mathbf{z}_0 \sim q(\mathbf{z}_0|\mathbf{x}_0), \tag{Encoder}$$

$$d\mathbf{z}(t) = \mathbf{f}(\mathbf{z}(t)) + \boldsymbol{\sigma}(\mathbf{z}(t)) \cdot d\mathbf{B}(t),\ \mathbf{z}(t_0) = \mathbf{z}_0, \tag{Dynamics}$$

$$\mathbf{x}_t|\mathbf{z}(t) \sim p(\mathbf{x}_t|\mathbf{z}(t)). \tag{Decoder}$$

We also explored such latent variable alternatives to the DJIN model in Chapter 5, showing that similar performance can be achieved with a slightly lower number of dimensions, particularly for predictions of relative risk of death (with the C-index) – at the cost of the interaction network. Although high-dimensional latent variables are still required for predictions of health outcomes. While the goal of the DJIN model

was to infer this interaction network, these latent variable alternatives are likely a more attractive approach when the network is not desired.

While the latent variables can be difficult to interpret, these models can be made more interpretable by incorporating structure in the latent variables [248–251]. In previous latent variable approaches to modelling aging data, the structure of the model enforced latent variables to be multi-dimensional constant rates of aging [173] or one dimensional summary measures such as a dynamical frailty index [203] or a biological age [75–77, 215, 252–254]. The dynamics of explicitly constructed latent variables such as Physiological Dysregulation have also been considered [20, 255].

In this chapter, I demonstrate one way to interpret such a latent variable model of aging. This approach includes structure by allowing the inference of distinct clusters of latent variables and their aging trajectories. This lets the model identify distinct aging phenotypes, without providing guidance on what these phenotypes may be (i.e. the process is unsupervised). Additionally, we assume linear dynamics for the latent variables leading to simpler dynamics on this latent space.

In this Chapter, the model uses data from the model organism *C. elegans* (open source data from [256]). We use this data here instead of human data as in previous chapters for multiple reasons. First, the ELSA human data in Chapter 5 contained short time-series for individuals, and so we could not use many time-points in the encoder of our model, which limits the ability of the model to infer clusters for unseen test individuals. This *C. elegans* data with long time-series allows the development of a model that uses many time-points as input to the encoder. Second, there has been previous work on understanding distinct aging phenotypes in *C. elegans* [83,256,257]. Third, this demonstrates the ability of these computational aging models to apply to other organisms with their own distinct characteristics of aging.

The work of this Chapter is not yet ready for publication, but is a follow-up to Chapter 5.

7.2 Introduction

The hallmarks of aging [4] and the pillars of aging [5] identify key physiological processes that underlie aging. While these identified processes represent a low dimensional description of aging, the number of observed changes to healthy functioning

during aging are enormous [54]. Latent variable models can be used to identify a low-dimensional set of key underlying attributes in a complex system.

One example of a latent variable model in aging is biological age [17], where the underlying progression of aging is inferred from biological variables, however it is limited to one-dimension. Other approaches in aging explicitly construct the latent variables describing the progression of aging e.g. the FI [14] or physiological dysregulation [19]. There has also been work on machine learning based one-dimensional latent variable models [21]. Few approaches have looked into higher dimensional latent descriptions of aging [173].

Previous work with *C. elegans* has used dimensional reduction techniques to identify latent trajectories of aging [83, 257]. In worms with different genotypes, this technique has revealed divergent aging trajectories [257]. Additionally, there has been the discussion of accelerated aging phenotypes in *C. elegans* [256], although in this work distinct clusters were not identified.

In this work we build a machine learning model to identify distinct aging phenotypes by unsupervised learning. This model describes the dynamics of aging on a simple latent space, where distinct phenotypes are more clearly identifiable.

7.3 Results

7.3.1 Bayesian latent-variable model for *C. elegans*

We use *C. elegans* data from Zhang *et al.*, [256]. In this dataset worms are tracked until death ($\sim$ 350 hours), with physiological measurements from image data taken every 3 hours. The observed variables measure movement ability [258], auto-fluorescence of material build-up in the intestine [259], tissue integrity [260], body size [256], and reproductive capability [256]. All of these variables have been shown to be associated with lifespan, and form a set of health variables that can be automatically measured with high-throughput image analysis. Further details of the specific variables in this dataset are in the Methods section. We denote the observed time-series measures as $\{\mathbf{x}_t\}_{t=t_0}^{a}$, where $\mathbf{x}_t$ is a 11-dimensional vector of the health state at time t and a is the death age.

We build a Bayesian latent-variable model, illustrated in Figure 7.1. In this model,

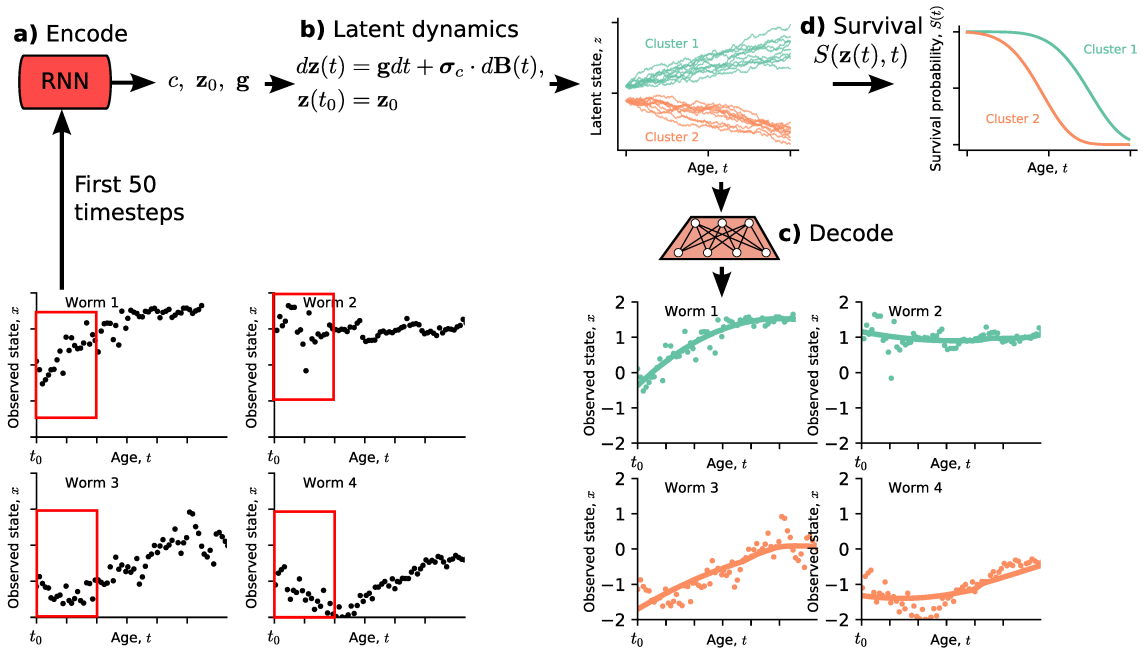


Figure 7.1: **A latent variable model for trajectory clustering.** **a)** The first 50 time-steps for each worm is input into an encoder recurrent neural network to infer the cluster label c, the baseline latent state $\mathbf{z}_0$, and the latent drift $\mathbf{g}$ with the approximate posterior $q(c, \mathbf{z}_0, \mathbf{g}|\{\mathbf{x}_t\}_t)$. **b)** Stochastic latent dynamics described by a stochastic differential equation with baseline state $\mathbf{z}_0$ and drift $\mathbf{g}$. **c)** Latent states are input into a decoder feed-forward neural network to output the predicted health state, $p(\mathbf{x}_t|\mathbf{z}(t))$. **d)** Survival probabilities are predicted from the latent states, $S(\mathbf{z}(t), t) = \exp\left(-\int_{t_0}^{t} h(\mathbf{z}(t'), t')dt'\right)$. The hazard rate $h(\mathbf{z}(t), t)$ is computed with a feed-forward neural network.

the dynamics of latent variables describe the underlying dynamics of aging, and are transformed to observed health variables and hazard rates for prediction. These latent variables are hierarchical, and dynamics are described by a continuous latent state $\mathbf{z}$ that depends on a discrete cluster label c. A simplified version of this model is,

$$
\begin{aligned}
c^{(m)}|\boldsymbol{\pi} &\sim \text{Categorical}(\boldsymbol{\pi}), && (7.1)\\
\mathbf{z}_0^{(m)}, \mathbf{g}^{(m)}|c^{(m)} &\sim p(\mathbf{z}_0^{(m)}|c^{(m)})p(\mathbf{g}^{(m)}|c^{(m)}) && (7.2)\\
d\mathbf{z}^{(m)}(t) &= \mathbf{g}^{(m)}dt + \boldsymbol{\sigma}_{\mathbf{z},c}d\mathbf{B}(t),\ \mathbf{z}^{(m)}(t_0) = \mathbf{z}_0^{(m)}, && (7.3)\\
S(\mathbf{z}^{(m)}(t), t)^{(m)} &= \exp\big(-\int_{t_0^{(m)}}^{t} h(\mathbf{z}^{(m)}(\tau), \tau)d\tau\big), && (7.4)\\
\mathbf{x}_t^{(m)}|\mathbf{z}^{(m)}(t) &\sim \mathcal{N}(\boldsymbol{\mu}(\mathbf{z}^{(m)}(t)), \boldsymbol{\sigma}(\mathbf{z}^{(m)}(t))). && (7.5)\\
&\quad \text{for } m = 1, ..., M \text{ individuals}
\end{aligned}
$$

The full detailed form of the model is shown in the Methods, which includes the full set of priors on the parameters. We perform inference with this model by variational Bayesian inference described in the Methods section. To summarize, for each individual m, we infer the posterior distribution of the individual specific cluster label, baseline latent state, and latent drift, $q(c^{(m)}, \mathbf{z}_0^{(m)}, \mathbf{g}^{(m)}|\{\mathbf{x}_t^{(m)}\}_t)$ with a recurrent neural network (RNN) that takes the first 50 time steps as input (Figure 7.1a). Then, latent trajectories $\mathbf{z}(t)$ are predicted using an SDE with baseline state $\mathbf{z}_0$ and constant drift $\mathbf{g}$ (Figure 7.1b, Equation 7.18). These latent states are decoded into predicted health states with a decoder neural network (Figure 7.1c, Equation 7.5) and predicted survival probabilities (Figure 7.1c, Equation 7.4). In Appendix C, we show that this model can accurately cluster data from a simulated dataset where we know the ground truth cluster labels.

This model requires choosing the dimension of the latent state $\mathbf{z}$. In Figure 7.2, model prediction accuracy on the test set is shown vs latent dimension for a) root-mean squared error (RMSE) for health predictions, b) C-index for survival predictions, and c) Integrated Brier Score for survival predictions. Since health predictions plateau at a dimension of 7 and survival predictions are similar for all dimensions, we have a latent state of size 7. Additionally in e), we show that the model infers 2 clusters on average regardless of the dimension, and that the distribution of clusters is distinct from the prior.

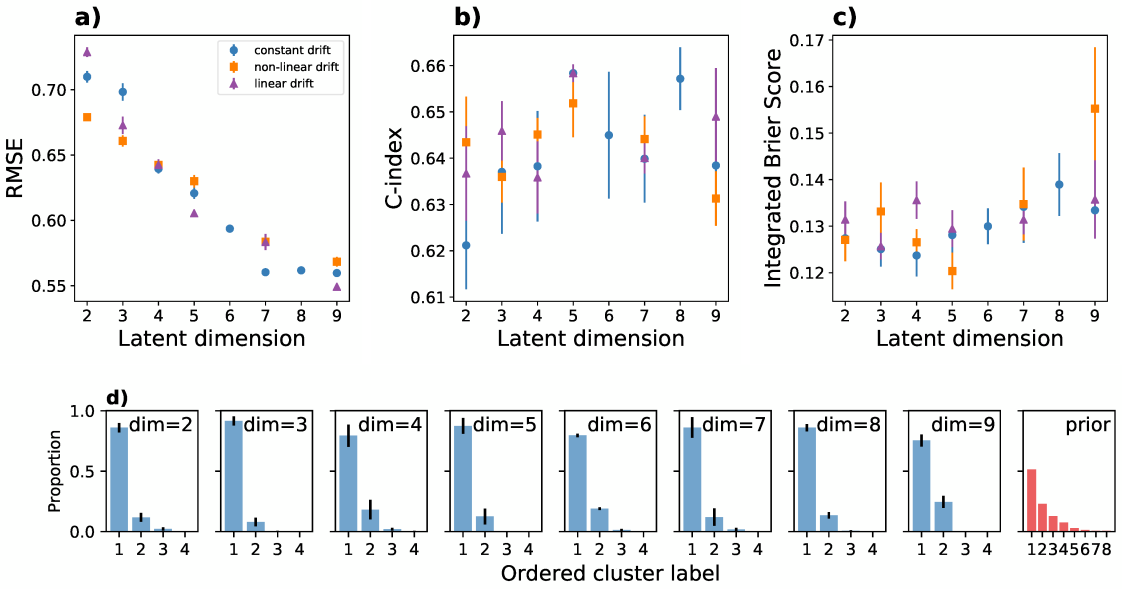


Figure 7.2: **Choosing model latent dimension.** In each plot, error bars are standard errors of the mean over 5 fits of the model. Blue circles show the constant drift model. Purple triangles show the linear drift model. Orange squares show the non-linear drift model. **a)** Root mean squared error (RMSE) vs latent dimension size. At a latent state of size 7, the constant drift model accuracy plateaus. **b)** Survival C-index vs latent dimension size. The C-index is noisy, but generally increases with latent dimension size. **c)** Survival Integrated Brier Score vs latent dimension size. Brier scores are better with increasing dimension until size 4 or 5, then begin to get worse for larger dimensions. **d)** Average proportion of worms in each cluster. Since cluster labels change with each fit, cluster labels correspond to the ordered size of the clusters. The model infers two clusters on average consistently, regardless of dimension. The red plot shows a histogram of samples from the Dirichlet process prior for the cluster probabilities $\boldsymbol{\pi}$.

We also show a comparison of the constant individual-specific drift model (blue) with a linear drift model (purple) and a non-linear drift model (orange), which use an individual-specific constant drift plus a cluster-specific weight matrix linearly coupling the latent variables (linear drift) or a cluster-specific feed-forward neural network for the drift as a function of the latent state (non-linear drift). While substantial hyper-parameter tuning of the neural network drift would be required for a full evaluation, we show here that the constant drift is a good simplification in comparison to these two approaches.

7.3.2 Distinct aging trajectories

In Figure 7.3a), we show our 7-dimensional latent space model for *C. elegans*. Each plot shows a different observed variable vs age, with red and blue colors indicating the two inferred clusters. Primarily, the blue cluster represents larger worms that lay more eggs, and the red cluster represents smaller worms that lay fewer eggs. Additionally we can see differences in the auto-fluorescence, intensity texture, and movement speed, where the smaller red cluster worms have lower intestinal auto-fluorescence, move slower and appear brighter.

These clusters also exhibit distinct short-lived and long-lived aging phenotypes. Figure 7.3b) shows the difference in lifespan for these worms. The red cluster worms cluster worms live longer, and a log-rank test (a standard statistical test to compare survival distributions) [261] is shown to highlight the difference in survival between these clusters.

A t-SNE dimensional reduction [262] is shown in Figure 7.3c), clearly showing these clusters on this lower dimensional space (note, this two-dimensional t-SNE reduction is not the same latent space as the 7-dimensional states inferred by the model). The average age-trajectory for these two groups through this space is shown with the bold lines. The color of the points corresponds to relative age t/a (age divided by death age), with lighter colors indicating early ages, and darker colors later ages. These trajectories are distinct, occupying separate regions of the space for the entire lifespan of the worms. This means that even in early life, these clusters have distinct health states, and continue to have distinct health states throughout their life.

These clusters are consistent with current the understanding of *C. elegans* aging biology. The finding of small worms with poor reproductive output being long-lived has been observed in other studies [256]. Similarly, there is a clear relationship between intestinal auto-fluorescence and lifespan [256, 259], with low auto-fluorescence worms living longer as seen here, as low auto-fluorescence worms are quickly clearing material from their intestines. The clustering is most clear in for these three variables. The relationship between movement speed and lifespan is more complex, and in the data we use very low speeds are associated with short lifespans, but the survival curves for medium to fast speeds overlap [256]. We have found that worms in

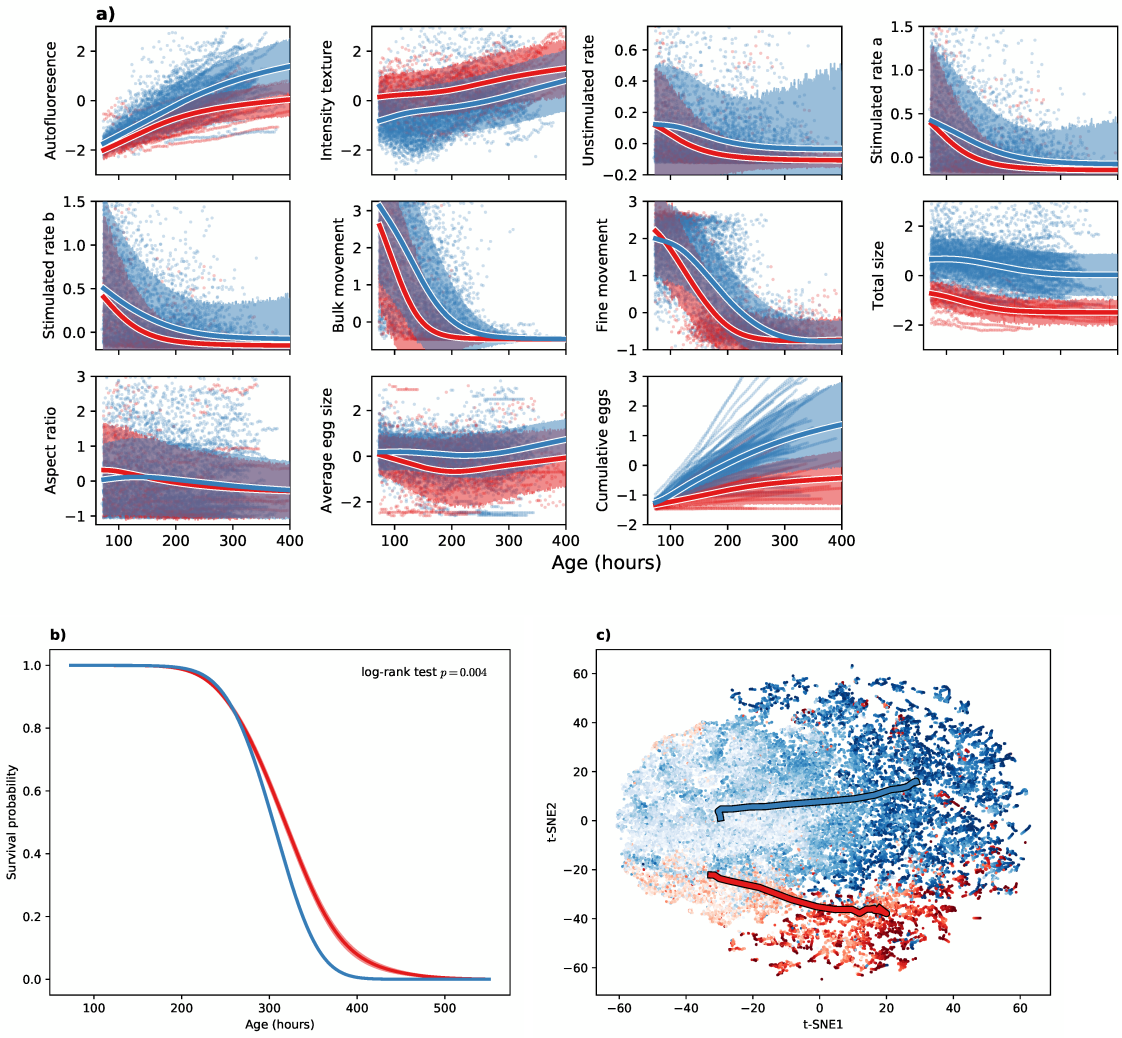


Figure 7.3: **Identifying distinct aging phenotypes.** **a)** Observed health variables vs age. Points show raw worm data, lines show the mean of the predicted trajectories from the model with the colored region showing 1 standard deviation from the mean. Colors show the two inferred clusters. All variables are in unit-less z-scores. **b)** Survival curves for the two clusters. Lines show the mean of the predicted survival probabilities from the model, with the shaded regions showing the standard error of the mean. A log-rank test is used to compare the lifespans for the two groups. **c)** t-SNE dimensionality reduction. Red/blue colors show the two clusters, with brightness accord to the relative age, which is age divided by death age t/a. Darker colors indicate higher relative age. The bold lines show the average of these points for binned ages.

the longer lived cluster are slightly slower. The cessation of vigorous worm movement may be a better signal for lifespan than simply movement speed [42], but this requires a cut-off to determine when this occurs.

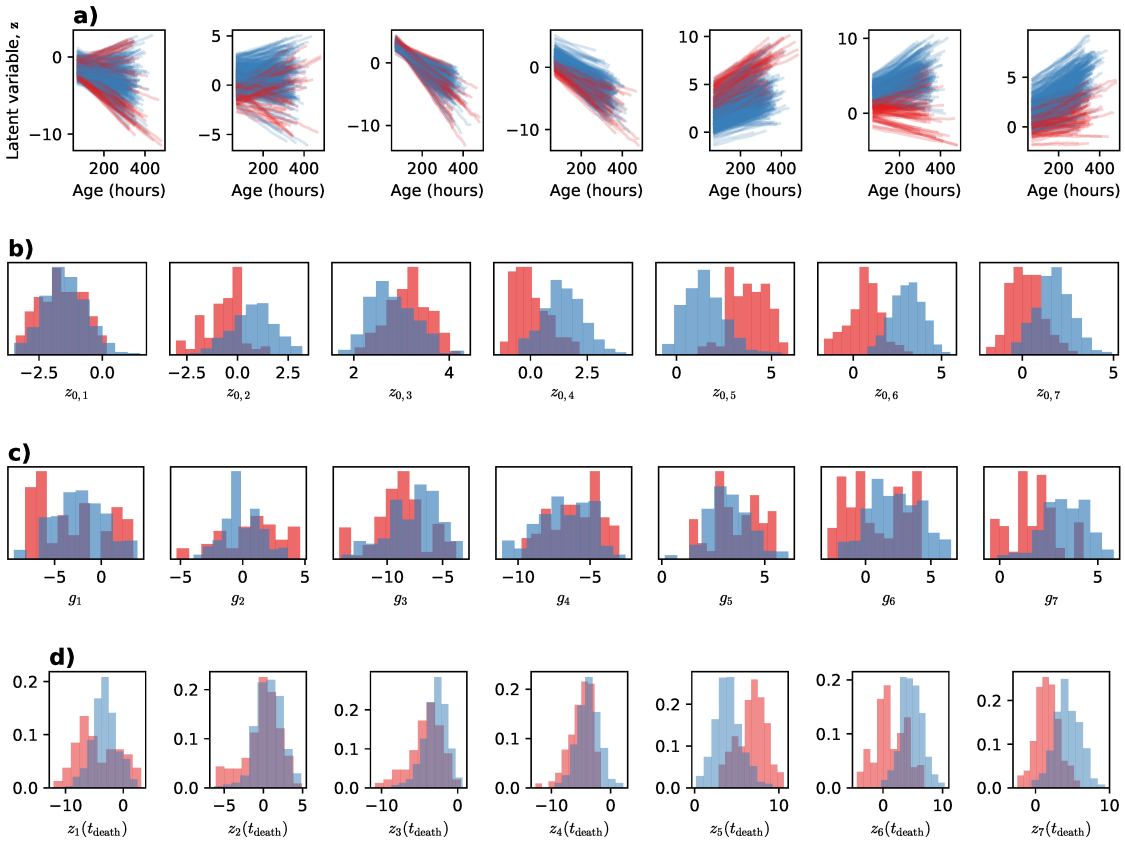


Figure 7.4: **Simple latent dynamics describe *C. elegans* aging. a)** Trajectories cleanly cluster on the latent space. Each plot shows a component of the latent state $\mathbf{z}$ vs age. For each worm in the test set, 5 sampled trajectories are shown. **b)** Distributions of the latent baseline state for the two clusters, $\mathbf{z}_0$. **c)** Distributions of the latent baseline drift for the two clusters, $\mathbf{g}$. **d)** Distributions of the latent state at death, $\mathbf{z}(t_{\text{death}})$. Latent states for the two clusters reach different end-points at death.

7.3.3 Aging dynamics are described by a simple low-dimensional dynamics with identifiable clusters

The complex dynamics of the observed variables are described by simple low-dimensional latent dynamics in this model. Given an individuals position and velocity in latent space, aging proceeds by simply moving in this direction. The observed health state and mortality rate are just complex transformations of this underlying latent state.

The dynamics of this 7-dimensional latent state are shown in Figure 7.4a). On this latent space, clusters are clearly identifiable. The distributions of the baseline state and linear drift are shown in Figures 7.4b) and c), showing that these clusters follow distinct aging trajectories. Additionally, in Figure 7.4d) we show that mortality in this model is determined by different points in latent space at death, because the only contribution to mortality is the latent state, rather than cluster-specific mortality rates. This puts all of the mortality information into the latent space.

7.4 Discussion

Flexible latent variable models can be hard to interpret. In this work, we developed an approach to interpret these latent variables by allowing the baseline latent states and dynamics to cluster. Doing so, the model can infer distinct divergent aging trajectory phenotypes. These clusters are inferred in an unsupervised way, and incorporate both the dynamics of the observed health state and survival.

Our approach is a probabilistic model based on a warped Gaussian mixture [263]. The advantage of this approach over simpler non-probabilistic clustering techniques such as k-means, agglomerative clustering, spectral clustering, or other traditional approaches [264] is that the criteria for clustering is setup by our generative model for the data (Figure 7.1), rather than constructing a distance or similarity metric. The data involves multivariate noisy irregularly-sampled time series of different lengths for health variables as well as observations of lifespan, and so it is not clear how to construct a good distance or similarity metric for which to cluster with. However, with our approach we need only to specify a flexible generative model for the dynamics and survival, and then clusters are naturally chosen by fitting the model. The use of SDEs

naturally allow irregularly-sampled time series of different lengths, and mortality is included by building a joint model of both health and survival. This approach also includes a metric for the quality of the clustering and choice of hyperparameters such as the size of the latent dimension – the quality of the model predictions on a test set.

Our model for the dynamics of aging is related to that of Pierson *et al.* [37], where multi-dimensional constant rates of aging are inferred. They model deterministic latent dynamics $\mathbf{z}(t) = \mathbf{r}t$, where $\mathbf{r}$ is a set of constant rates of aging inferred for each individual, while we model aging dynamics by constant drift SDEs for each individual. Our approach is distinct in that it includes clustering of the individual aging trajectories, stochasticity in the evolution of the aging trajectories, and includes mortality. Additionally, this approach only used a single baseline time-point to infer the rates of aging, similar to the DJIN model (Chapter 5), where here we include time-series input with a recurrent neural network.

Other approaches to analyzing aging data have shown that a low-dimensional latent state, such as biological age and similar approaches, can describe the progression of aging [18, 77, 203, 216, 265], although we have shown this is not sufficient to predict individual heterogeneous aging outcomes (Chapter 5). Here, we chose the dimensionality of the latent state by the quality of the model predictions.

It has been previously seen that genetic variants of *C. elegans* follow distinct trajectories in a 2-dimensional space reduced by t-SNE [257]. However, to do this they did not perform the dimensionality reduction on the full set of variables, but extracted a subset of the variables that showed the most significant variation between the two groups. Instead of this ad-hoc approach, our model directly clusters from the observed data.

In future work, our approach can be applied to such genetic variants of organisms or to groups having undergone interventions. This can be used to answer questions such as: are genetic mutants or aging interventions merely changing the rate at which aging occurs, or inducing a healthier divergent aging trajectory? Can we identify and understand why individuals within an intervention group that are outliers do not respond well to the intervention? We see a promising future for approaches such as this to answer these questions.

7.5 Methods

7.5.1 *C. elegans* data

We use *C. elegans* data from [256]. We use data from 672 isolated worms on bacterial food pads that are monitored with automated image acquisition and processing. Data for individuals are collected at $\sim$ 3 hour intervals (although the start time for each individual can vary by $\pm$ 3 hours). The data taken consists of locomotary ability (Bulk movement [pixel displacement over 3 hours], unstimulated movement rate [pixels/second], stimulated movement rate immediately after blue-light stimulus (a) [pixels/second], stimulated movement rate 1.5 seconds after blue-light stimulus (b) [pixels/second]), tissue integrity (average brightness of worm pixels), homeostatic ability to clear material from the intestine (80th percentile of worm pixel auto-fluorescence), body size and shape (total cross-sectional body size [pixels] and aspect ratio [area of rectangle bounding worm]), and reproductive output (Average egg size [pixels], cumulative eggs laid [count]). Worms are tracked until death. We drop the first 25 time-points for each worm, corresponding to the point in time when the average number of eggs laid by the worms goes above zero, i.e. we only use data for adult worms.

We split the data into a training set of 472 worms, a validation set of 100 worms and a test set of 100 worms. Training data is used to train the model, validation data is used to monitor training and decrease the learning rate during training, and test data is used to test model predictions.

7.5.2 Latent cluster trajectory model

We build a Bayesian latent cluster trajectory model. In this model, the dynamics of worm aging are described by an underlying latent health state evolving stochastically according to stochastic differential equations. For health predictions, this latent state is mapped to the observed variables with a decoder neural network, and for survival predictions the latent state is mapped to the hazard rate with another neural network.

This model is written,

$$\begin{aligned}
\boldsymbol{\pi} &\sim \text{GEM}(1), && \text{(Cluster probabilities)}\\
c|\boldsymbol{\pi} &\sim \text{Categorical}(\boldsymbol{\pi}), && \text{(Individual cluster label)}\\
\{\boldsymbol{\mu}_c\}_c &\sim \mathcal{N}(0, 25), && \text{(Cluster baseline latent state)}\\
\mathbf{z}_0|c &\sim \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_{\mu,c}), && \text{(Individual baseline latent state)}\\
\{\mathbf{f}_c\}_c &\sim \mathcal{N}(0, 25), && \text{(Cluster latent drift)}\\
\mathbf{g}|c &\sim \mathcal{N}(\mathbf{f}_c, \boldsymbol{\Sigma}_{f,c}), && \text{(Individual latent drift)}\\
d\mathbf{z}(t) &= \mathbf{g}dt + \boldsymbol{\sigma}_{\mathbf{z},c}d\mathbf{B}(t),\ \mathbf{z}(t_0) = \mathbf{z}_0, && \text{(Latent dynamics)}\\
\mathbf{x}_t|\mathbf{z}(t), t &\sim \mathcal{N}(\boldsymbol{\mu}(\mathbf{z}(t)), \boldsymbol{\sigma}^2_{\mathbf{x}}(\mathbf{z}(t)), && \text{(Health observations)}\\
S(t, \mathbf{z}(t)) &= \exp\left(-\int h(t', \mathbf{z}(t'))dt'\right), && \text{(Survival)}
\end{aligned}$$

For the cluster probabilities $\boldsymbol{\pi}$ we use a GEM stick-breaking prior [266] truncated to 8 clusters, this is a method of constructing a Dirichlet process which is typically used for clustering with an unknown number of clusters. We use broad normal priors for the cluster means and drifts, $\{\boldsymbol{\mu}_c\}_c$ and $\{\mathbf{f}_c\}_c$. Individual baseline latent states $\mathbf{z}_0$ and drifts $\mathbf{g}$ are assumed to follow a Gaussian mixture with the corresponding cluster means for the latent state $\{\boldsymbol{\mu}_c\}_c$ and drift $\{\mathbf{f}_c\}_c$. Latent dynamics are modelled with an SDE with drift $\mathbf{g}$ and initial latent state $\mathbf{z}_0$. Since the drift, baseline latent state and diffusive noise strength $\boldsymbol{\sigma}_{\mathbf{z},c}$ are all cluster dependent, these dynamics are clustered. A graphical representation of this model is shown in Figure 7.5a). All other parameters not discussed have uniform priors.

The latent state is mapped to the observed health state with a Gaussian observation model with mean $\boldsymbol{\mu}(\mathbf{z}(t)$ and standard deviation $\boldsymbol{\sigma}(\mathbf{z}(t))$, where these functions are neural networks. We impose that the mean $\boldsymbol{\mu}(\mathbf{z}(t)$ be non-negative, preventing the prediction of negative movement speeds. Additionally, we impose that the standard deviation $\boldsymbol{\sigma}(\mathbf{z}(t))$ of the movement speed variables goes to zero as the mean of the movement speed variables go to zero,

$$\sigma_i(\mathbf{z}(t)) = \max\left(\tilde{\sigma}_i(\mathbf{z}(t)) \times \mu_i(\mathbf{z}(t)),\ 0\right), \tag{7.6}$$

where $\tilde{\sigma}_i(\mathbf{z}(t))$ is a neural network.

The stick-breaking process is an iterative process defining the cluster probabilities

$\boldsymbol{\pi}$ as,

$$\begin{aligned}
\mathbf{v} &\sim \text{Beta}(1, \alpha = 1), && (7.7)\\
\pi_1 &= v_1, && (7.8)\\
\pi_i &= v_i \prod_{k=1}^{i-1}(1 - v_k),\ i = 1, ..., 7, && (7.9)\\
\pi_8 &= \prod_{k=1}^{8}(1 - v_k). && (7.10)
\end{aligned}$$

This is used to define the posterior in terms of $\mathbf{v}$ instead of $\boldsymbol{\pi}$.

7.5.3 Inference

We perform variational Bayesian inference with a posterior approximation

$$\begin{aligned}
q(\mathbf{z}(t), \mathbf{z}_0, \mathbf{g}, c, \boldsymbol{\mu}_c, \mathbf{f}_c, \mathbf{v}|\{\mathbf{x}_t\}_t) = q(\mathbf{z}(t)|\mathbf{z}_0, \mathbf{g})q(\mathbf{z}_0|\{\mathbf{x}_t\}_t)q(\mathbf{g}|\{\mathbf{x}_t\}_t) \quad (7.11)\\
\times q(c|\{\mathbf{x}_t\}_t)q(\boldsymbol{\mu}_c)q(\mathbf{f}_c)q(\mathbf{v}).
\end{aligned}$$

For simplicity, we leave out the parameters of these distributions $\boldsymbol{\phi}$ from the notation.

The specific distributions are defined as,

$$\begin{aligned}
q(\mathbf{z}_0|\{\mathbf{x}_t\}_t) &= \mathcal{N}(\boldsymbol{\mu}_z(\{\mathbf{x}_t\}_t), \boldsymbol{\Sigma}_z(\{\mathbf{x}_t\}_t)), && (7.12)\\
q(\mathbf{g}|\{\mathbf{x}_t\}_t) &= \mathcal{N}(\boldsymbol{\mu}_g(\{\mathbf{x}_t\}_t), \boldsymbol{\Sigma}_g(\{\mathbf{x}_t\}_t)), && (7.13)\\
q(c|\{\mathbf{x}_t\}_t) &= \text{Categorical}(\boldsymbol{\pi}(\{\mathbf{x}_t\}_t)), && (7.14)\\
q(\boldsymbol{\mu}_c) &= \mathcal{N}(\mathbf{m}_{\mu,c}, \mathbf{s}^2_{\mu,c}), && (7.15)\\
q(\mathbf{f}_c) &= \mathcal{N}(\mathbf{m}_{f,c}, \mathbf{s}^2_{f,c}), && (7.16)\\
q(\mathbf{v}) &= \text{Beta}(\boldsymbol{\alpha}, \boldsymbol{\beta}). && (7.17)
\end{aligned}$$

Functions of $\{\mathbf{x}_t\}_t$ use a recurrent neural network for on the first 50 time points. Values $\mathbf{m}_{\mu,c}, \mathbf{m}_{f,c}, \mathbf{s}_{\mu,c}, \mathbf{s}_{f,c}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ are the variational parameters of the corresponding distributions. Figure 7.5(left) shows the construction of the encoder network for $q(\mathbf{z}_0, \mathbf{g}, c, |\{\mathbf{x}_t\}_t)$. The structure of this encoder is called a Variational Ladder Auto-Encoder (VLAE) [267]. This structure places more abstract features deeper in the hierarchy, and prevents the model from ignoring these deeper levels of the hierarchy.

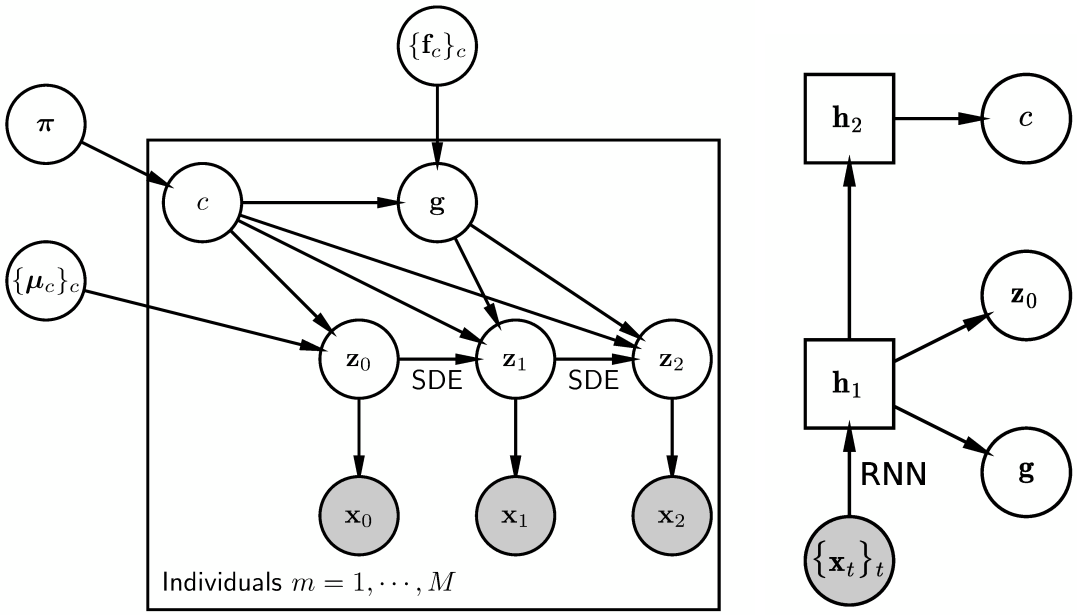


Figure 7.5: **Graphical representation of the clustering model**. **(left)** Structure of the generative model for clustering. This network depicts the factorization of the joint distribution $p(\{\mathbf{x}\}_t, \{\mathbf{z}(t)\}_t, \mathbf{z}_0, \mathbf{g}, c, \{\boldsymbol{\mu}_c\}_c, \{\mathbf{f}_c\}_c, \boldsymbol{\pi})$ for the model. Variables inside the rectangle are all indexed by m, representing each individual from the data. In this model there are global cluster probabilities $\boldsymbol{\pi}$, global cluster baseline latent state means $\{\boldsymbol{\mu}_c\}_c$, and global cluster latent drift means $\{\mathbf{f}_c\}_c$. An individual specific cluster label c is inferred, which leads to individual specific baseline $\mathbf{z}_0$ and drift $\mathbf{g}$. Using these individual specific variables, latent dynamics are run with a stochastic differential equation for $\mathbf{z}(t)$. These latent states are then decoded into observed states $\mathbf{x}_t$. **(right)** Structure of the encoder network for $q(\mathbf{z}_0, \mathbf{g}, c, |\{\mathbf{x}_t\}_t)$. This structure is known as a Variational Ladder Auto-encoder [267]. All arrows represent feed-forward neural network layers, except for the arrow specifically labeled as a recurrent neural network (RNN). The grey circle for $\{\mathbf{x}_t\}_t$ represents the observed time series of health measurements. The squares $\mathbf{h}_1$ and $\mathbf{h}_2$ are deterministic intermediate hidden states. $\mathbf{z}_0$, $\mathbf{g}$, and c are random variables from $q(\mathbf{z}_0, \mathbf{g}, c, |\{\mathbf{x}_t\}_t)$. The variational parameters $\boldsymbol{\phi}$ are not shown, but connect to every white node. This network represents the factorization $q(\mathbf{z}_0, \mathbf{g}, c|\{\mathbf{x}_t\}_t) = q(\mathbf{z}_0|\{\mathbf{x}_t\}_t)q(\mathbf{g}_0|\{\mathbf{x}_t\}_t)q(c|\{\mathbf{x}_t\}_t)$.

Parameters not shown here have delta function posterior distributions, as discussed in Chapter 2. Note that $\boldsymbol{\Sigma}$ indicate covariance matrices, while $\mathbf{s}^2$ indicate diagonal covariance matrices only parameterized by the variance of each component.

To sample the latent trajectories from the posterior $q(\mathbf{z}(t)|\mathbf{z}_0, \mathbf{g})$, we follow previous work [234–236] and solve the auxillary SDE,

$$d\mathbf{z}(t) = \bar{\mathbf{g}}dt + \mathbf{g}_1(\mathbf{z}(t), t)dt + \boldsymbol{\sigma}_{\mathbf{z},c} d\mathbf{B}(t),\ \mathbf{z}(t_0) = \mathbf{z}_0, \tag{7.18}$$

where the drift consists of the posterior mean of the individual drift $\bar{\mathbf{g}} = \boldsymbol{\mu}_{g,c}(\{\mathbf{x}_t\}_t)$ plus an additional feed-forward neural network $\mathbf{g}_1(\mathbf{z}(t), t)$.

The loss (negative ELBO) for this model is,

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\phi}) \quad = \quad & -\sum_c q(c|\{\mathbf{x}_t\}_t)\mathbb{E}_{\mathbf{z}_0,\mathbf{g},\boldsymbol{\mu}_c,\mathbf{f}_c,\mathbf{z}(t)\sim q}\Big[\log \prod_k p(\mathbf{x}_{t_k}|\mathbf{z}(t_k)) + \\
& + \int_{t_0}^{a} \log S(t, \mathbf{z}(t))dt + \log h(a, \mathbf{z}(a)) + \log S(a, \mathbf{z}(a)) \\
& + \int_{a}^{a+a_{\max}} \log\left(1 - S(t, \mathbf{z}(t))\right)dt \\
& + \log \frac{p(\mathbf{z_0}|c, \boldsymbol{\mu}_c)p(\mathbf{g}|c, \mathbf{f}_c))p(\boldsymbol{\mu}|c)p(\mathbf{f}|c)p(c|\boldsymbol{\pi})p(\boldsymbol{\pi})}{q(\mathbf{z}_0|\{\mathbf{x}_t\}_t)q(\mathbf{g}|\{\mathbf{x}_t\}_t)q(c|\{\mathbf{x}_t\}_t)q(\boldsymbol{\mu}_c)q(\mathbf{f}_c)q(\mathbf{v})} \\
& - \frac{1}{2}\int ||(\bar{\mathbf{g}} + \mathbf{g}_1(\mathbf{z}(t), t) - \mathbf{g})/\boldsymbol{\sigma}_{\mathbf{z},c}||^2 dt\Big].
\end{aligned} \tag{7.19}$$

Note that since c is a discrete variable, we cannot compute gradients with samples of $c \sim q$. Thus, we need to sum over all cluster labels, weighted by the probability $q(c|\{\mathbf{x}_t\}_t)$. The derivation of this loss is similar to the derivation of the loss for the DJIN model in Chapter 5, which is shown in Appendix B. The first 3 lines of this loss are the likelihood for the data, including both health and survival. We penalize the survival probability by integrating the probability of being dead from the death age a to $a_{\max}$, which better estimates survival probabilities [238] (the integral on the third line). We set $a_{\max} = 45$ hours. Otherwise, it is difficult for the model to learn $S \to 0$ for large t. The last 2 lines are the KL-divergence terms for variational inference.

To compute the term involving the cluster labels and the stick-breaking prior $\log p(c|\boldsymbol{\pi})$, we follow Blei *et al.*, [268]

$$\begin{aligned}
\mathbb{E}[\log p(c|\boldsymbol{\pi})] \quad = \quad & \sum_{k=1}^{8}\Big[q(c > k|\{\mathbf{x}_t\}_t)(\Psi(\alpha_k) - \Psi(\alpha_k + \beta_k)) \\
& + q(c = k|\{\mathbf{x}_t\}_t)(\Psi(\beta_k) - \Psi(\alpha_k + \beta_k))\Big],
\end{aligned} \tag{7.20}$$

where Ψ is the digamma function, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the variational parameters of the posterior distribution of $\mathbf{v}$ in Equation 7.17.

7.5.4 Summary of training procedure

1. Sample batch of data. Run RNN backwards through the first 50 time-points to output cluster labels, individual baseline state states, and drifts

$$\{\mathbf{x}_t\}_{t=-50}^{0} \rightarrow \text{RNN} \rightarrow c,\ \mathbf{z}_0,\ \mathbf{g} \sim q(c, \mathbf{z}_0, \mathbf{g}|\{\mathbf{x}_t\}_{t=-50}^{0}). \tag{7.21}$$

2. Set baseline state $\mathbf{z}(t_0) = \mathbf{z}_0$ and solve latent SDE dynamics with the SDE in Equation 7.18,

$$d\mathbf{z}(t) = \bar{\mathbf{g}}dt + \mathbf{g}_1(\mathbf{z}(t), t)dt + \boldsymbol{\sigma}_{\mathbf{z},c}d\mathbf{B}(t). \tag{7.22}$$

3. Compute mean $\boldsymbol{\mu}(\mathbf{z}(t))$ and standard deviation $\boldsymbol{\sigma}(\mathbf{z}(t))$ of observed health state from latent state.

4. Compute hazard rates $h(t, \mathbf{z}(t))$. Compute survival functions from hazard rates,

$$S(t, \mathbf{z}(t)) = \exp\left(-\int h(t', \mathbf{z}(t'))dt'\right). \tag{7.23}$$

5. Compute gradient of loss in Equation 7.19 and update variational parameters. Return to step 1.

7.5.5 Hyperparameters and network architectures

We use a latent state of size 7. The encoder uses a GRU RNN [112], with an internal hidden state size of 30 followed by a Batch-Norm ($\mathbf{h}_1$ in Figure 7.5). The hidden state $\mathbf{h}_2$ has size 15 and follows from an ELU activition, linear layer, and Batch-Norm applied to $\mathbf{h}_1$.

Linear transformations of $\mathbf{h}_1$ are used to compute the mean and covariance of $\mathbf{z}_0$ and the mean and covariance of $\mathbf{g}$. A linear transformation of $\mathbf{h}_2$ and a soft-max activation is used to compute probabilities of each cluster label c.

The hazard rate is estimated from the latent state $\mathbf{z}$ and age t with a feed-word neural network that has input size 8 (latent state plus age), hidden layer size 7, an ELU activation, and output layer size 1. The mean for the observed health variables $\boldsymbol{\mu}(\mathbf{z}(t))$ has an input layer of size 5, a hidden layer of size 8, an ELU activation, and an output layer of size 11.

Covariance matrices are parameterized with a low-rank approximation of rank 3 (compared to the full rank size of 7). This is done by fitting a matrix $\mathbf{C} \in \mathbb{R}^{3\times 7}$ and a vector $\mathbf{d} \in \mathbb{R}^7$ and computing the covariance $\mathbf{\Sigma} = \mathbf{d}^2 + \mathbf{C}^T\mathbf{C}$.

We use the ADAM optimizer to update the parameters [111], with a learning rate of 10^{-2}. A learning rate schedule is used to halve the learning rate when loss plateaus last for longer than 10 epochs. We use KL annealing and increase the weight of the KL-divergence terms in the loss function from 0 to 1 linearly over the first 100 epochs. All models are run to 300 epochs — all training set and validation set ELBOs are plateaued at this point.

7.5.6 Non-constant drift models

To validate our simplification of the dynamics with a constant drift, we compare the constant drift dynamics with linear and non-linear drift dynamics. The form of these two approaches are,

$$d\mathbf{z}(t) = \mathbf{g}dt + \mathbf{W}_{\text{lin},c}\mathbf{z}(t)dt + \boldsymbol{\sigma}_{\mathbf{z},c}d\mathbf{B}(t), \qquad \text{(linear drift)}$$

$$d\mathbf{z}(t) = \mathbf{g}dt + \mathbf{g}_{\text{non-lin},c}(\mathbf{z}(t))dt + \boldsymbol{\sigma}_{\mathbf{z},c}d\mathbf{B}(t), \qquad \text{(non-linear drift)}$$

where $\mathbf{W}_{\text{lin},c}$ is a matrix linear coupling the latent variables, and $\mathbf{g}_{\text{non-lin},c}$ is a feed-forward neural network with one hidden layer of size 5.

Note that while the constant drift $\mathbf{g}$ is inferred for each individual (i.e. the drift is individual-specific), $\mathbf{W}_{\text{lin},c}$ and $\mathbf{g}_{\text{non-lin},c}$ are only cluster-specific. Including individual-specific versions of these would greatly increase the parameters of the model.

While a full test of the the non-linear approach would require substantial hyper-parameter tuning, we show that the constant drift dynamics work well in comparison to these two models, representing an enormous simplification of the dynamics.

7.5.7 Evaluation metrics

RMSE scores

Health predictions are assessed with the Root-Mean-Square Error (RMSE) of the predictions with respect to the observed values. We compute these RMSE values for

predictions for a variable i and a specific age t_k,

$$\mathrm{RMSE}_i(t_k) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{x}_i^{(m)}(t_k) - x_{i,t_k}^{(m)})^2}. \tag{7.24}$$

The observed variables are denoted $x_{i,t_k}^{(m)}$, and the predicted variables $\hat{x}_i^{(m)}(t_k)$.

Time-dependent C-index

The C-index measures the probability that the model correctly identifies which of a pair of individuals live longer. Our model contains complex time-dependent effects where survival curves can potentially intersect, so we use a time-dependent C-index [177],

$$\begin{aligned} C_{\mathrm{td}} &= \Pr(S^{(m_1)}(a^{(m_1)}) < S^{(m_2)}(a^{(m_1)}) | a^{(m_1)} < a^{(m_2)}) \\ &= \frac{\sum_{m_1,m_2}\delta[S^{(m_1)}(a^{(m_1)}) < S^{(m_2)}(a^{(m_1)})]\delta[a^{(m_1)} < a^{(m_2)}]}{\sum_{m_1,m_2}\delta[a^{(m_1)} < a^{(m_2)}]}, \end{aligned} \tag{7.25}$$

We denote death ages by a. Note, there is no censoring for the worm data, in contrast to the data in other chapters. The indexes (m_1) and (m_2) indicate the pair of individuals that are being compared. Delta functions $\delta[]$ have value 1 if the argument is true, otherwise have value 0.

Brier score

The Brier score compares predicted individual survival probabilities to the exact survival curves, i.e. a step function where $S = 1$ while the individual is alive, and $S = 0$ when the individual is dead. Then the Brier score is computed for all possible death ages,

$$\mathrm{BS}(t) = \frac{1}{M}\sum_{m}\Big[\delta[a^{(m)} \le t]\big(S^{(m)}(t)\big)^2 + \delta[a^{(m)} > t]\big(1 - S^{(m)}(t)\big)\Big]. \tag{7.26}$$

Note there is no censoring, in contrast to the data in other chapters.

Chapter 8

Resilience and robustness decrease with age, are dynamic over broad timescales, and can be attenuated with interventions in aging mice and humans

8.1 Background

Chapter 3 (generic network model) and Chapter 4 (weighted network model) both considered discrete transitions between health states consisting of binary deficits, with damage and repair rates describing these transitions. In the weighted network model, we fit rate parameters from cross-sectional data and in the generic network model the rate parameters were tuned so that simulated populations agree with the population mortality rate and average Frailty Index scores vs age. In both of these cases the primary focus was not on the actual repair and damage processes, but on the mortality and FI scores, and the parameters of the rates were inferred indirectly from the observed data as part of the overall fitting of the model.

However, with longitudinal data for binary deficits we can directly extract damage and repair rates from the data to study the processes of damage and repair in aging. Additionally since the damage rate is the rate of acquiring new damage and the repair rate is the rate of recovering damage, we can interpret these damage and repair processes as (inverse) robustness and resilience. These are important concepts in aging research, though there has been little attempt to directly observe these processes, despite the popularity of theories of aging and review articles based on these processes [8, 90, 93, 94].

Here repair and damage are extracted directly from the data, with no induced stressor or intervention targeted at the specific damage, representing a measurement of natural robustness and resilience from a stochastic damage and repair. We use 3 different mouse datasets and 1 human dataset. In two of these mouse datasets we have interventions – mice treated with the Angiotensin-Converting Enzyme (ACE)

inhibitor enalapril, and voluntary exercise. Both of these interventions have shown to slow the increase of the Frailty Index in mice [87, 88]. In humans we stratify by net household wealth, representing a measure of socioeconomic status, which is associated with lower degrees of frailty [269]. The net household wealth serves as a proxy for medical or behavioral interventions that are not observed in the data, and do not exist in the mice data. These distinct datasets strengthen the conclusions by replicating our findings.

This work is the first time this type of study has been done, opening up a new method of approaching resilience and robustness. This chapter presents the paper "Resilience and robustness decrease with age, are dynamic over broad timescales, and can be attenuated with interventions in aging mice and humans", in preparation for submission.

8.2 Introduction

As organisms age, they can be described by a health state that evolves according to dynamical processes of damage and repair. The health state is the net result of accumulated damage and repair, and studies of aging have mostly focused on the health state rather than the underlying dynamic processes, due to the difficulty of their measurement. Two common approaches to measuring health-states, the Frailty Index (FI) [14] and the Frailty Phenotype [16], are constructed from cross-sectional data and so cannot assess damage and repair processes directly. Nevertheless, strong associations between frailty measures and adverse health outcomes [3, 270] indicate that frailty has a strong effect on these underlying dynamical processes. This is supported by the increasing net accumulation of health deficits with worsening health [162, 271].

Reduced resilience, or the decreasing ability to recover from stressors, is increasingly seen as a key manifestation of organismal aging [8, 90, 91, 93, 94, 272]. Resilience is often assessed by the ability to recover from an acute stressor, such as a heat/cold shock, viral infection, or anesthesia, within a short experimentally-accessible timeframe [93, 95, 96, 273]. Robustness, or an organism's resistance to damage, has not been as well studied – but there is also evidence for its decline with age [20,274]. Both resilience and robustness sustain organismal health during aging, but their relative

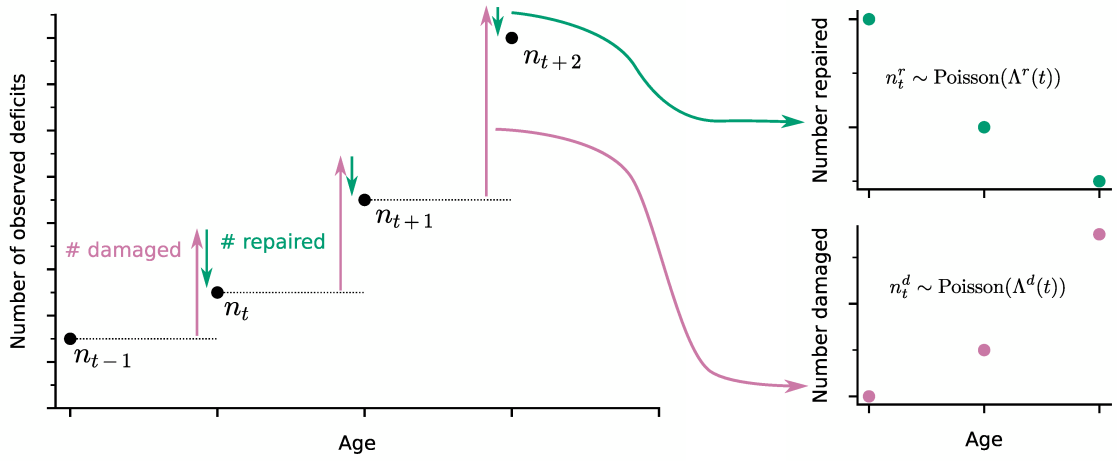
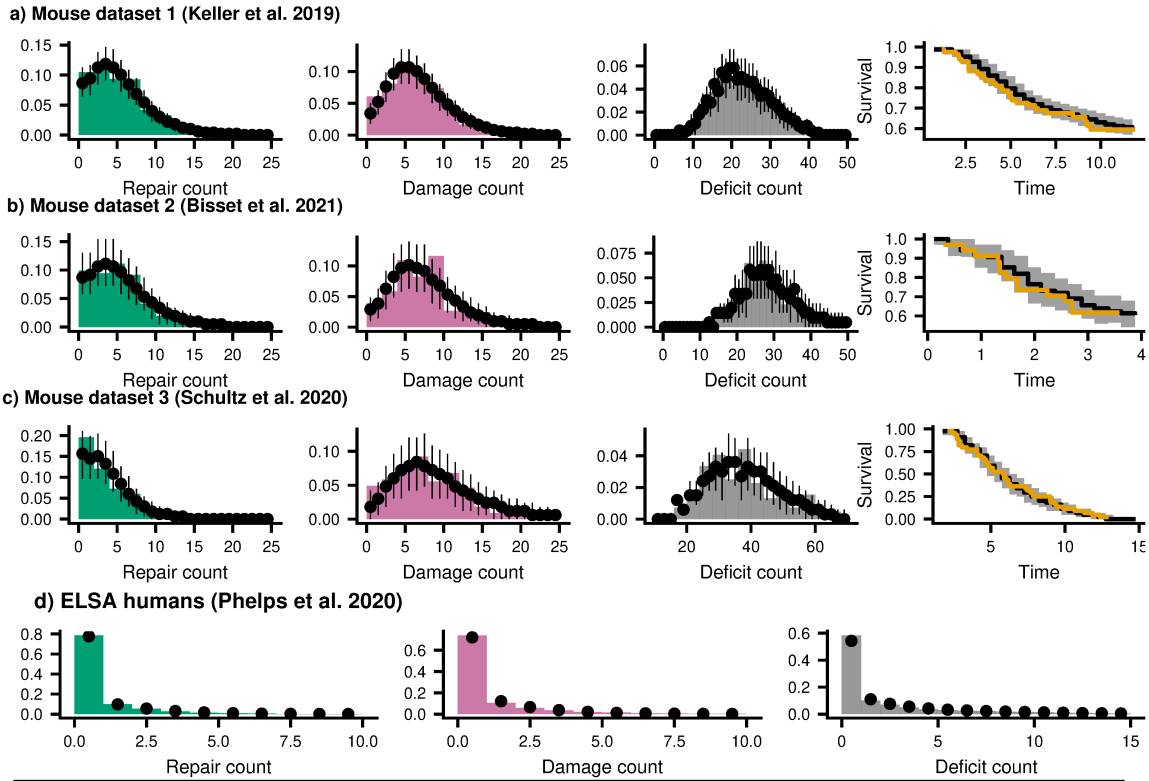


Figure 8.1: **Extracting damage and repair from the longitudinal observation of binary health deficits.** Instead of just considering the Frailty Index (FI) with the proportion of deficits n_t/N, we consider the number of deficits damaged n_t^d (pink) or repaired n_t^r (green) between time intervals of observed changes in the FI. Time-dependent damage and repair rates are extracted with Poisson models.

importance and their timescales of action remain unexplored.

To study resilience and robustness in aging, we have developed a novel method of analysis that uses longitudinal binarized health-deficit data from mice and humans to obtain summary measures of organismal damage and repair processes over time. This approach can be adapted to use any other biomarkers, rather than being restricted to biomarkers specifically associated with resilience [94]. We apply our method to study how resilience and robustness evolve with age and how they differ between species, between sexes, and under different health interventions.

Geroscience is aimed at identifying and developing interventions to extend both lifespan and healthspan [5, 275, 276]. While some interventions that affect aging health have been identified, how they differentially affect damage and repair, and their timescales of action, is less understood. We consider interventions in mice that have previously been shown to have a positive impact on frailty, the ACE inhibitor enalapril [87] and voluntary exercise [88]. In humans, we stratify individuals by net household wealth within the English Longitudinal Study of Aging [201]. Wealth is a socioeconomic factor associated with aging health [269]. Understanding how various interventions affect aging health by affecting resilience and robustness will better enable us to improve and combine interventions to fulfil the geroscience agenda.



Dataset	Outcome	Median R^2 (95% CrI)	Proportion within 95% CrI
Mouse group 1 (Keller *et al.* 2019)	Repair counts	$0.42(0.48, 0.57)$	0.96
	Damage counts	$0.46(0.42, 0.50)$	0.91
	Deficit counts	$0.48(0.45, 0.50)$	0.96
Mouse group 2 (Bisset *et al.* 2021)	Repair counts	$0.51(0.45, 0.56)$	0.96
	Damage counts	$0.54(0.48, 0.60)$	0.94
	Deficit counts	$0.50(0.46, 0.53)$	0.97
Mouse group 3 (Schultz *et al.* 2019)	Repair counts	$0.49(0.41, 0.58)$	0.96
	Damage counts	$0.46(0.39, 0.52)$	0.95
	Deficit counts	$0.48(0.45, 0.51)$	0.94
ELSA humans	Repair counts	$0.52(0.50, 0.53)$	0.98
	Damage counts	$0.44(0.43, 0.45)$	0.96
	Deficit counts	$0.47(0.46, 0.47)$	0.95

Figure 8.2: **Joint models fit quality.** Quality of model fits for repair counts (green), damage counts (pink), total deficit counts (grey), and survival probability (orange) for **a)** Mouse group 1 (Keller *et al.*, 2019) [87], **b)** Mouse group 2 (Bisset *et al.*, 2021) [88], **c)** Mouse group 3 (Schultz *et al.*, 2020) [253], and **d)** ELSA humans [277]. These distributions show the distributions of these counts and survival for all individuals, all time-points, and all sex and treatment groups. The table shows R^2 statistics and the proportion of posterior 95% credible intervals (CrI) for data-points where the residual includes zero (expected to be 0.95). Full plots of these residuals are shown in Appendix D.

8.3 Results

8.3.1 Measuring resilience and robustness

A well-established approach to quantify health in both humans and in animal models is to count binarized deficits in a Frailty Index (FI) [14, 73]. In longitudinal studies, the health state of individuals can be assessed at each follow-up. Shown in the Figure 8.1 schematic, the change in of number of deficits from one follow-up to the next is determined by the number of new deficits (e.g. damage, deficit value transitioning from 0 to 1, pink arrow) minus the number of repaired deficits that were previously damaged (deficit value transitioning from 1 to 0, green arrow). These counts of damaged and repaired deficits between follow-ups represent summary measures of the underlying damage and repair processes. We use a Poisson model with age-dependent rates to model the age-dependent damage and repair rates using these longitudinal counts. This is a joint model coupling the damage and repair rates together with mortality.

In Figure 8.2, the quality of the model fits are shown with posterior predictive distributions of the repair counts, damage counts, total deficit counts, and survival probability, as well as summary statistics of an analysis of residuals. This shows that the model fits the data well, and posterior credible intervals (CrI) accurately reflect the uncertainty in the data, shown by the correct proportion of residuals containing zero within 95% CrI. While we can compute binned-averages of repair and damage rates from the raw data, this model is used to compute additional information, such as the posterior distribution of correlation coefficients to determine the increase/decrease of the rates with age, slopes of the rates vs age, and hazard rates of morality. Since our model is Bayesian, we can compute posterior credible intervals of these quantities, representing the probability that the true value falls within the interval.

In our approach, damage rates are the probability of acquiring a new deficit per unit of time, and repair rates are the probability of repairing a deficit per unit time. These are measures of susceptibility to damage (lack of robustness), and ability to repair (resilience). Since the FI is a whole organism-level summary measure of health, these measures are also whole organism-level measures of robustness and resilience.

8.3.2 Natural resilience and robustness in control aging populations

We first establish the trends of repair and damage rates in aging. In Figure 8.3, we plot the age-dependence of the repair and damage processes in mice and humans along with the Frailty Index for mouse datasets from a) from Keller *et al.*, 2019 [87], b) Bisset *et al.*, 2021 [88], and c) Schultz *et al.*, 2020 [253], and humans from the ELSA study [277]. Humans are separately plotted by decade of baseline age at entry to the study, to separate out recruitment effects. Points are binned averages from the data, and lines are posterior samples from the model of the rates. Both a decrease in repair rates and an increase in damage rates result in an increasing Frailty Index with age.

In both humans and mice, as individuals grow older, there is a strong decrease in repair rates and increase in damage rates (except in mouse group 2 for damage rates). These effects are replicated in three separate mouse groups, and an observational human study for older individuals. Spearman rank correlations ρ for each plot are also shown, highlighting the increase or decrease of the rates with age. The Bayesian model of the rates is used to compute 95% posterior credible intervals of these correlations (in brackets).

The observed damage occurs due to natural stochastic transitions, rather than an experimentally induced or observed stressor [90, 95]. The observed resilience occurs after no interventions, representing natural resilience to the stochastic damage. This natural resilience can be thought of as resilience to the natural stressors of life, which continually occur during aging. While errors in deficit assessment could contribute to the damage or repair assessment, we would expect these errors to be constant with age. In contrast, we observe decreasing repair rates and increasing damage rates with age. Therefore these age-dependent rates signify decreasing resilience and robustness with age, in both mice and humans.

One caveat with this approach is that we may miss fast damage and repair dynamics that occur on time-scales shorter than the observed time-points. For example, we cannot observe daily or weekly fluctuations in deficit states. Therefore, our measurements of damage and repair are only looking at the net damage and net repair between time-points. Ultimately, our approach results in *summary* measures of damage and repair.

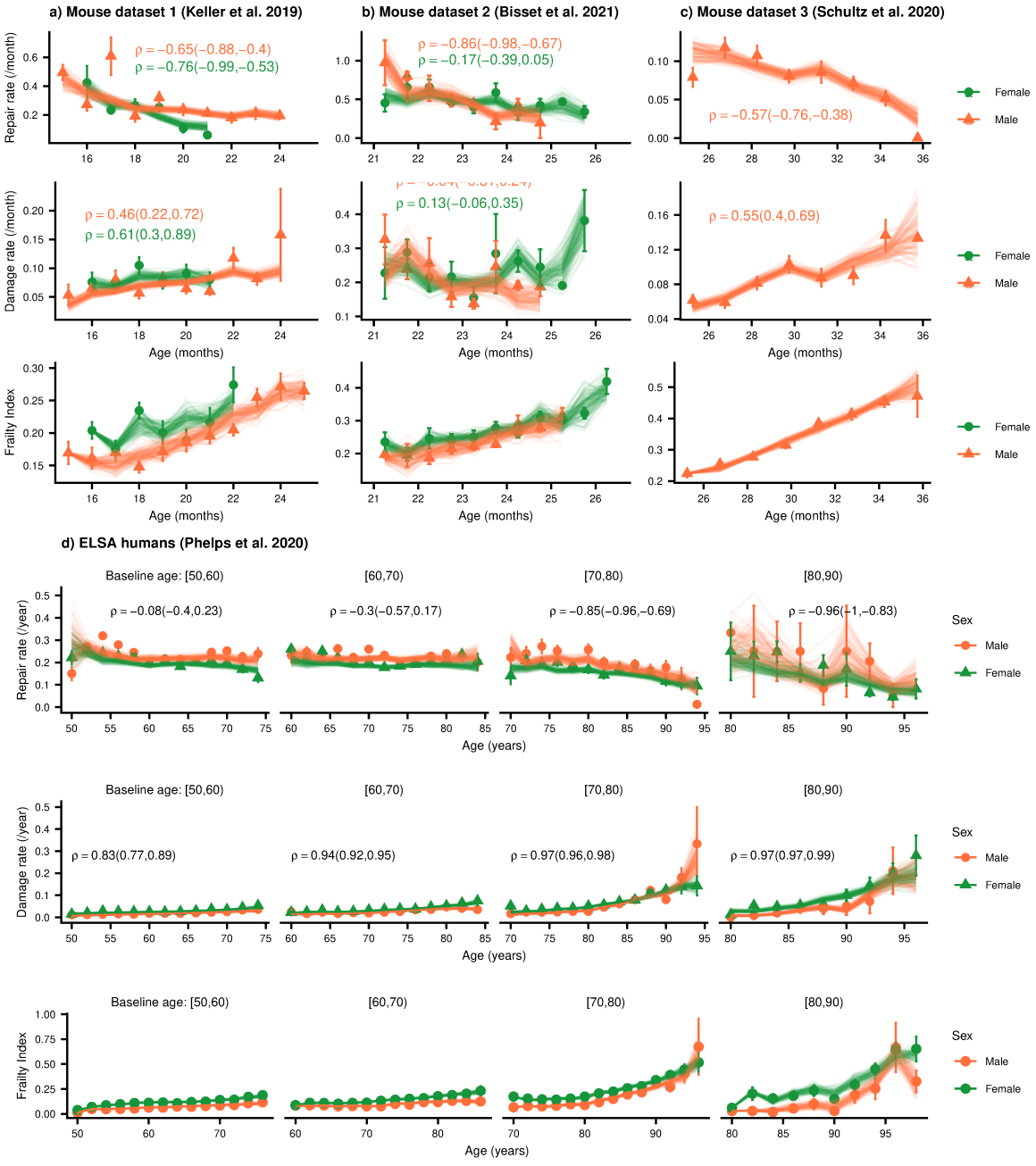


Figure 8.3: **Separating Repair and damage from the Frailty Index.** Repair rates vs age (top), damage rates vs age (middle), and Frailty Index vs age (bottom) for **a)** Mouse dataset 1 (Keller *et al.*, 2019) [87], **b)** Mouse dataset 2 (Bisset *et al.*, 2021) [88], **c)** Mouse dataset 3 (Schultz *et al.*, 2020) [253] and **d)** ELSA humans [277]. For each plot the mean Spearman's rank correlation ρ between the rate and age is shown by the median of the posterior and a 95% posterior credible interval in parenthesis. Humans are separately plotted by decades of baseline age. Points in all plots represent binned averages of rates from the data with standard errors, and lines represent posterior samples from a Bayesian model of the rates (see methods).

In Appendix D Figure D.1, we highlight the sex differences in the mouse cohorts. We observe higher Frailty Index in female mice [57] and humans [278], with corresponding differences in repair and damage rates.

8.3.3 Interventions modify damage and repair rates in mice

Mouse datasets 1 (Keller *et al.*, 2019) and 2 (Bisset *et al.*, 2021) additionally have corresponding intervention cohorts treated with either the ACE inhibitor enalapril, or voluntary exercise. In Figure 8.4a and b, we show that these interventions target both repair and damage processes, resulting in lower FI damage accumulation for the treated groups. This shows that these interventions are not just reducing the susceptibility to damage (increasing robustness) or increasing resilience, but are impacting both processes in tandem.

The effect of these interventions is clearer seen when looking at how the rates change with time, shown with the age-slopes of the rates in Figure 8.4c and d. These slopes are computed with our Bayesian model for the rates, and the 95% posterior credible interval for the slopes are shown. These slopes show the rate of increase or decrease of the repair and damage rates as age increases. The interventions are strongly acting on the rate of decrease of repair rates and rate of decrease of damage rates with time, resulting in less cumulative damage over time.

We see similar effects between male and female mice treated with enalapril, the rate of decrease of repair rates and the rate of increase of damage rates are both attenuated, resulting in rates closer to zero. We see a sex-specific effect on mice treated with voluntary exercise, such that repair and damage rates appear to be effected differently by exercise in each sex. There is a complete stoppage or reversal of the decrease in repair rate for female mice, and a reduction of the increase in damage rate with age. For male mice, there is a reduction in the rate of decrease of repair rates (but not a stoppage or reversal), and a complete stoppage of the increase in damage rates. This suggests that, in mice, exercise acts more to increase repair in females, and to decrease damage in males. Appendix Figure D.2 highlights these sex differences.

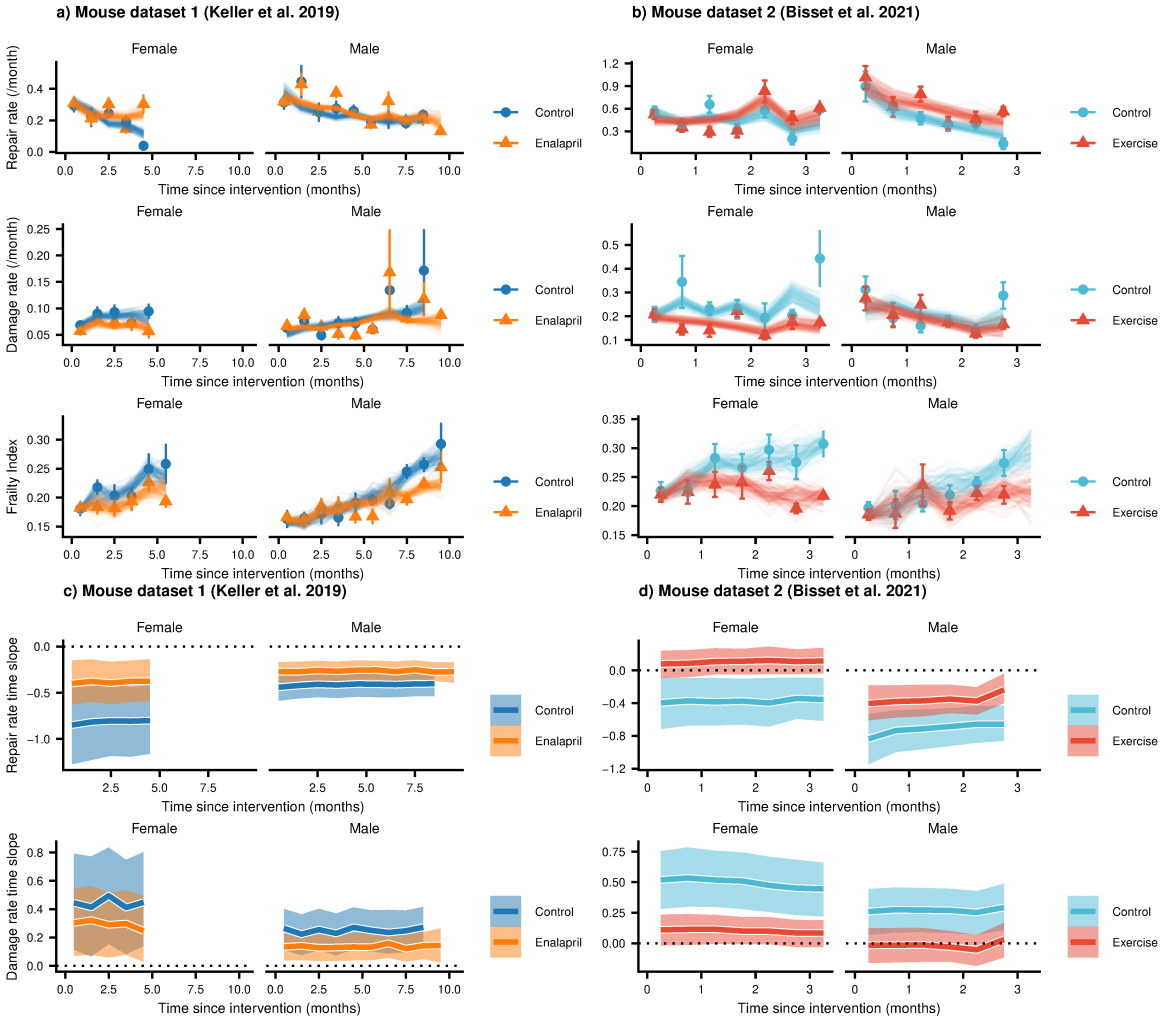


Figure 8.4: **Interventions both increase resilience and decrease damage.** Repair rates , damage rates, Frailty Index vs age for **a)** control mice and mice treated with enalapril from Keller *et al.*, 2019 [87] and for **b)** control mice and mice with voluntary exercise from Bisset *et al.*, 2021 [88]. Points in all plots represent binned averages of rates from the data with standard errors, and lines represent posterior samples from a Bayesian model of the rates. Repair rate and damage rate slopes vs age are shown for **c)** control mice and mice treated with enalapril from Keller *et al.*, 2019 [87] and for **d)** control mice and mice with voluntary exercise from Bisset *et al.*, 2021 [88]. 95% credible intervals of repair and damage rate time-derivatives are shown as the color filled regions.

8.3.4 Household wealth modifies human damage and repair rates

For humans, we use net household wealth as a socioeconomic environmental factor that serves as a proxy for medical and behavioral interventions that are not individually tracked. In Figure 8.5a), we show rates stratified by terciles of net household

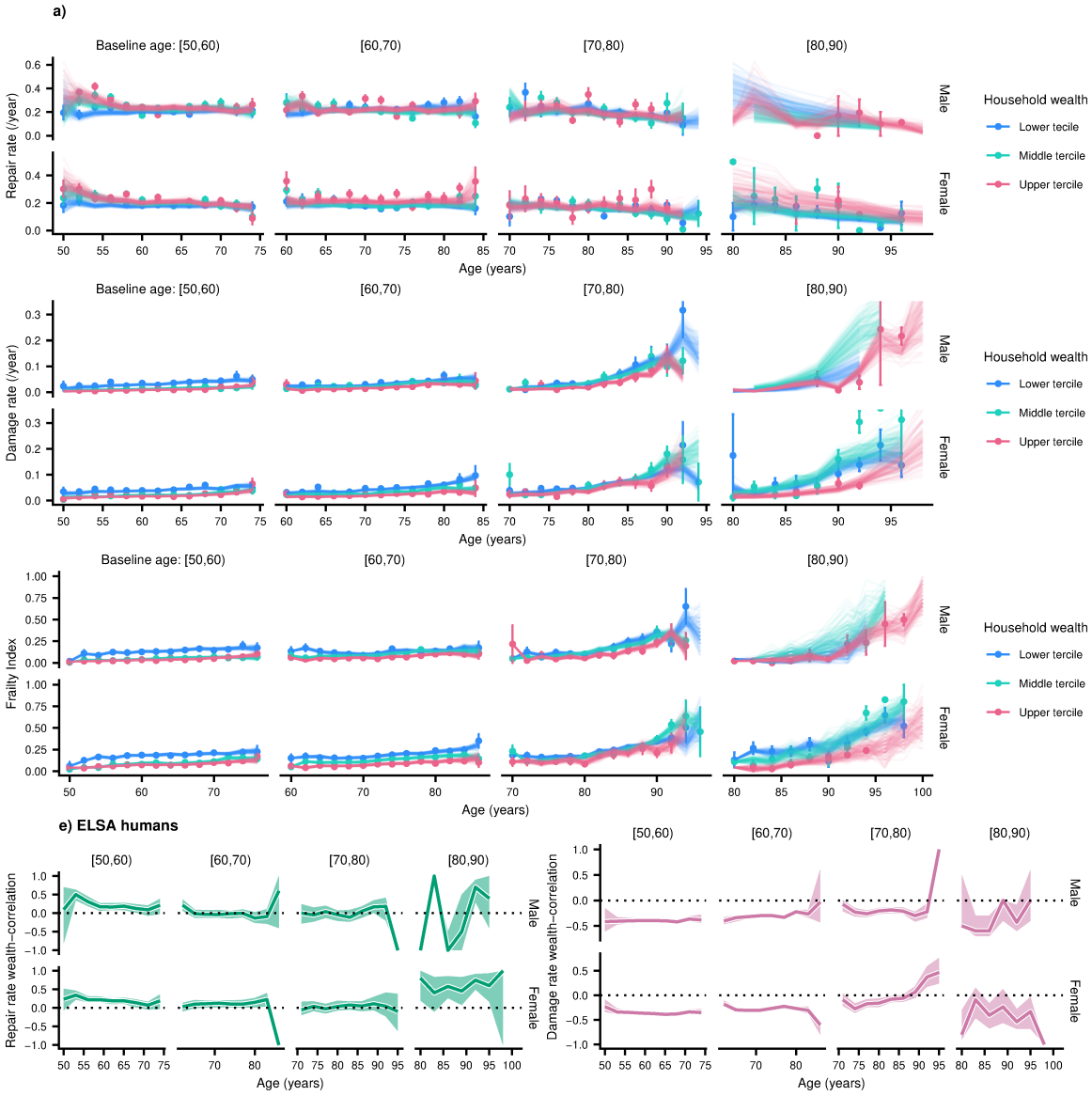


Figure 8.5: **Household wealth differentiates repair and damage rates in humans. a)** Repair rate, damage rate, and Frailty Index vs age for humans from the ELSA dataset, stratified by terciles of net household wealth. **b)** Spearman rank correlation ρ between repair rate and wealth and damage rate and wealth (wealth is continuous-valued, not terciles), separated by baseline age. Lines show the median of the posterior distribution of the correlation, colored regions show 95% credible intervals of the correlation.

wealth, where the lowest tercile has lower repair rates and higher damage rates for younger ages. Correspondingly, the Frailty Index is lower for individuals with a higher net household wealth. Since there is no intervention at time = 0, we treat wealth as a continuous variable and compute correlations between wealth and repair rate

rather than age-slopes (as was done for mice), shown in Figure 8.5b). We see that repair rates are positively correlated with net household wealth for younger ages, while damage rates are negatively correlated for younger ages.

These results replicate the findings in mice, where interventions impact both damage and repair rates. However, in humans we see that the effect of net household wealth is strongly age-dependent, having the largest effects at younger ages.

8.3.5 Damage and repair have broad time-scales

In the results above, we considered the average repair and damage rates vs age. Since individual deficits undergo stochastic transitions between damaged and repaired states (e.g. Figure 8.6a) we can also measure the time-scale of these individual states. However, observations are only taken at specific time-points and not all subjects were followed to death. This represents a mixture of interval-censored and right-censored time-to-event data, where the "event" is a damage or repair transition, rather than death in typical survival analysis. We estimate survival curves using an interval censored-analogue to the standard Kaplan-Meier estimator for right censored data (details in methods). These survival curves represent the probability of a deficit remaining undamaged after a repair event, which measures robustness, or remaining damaged after a damage event, which measures resilience.

The corresponding survival curves for the repaired or damaged states or shown in Figure 5, combining all deficits. Generally there is a large drop in probability at early times, indicating the states are rapidly switching at or below the interval between the measurements. However all of the curves also extend to very long times, indicating that both robustness and resilience have a broad range of time-scales, up to the remaining life of the organism.

Small effects due to enalapril and exercise are seen in mice on these time-scales, despite the effects seen on the average damage and repair rates vs age. This only small effect may be because we do not have sufficient data to stratify these curves by the time of initial damage or repair. For humans the effects due to household wealth are strong for females, and females with higher household wealth have a longer time-scale of damage and a shorter time-scale of repair.

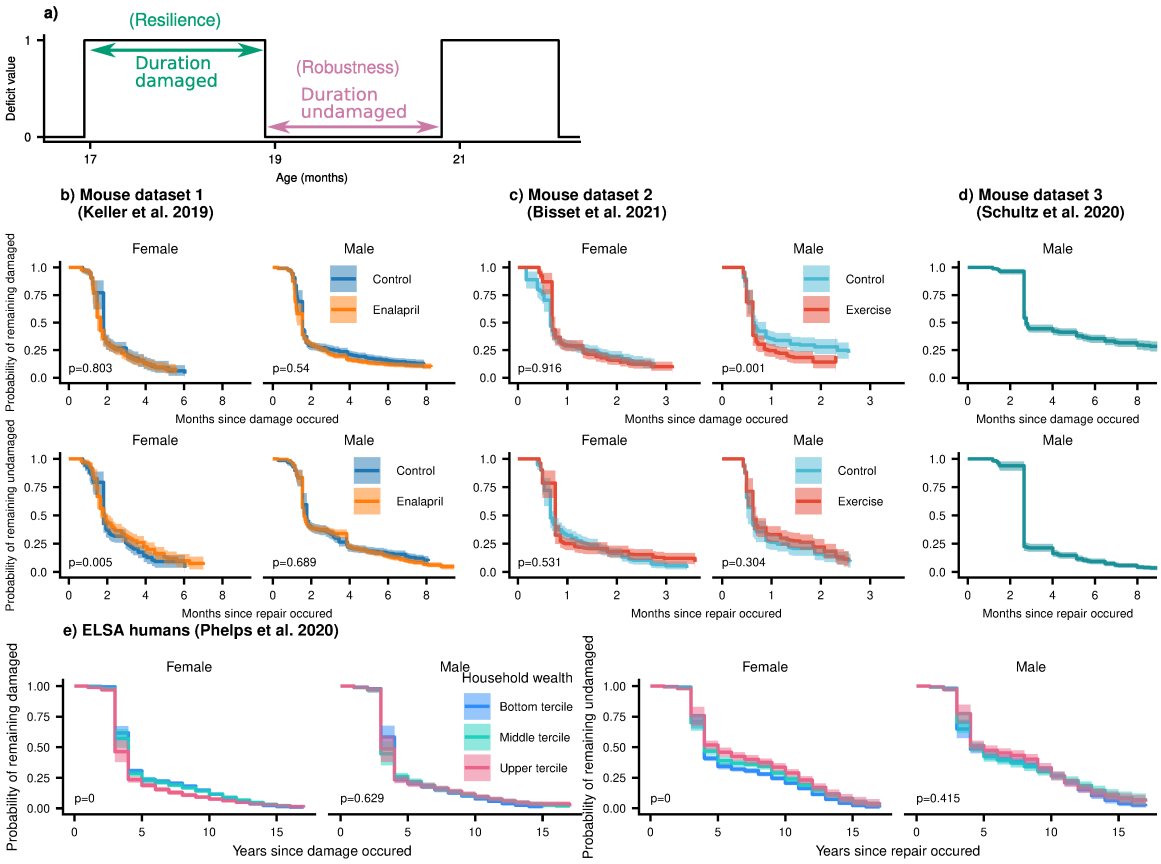


Figure 8.6: **Resilience occurs over both short and long time-scales in both mice and humans. a)** We measure the timescale of resilience of robustness with the lifetime of the damaged and undamaged deficit states. The corresponding state-survival curves are shown, which are the probability of remaining damaged vs months since damage occurred (resilience) and probability of remaining undamaged vs months since repair occurred (robustness). The colored regions show 95% posterior credible intervals for the survival curves. We show p-values in the left lower corner of each plot, which evaluate the difference between intervention/wealth groups with a generalized log-rank test for interval censored data [279]. These plots show **b)** Mouse dataset 1 (Keller *et al.*, 2019), **c)** Mouse dataset 2 (Bissett *et al.*, 2021), **d)** Mouse dataset 3 (Schultz *et al.*, 2020), and **e)** ELSA humans (Phelps *et al.* 2020).

These curves are also shown for each deficit and each dataset separately in Appendix Figures D.3 to D.8. In mice, deficits have distinct time-scales of resilience, with some deficits having very broad time-scales and other deficits very short. In humans, all deficits have similar time-scales of resilience, owing to the fact that all deficits are either ADLs or IADLs whereas the deficits in mice are quite distinct.

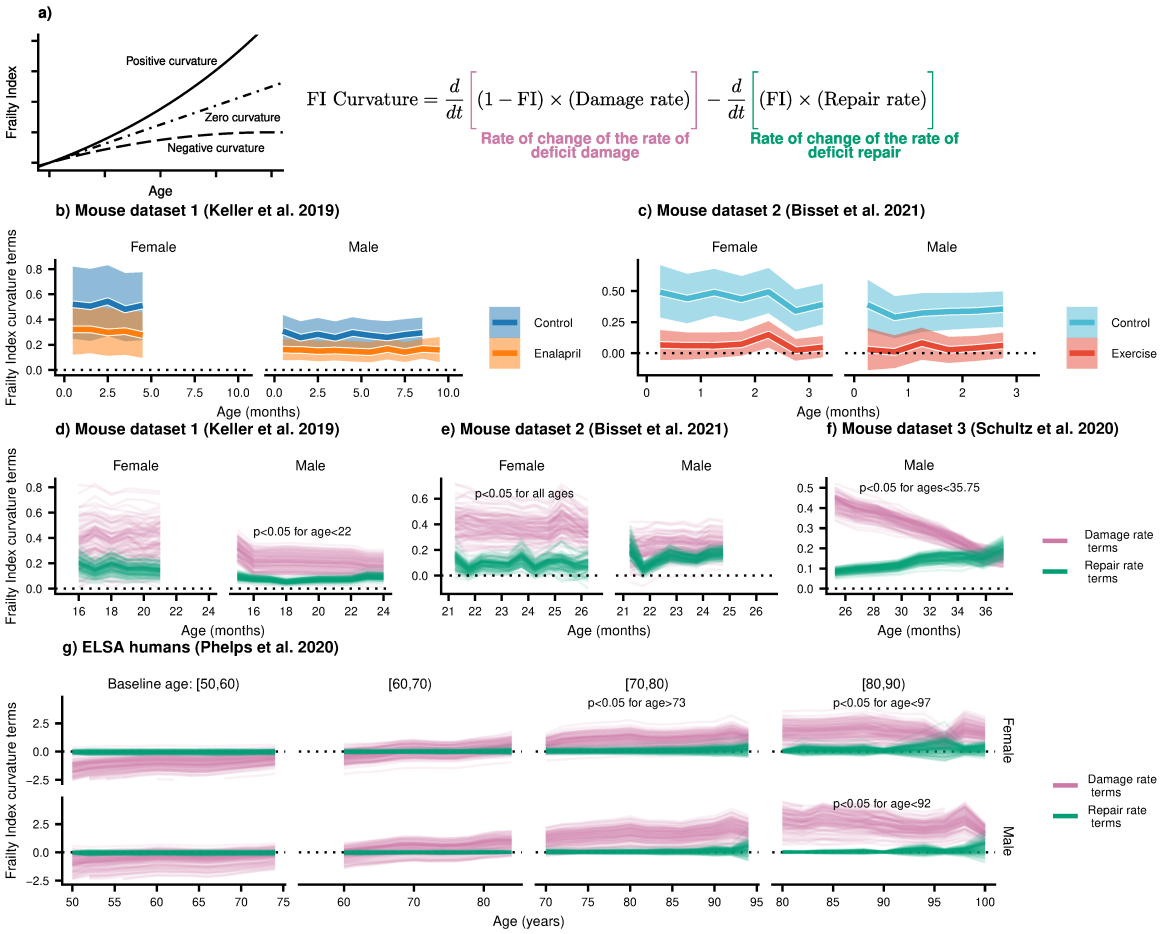


Figure 8.7: **Frailty Index curvature.** **a)** Frailty Index curvature measures the rate of accumulation of damage. Positive curvature indicates an acceleration of damage accumulation, zero curvature indicates a constant accumulation of damage, and negative curvature indicates a decelerating accumulation of damage. Curvature is computed with the second time-derivative of the Frailty Index (Equation 8.1). Terms of the curvature involving the repair rate (green) and the damage rate (pink) are separately shown. **b)** Enalapril treatment effect on Frailty Index curvature in Mouse dataset 1. Coloured regions show 95% credible intervals. **c)** Exercise treatment effect on Frailty Index curvature for Mouse group 2. Coloured regions show 95% credible intervals. **d)-g)** Frailty Index curvature terms (repair in green, damage in pink) for all 3 Mouse datasets and ELSA humans. Lines represent posterior samples from our Bayesian models.

8.3.6 Decreasing robustness determines Frailty Index curvature

The Frailty Index in Figures 8.3 and in other studies [14, 41, 66, 183] has been observed to have an acceleration or positive curvature, accelerating near death [280]. This positive curvature is also seen in other summary measures such as Physiological

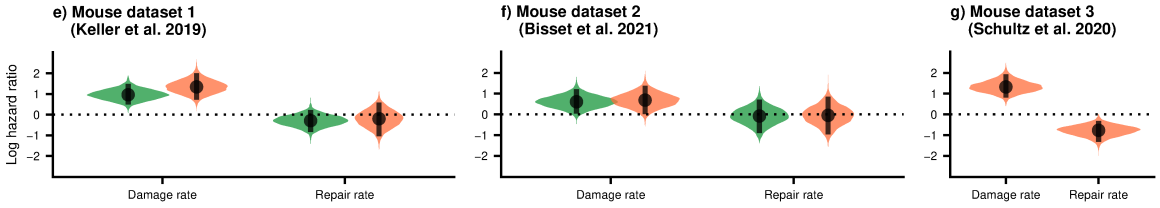


Figure 8.8: **Damage rates strongly impact survival in mice.** Log hazard ratio of death for the mice data sets. Hazard ratios are scaled so that a unit increase in damage rate or repair rate represents an increase by 1 standard deviation of the respective rate. Points show the median of the posterior distribution, violin distributions show the full posterior distributions, with the line showing the 95% posterior credible interval. **a)** Mouse dataset 1 (Keller *et al.*, 2019), **b)** Mouse dataset 2 (Bissett *et al.*, 2021), and **c)** Mouse dataset 3 (Schultz *et al.*, 2020).

Dysregulation [20]. However, there has been little work done to understand the origin of this curvature – whether it is due to a late-life decrease in resilience or a decrease in robustness.

We measure the curvature of the Frailty Index with the second time-derivative, which involves first derivatives of the damage and repair rates,

$$\frac{d^2}{dt^2}f(t) = \left[(1-f(t))\frac{d\lambda^d(t)}{dt} - \frac{df(t)}{dt}\lambda^d(t)\right] - \left[f(t)\frac{d\lambda^r(t)}{dt} + \frac{df(t)}{dt}\lambda^r(t)\right]. \quad (8.1)$$

In Figure 8.7, we observe the typical positive curvature, indicating an acceleration of damage accumulation, which is strongly effected by the interventions for mouse datasets 1 and 2 (Figure 8.7b) and c). In particular, we note that exercise nearly stops the acceleration in damage accumulation for mice.

In mouse datasets 1, 2, and 3, and humans (Figure 8.7d-g), we find that robustness or damage rates have the strongest effect on the FI curvature, particularly in the oldest age groups for humans (no effect of wealth is seen, and so is not shown). This indicates that it is the increase in susceptibility to damage with age rather than the decline of repair, that causes this acceleration of damage accumulation. These results highlight the role of declining robustness in aging, and not just declining resilience [94].

8.3.7 Robustness strongly impacts survival

The Frailty Index, as a measure of health, has been shown as a strong predictor of mortality for mice and humans [130]. We use our joint survival model to evaluate

the contributions of damage rate and repair rate to survival. With our model, we can evaluate the hazard ratio for damage rates and repair rates. Log hazard ratios represent the difference between the log hazard rate when damage or repair rates are increased by 1 standard deviation from the mean. For zero log hazard ratios, there is no effect of the variable on survival, and for positive or negative log hazard ratios there is a decrease or an increase in survival respectively.

In Figure 8.8, we find that the damage rate has significant (with 95% credible intervals around the median log hazard ratio above zero) and larger magnitude of log hazard ratios than repair rate, adjusted for Frailty Index score held constant. This shows that an increasing susceptibility to damage leads to larger decreases in survival than decline in resilience. This means that individuals survive longer when damage is avoided altogether, rather than repairing the damage after it occurs. This is an intuitive result, and a preliminary suggestion that interventions that focus on robustness should be favoured, rather than resilience.

In Figure 8.7 we had also shown that the decline in robustness has the strongest effect on the acceleration of damage accumulation for older ages. Both of these results demonstrate the importance of robustness for acceleration of damage at older ages, resulting in death [181].

While enalapril and exercise have been shown to slow the increase in the FI (Figure 8.4, [87,88]), no effects on survival have been found, although this may be due to the short period of time the mice were given these interventions [87,88]. Additionally, no survival information is available for the ELSA humans used in this work.

8.4 Discussion

Both humans and mice exhibit increasing accumulation of health related deficits with age [14,41,66,130,183]. Underlying this, we have demonstrated that both organisms have increasing damage rates and decreasing repair rates with age. Furthermore, the timescales of both repair and damage were broadly distributed from the earliest measurement interval to lifetime scales, and both damage and repair rates were able to be manipulated with interventions. We present a new approach for the assessment of damage (robustness) and repair (resilience) rates in longitudinal aging studies, and highlight fundamental similarities between mouse and human aging. Overall

these results imply that health in aging can be improved by both decreasing the accumulation of damage (or increasing robustness) and increasing the ability to repair (or increasing resilience), and that these factors may be observed and manipulated over longer time frames than previously expected.

Previous work has modeled the transitions between counts of deficits [38–41], however this work only modelled the mean number of deficits – they did not separate the damage and repair processes. These transition models have also been used for changes in indexes of cognition, again without separating repair and damage [281].

Our approach has similarities with some recent approaches to measuring resilience or robustness. In Pyrkov *et al.*, 2021 [96], resilience is measured by the ability of the organism to recover from stochastic fluctuations by computing the characteristic time-scale of the auto-correlation of physiological state variables. Other studies have also used auto-correlation as measures of resilience [273, 282]. In Arbeev *et al.*, 2019 [20], onset of disease is used to indicate a lack of robustness. In both of these approaches, there are continuous fluctuations or discrete transitions away from the healthy state, as in our approach. Other approaches such as Colon-Emeric *et al.*, 2021 [95] measure the difference between the recovered state after a specific acute stressor (hip fracture or viral respiratory infection) and the baseline state before the stressor, which is unlike our approach where we measure the tendency to recover after stochastic natural damage. An additional advantage to our approach in comparison to these other approaches is that we can observe both resilience and robustness together.

Summary measures of health such as the FI exhibit an accelerating accumulation of health deficits with age [280]. This universal behavior must be reflected in either increasing damage rates with age, or decreasing repair rates [96], or – as we find – both. Both increasing damage and decreasing repair rates with age are qualitatively consistent with common theories of aging [8, 52, 124]. However, the question of whether, and by what mechanisms, damage and repair processes are coupled during aging remains unanswered. Both damage and repair rates have been typically modelled as functions of the health state in descriptive models of aging [11, 44, 45], but without a mechanistic relationship between them apart from that imposed statistically by the observed accumulated damage. Here, we confirm that both mice and humans exhibit accelerating damage accumulation with age, but that damage and repair processes

are separately dynamic and targeted by interventions. This opens significant room for developing targeted or systematic interventions for both damage and repair.

Resilience characterizes organismal recovery from damage or disequilibrium, and can be measured as the rate of repair. We have shown that time-scales of repair appear broadly distributed - up to lifespan-scales. This was observed both in mice, where medical/lifestyle interventions are minimal, and humans where there are medical, dietary, exercise and other interventions over the lifespan that may be implicitly captured in our data. This implies that we may be able to target the increase of repair with interventions over a longer timeframe than just acutely when the damage occurs. The time scales we have studied are not easily accessible in challenge-response studies of resilience that typically occur over a much short timeframe [90]. As we have shown (Appendix Figures D.3, D.5, and D.7), different health attributes can have quite different recovery times, and future resilience studies across a range of health attributes should focus on timescales between our results and studies of controlled challenges. It is also possible that dichotomized deficits, which we have used, probe qualitatively different timescales than continuous measures studied in other studies of resilience. Nevertheless, we highlight that repair for many health attributes occurs at long timescales for both mice and humans.

Damage rates are much less directly studied, despite early definitions of frailty as susceptibility to damage. Robustness, or resistance to damage, is complementary to damage rate. As assessed in mice, damage rates appear to have a stronger effect on mortality than repair rates, indicating that reducing damage will have greater impact on mortality than systemically increasing repair however whether this is generally so, and the mechanisms behind this observation remain open questions. In humans, the effect of wealth is also stronger on damage rates than repair rates, indicating that the known strong effect of wealth on health and mortality, may be related to increased robustness due to environmental factors of higher socioeconomic status, rather than increased resilience. We have shown that, like repair processes, damage processes occur over a broad range of timescales from the shortest assessed to organismal lifespan scales, indicating a broad range of time for possible intervention. Further studies of damage rates are indicated in other model organisms over their life-course.

The effects of age on both damage and repair, in both mice and humans, are

qualitatively similar in male and female populations. Since is somewhat surprising, given the known sex effects in aging where, at the population level, females have poorer health than males, but are at lower risk of mortality [183]. At first thought, this paradox might imply that females have less robustness that males, but greater resilience. Instead in humans we find that males have both increased repair rates and decreased damage rates. Our findings imply, however, that this frailty-mortality sex paradox is not related to sex specific differences in resilience and robustness and that both sexes are equally vulnerable to increasing damage and decreasing repair in aging. Nevertheless, we have found that systemic interventions can have qualitatively distinct sex effects. The ACE inhibitor enalapril has a strong effect on reducing the timescales of repair in female, but not male mice. Previous work demonstrating that enalapril slows the increase of the FI in mice also found sex-specific effects on inflammation, possibly contributing to this difference between male and female mice [87]. Similarly, exercise prevents the decline of repair rates with age in female mice, but not in male mice. This preliminary evidence suggests that perhaps targeting resilience in females, and robustness in males will provide the most benefit in aging, although more studies are needed to support this claim. We show that assessing both damage and repair rates, and not just the health-state, in interventional aging studies can provide a clearer assessment of sex differences and suggest future studies adopt this approach.

The increasing availability of longitudinal health data over the lifespan of model aging organisms facilitates the analysis of damage and repair rates, and how they extend and change over the organismal lifespan. These damage and repair rates underlie the accumulation of damage that describes aging. While they have been studied as frailty or as acute resilience, they are seldom studied over the organismal lifespan. Here we have shown the value of considering both resilience and robustness over the lifespan. Further studies will be able to determine how widespread organismal and sex differences in these effects are, or how universal they may yet prove to be. It remains to be seen whether our results in mice and humans generalize to other portfolios of health attributes and whether the same universality is exhibited in damage and repair rates that has been observed with deficit accumulation. The mechanisms underlying differing age effects in different organisms, and behind the differing timescales of recovery in different health attributes will be important to study in detail. Studies on

the effects on damage and repair rates of both targeted and systemic interventions will also be crucial. We have studied only three (enalapril and exercise in mice, and wealth in humans), but there are many more.

8.5 Methods

8.5.1 Mouse data

For the mouse portion of this manuscript, published data on longitudinal health-related deficits in C57BL/6 mice from three papers was used (Keller *et al.*, 2019 [87]; Bisset *et al.*, 2021 [88]; Schultz *et al.*, 2020 [253]). A brief summary of the methods of each paper is below:

Mouse group 1

Keller *et al.*, 2019: Male and female C57BL/6 mice were assessed for deficits approximately every 4 weeks from 16 to either 21 months of age (females) or 25 months of age (males). Mice were fed either a diet containing enalapril (280 mg/kg) or control diet for the duration of the experiment.

After pre-processing (below), this data contains 21 female control mice, 25 female enalapril mice, 13 male control mice, and 25 male enalapril mice.

Mouse group 2

Bisset *et al.*, 2021: Male and female C57BL/6 mice were assessed for deficits approximately every 2 weeks from 21 to 25 months of age. Mice were all singly housed, and half were provided a running wheel for voluntary exercise.

After pre-processing (below), this data contains 11 female control mice, 11 female exercise mice, 6 male control mice, and 6 male exercise mice.

Mouse group 3

Schultz *et al.*, 2020: Male C57BL/6Nia mice were assessed for deficits approximately every 6 weeks from 21 months of age until their natural deaths.

After pre-processing (below), this data contains 44 male control mice.

Mouse clinical frailty index assessment

Each of the papers above assessed health deficits using the mouse clinical frailty index as described previously [283]. Briefly, this assessment involves scoring 31 non-invasive health-related measures in mice. Most measures are scored as 1 for a severe deficit, 0.5 for a moderate deficit and a 0 for no deficit. Deficits in body weight and temperature were scored based on deviation from reference values in young adult animals, such that a difference of less than 1 SD was scored 0, a difference of ± 1 SD was scored 0.25, a difference of ± 2 SD was scored 0.5, a difference of ± 3 SD was scored 0.75, and a difference of more than 3 SD received the maximal deficit value of 1 [283]. The deficits of malocclusions and body temperature were not assessed in mouse group 3 [253], leaving only 29 deficits for this group.

Mouse data pre-processing

For mouse dataset 1, we impute missing deficit values by propagating the last observed value forward. If the first observed deficit is missing, it is imputed by propagating the first observed value backward. Less than 1% of all total deficit values are missing in this dataset. No values in the other datasets are missing.

Deficits are scored on a fractional scale, with deficit i having values $d_i \in \{0, 0.25, 0.5, 0.75, 1\}$. To treat these as binary, we represent each fractional deficit d_i by 4 ordered binary deficits, $[d_i^{(1)}, d_i^{(2)}, d_i^{(3)}, d_i^{(4)}]$. Fractional deficits are then represented by setting $4 \times d_i$ of these ordered binary deficits to 1. For example if $d_i = 0.75$, this is represented as $[1, 1, 1, 0]$.

A new Frailty Index is then created by taking all of these new binary deficits, representing a $4 \times 31 = 124$ item Frailty Index ($4 \times 29 = 116$ for mouse dataset 3). This process preserves the FI scores, and a single repair or damage event can be interpreted as taking a step of size 0.25 on the fractional deficit scale.

Measurement times with abnormally short or long intervals are removed. In mouse dataset 2, measurement times less than 0.1 months from the previous time are removed. In mouse dataset 3, measurement times more than 2 months from the previous time are removed.

In each dataset, mice with less than 2 observed time-points are removed.

8.5.2 Human data and pre-processing

We use human data from the English Longitudinal Study of Aging [277]. We select individuals that have full data for net household wealth and activities of daily living (ADL) and instrumental activities of daily living (IADL). A Frailty Index is created from the count of 10 possible ADLs and 13 possible IADLs, giving a fraction out of 23. We restrict individuals to those that were recruited to the study between the ages of 50 and 89 years. We drop individuals with follow-up time intervals above 4 years and individuals with fewer than 6 follow-ups.

We use net household wealth, as determined in the financial assessment in wave 5 of the ELSA data. We drop individuals that have parts of this assessment imputed. The raw value of net household wealth spans several orders of magnitude (and includes negatives for individuals in debt), and so is transformed by $w = \log\big(w_{\text{raw}} + \text{mean}(w_{\text{raw}})\big)$.

After pre-processing, this data contains 1074 males and 1422 females with time-intervals of approximately 2 years between observations. There are 1211 individuals from baseline ages in $[50, 60)$, 833 with baseline ages in $[60, 70)$, 280 with baseline ages in $[70, 80)$, and 21 with baseline ages in $[80, 90)$.

8.5.3 Extracting raw damage and repair counts from datasets

In each dataset, we observe the state of N binary health deficits $\{d_{jk}\}_{k=1}^{N}$ for each subject at a set of observation times $\{t_j\}_{j=1}^{J}$. Summing up the number of deficits at each time, we get deficit counts for each observation time, $\{n_j\}_{j=1}^{J}$, which is used to compute the Frailty Index $f_j = n_j/N$.

We compute the number of deficits damaged ($0 \to 1$ transitions) and repaired ($1 \to 0$ transitions) between two time points t_j and t_{j+1}, denoted as $n^d(t_j)$ or $n^r(t_j)$. These counts satisfy $n(t_{j+1}) = n(t_j) + n^d(t_j) - n^r(t_j)$, linking these damage and repair processes with the Frailty Index.

8.5.4 Modelling

We model deficit repair and damage as Poisson point processes with time-dependent rates, $\lambda^r(t)$ and $\lambda^d(t)$. The count of deficits repaired or damaged in an interval $[t_1, t_2]$

is assumed to follow a Poisson distribution, with mean equal to the instantaneous rate $\lambda^r(t)$ or $\lambda^d(t)$ integrated over this interval times the number of possible deficits available to be repaired $\Lambda^r(t) = \int \lambda(t)^r n_t dt$, or damaged $\Lambda^d(t) = \int \lambda^d(t)(N - n_t)dt$. We assume these time intervals are small so that we use constant rates within each time-interval to approximate these integrals, $\Lambda^r(t) \approx \lambda^r(t) n_t \Delta t$ or $\Lambda^d(t) \approx \lambda^d(t)(N - n_t)\Delta t$.

Joint longitudinal-survival model for mice datasets

We use a joint modelling framework to model repair and damage rates, while assessing their effect on survival. We decompose the multivariate joint distribution of the observed longitudinal damage and repair counts and survival outcome by coupling survival with the repair and damage rates $\lambda_i^r(t)$ and $\lambda_i^d(t)$ [121, 284],

$$p(T_i, c_i, \{n_i^r(t)\}, \{n_i^d(t)\} | \lambda_i^r(t), \lambda_i^d(t)) = p(T_i, c_i | \lambda_i^r(t), \lambda_i^d(t)) p(\{n_i^r(t)\}, \{n_i^d(t)\} | \lambda_i^r(t), \lambda_i^d(t)). \tag{8.2}$$

A graphical version of this model is shown in Figure 8.9.

Longitudinal component

We use a linear Poisson model for the longitudinal damage and repair rates. A Softplus function, $\log(1 + e^x)$, is used to enforce positive rates. This function is chosen because it is approximately linear for larger values of x, in contrast to e^x which is often used for Poisson models. The form of this model is,

$$\lambda_i^r(t) = \text{Softplus}\big(\boldsymbol{\beta}^r \cdot \mathbf{x}_i(t) + b_{i,0}^r + b_{i,1}^r t\big), \tag{8.3}$$

$$\lambda_i^d(t) = \text{Softplus}\big(\boldsymbol{\beta}^d \cdot \mathbf{x}_i(t) + b_{i,0}^d + b_{i,1}^d t\big), \tag{8.4}$$

$$n_i^r(t_j) | \lambda_i^r(t_j), n_i(t_j) \sim \text{Poisson}\Big(n_i(t_j) \lambda_i^r(t_j)(t_{j+1} - t_j)\Big), \tag{8.5}$$

$$n_i^d(t_j) | \lambda_i^d(t_j), n_i(t_j) \sim \text{Poisson}\Big((N - n_i(t_j)) \lambda_i^d(t_j)(t_{j+1} - t_j)\Big), \tag{8.6}$$

$$n_i(t_{j+1}) = n_i(t_j) + n_i^d(t_j) - n_i^r(t_j). \tag{8.7}$$

The first two equations describe the time-dependent repair and damage rates, $\lambda^r(t)$ and $\lambda^d(t)$. These rates represent the probability of repair or damage, per deficit per unit time. These rates are multiplied by the number of deficits that can repair $n(t_j)$

or the number that can damage $N - n(t_j)$ and the time-interval $t_{t_{j+1}} - t_j$ to compute the mean count of repaired or damaged deficits for Poisson distributions. The last Equation 8.7 shows how we can compute the total count of deficits from this model; allowing this model to be used for the Frailty Index as well.

The full-cohort parameters are denoted $\boldsymbol{\beta}$ and the subject-specific intercept and time-slopes $b_{i,0}, b_{i,1}$. The variables $\mathbf{x}_i(t)$ include the covariates and their interactions,

$$\begin{aligned} \mathbf{x}_i(t) &= (1,\, t,\, \text{sex},\, \text{treatment},\, f,\, a_0,\, \text{sex} \times \text{treatment}, \\ &\qquad \text{sex} \times t,\, \text{treatment} \times t, \text{sex} \times \text{treatment} \times t). \end{aligned} \tag{8.8}$$

The "treatment" variable is a 0/1 indicator for enalapril in mouse group 1 or exercise in mouse group 2. The other variables are the time from baseline t, the frailty index f, baseline age a_0, and sex (M/F). These interactions allow sex and treatment group specific age slopes.

The repair and damage processes are linked by including correlation between the subject-specific parameters $[\boldsymbol{b}_i^r, \boldsymbol{b}_i^d] \sim \mathcal{N}(0, \boldsymbol{\Sigma})$.

Survival component

We jointly model these repair and damage processes with survival, with proportional hazards and a baseline hazard parameterized with M-splines [285] (which are always non-negative). The damage and repair processes are linked with survival by including damage rate and repair rate in the hazard rate,

$$h_i(t) = h_0(t, \text{sex}) \exp\Big(\boldsymbol{\gamma} \cdot \mathbf{u}_i(t) + \gamma^r \text{Softplus}^{-1}\lambda_i^r(t) + \gamma^d \text{Softplus}^{-1}\lambda_i^d(t)\Big), \tag{8.9}$$

$$\begin{aligned} h_0(t) &= (\text{male}) \cdot \sum_{l=1}^{L} a_{l,\text{male}} M_{l,3}(t|\mathbf{k}) \\ &+ (\text{female}) \cdot \sum_{l=1}^{L} a_{l,\text{female}} M_{l,3}(t|\mathbf{k}), \;\; \sum_{l=1}^{L} a_l = 1, \;\; a_l \geq 0, \end{aligned} \tag{8.10}$$

$$S_i(t) = \exp\big(-\int_{t_0}^{t} h_i(s)ds\big). \tag{8.11}$$

The first equation describes the hazard rate $h_i(t)$ in terms of the covariates $\mathbf{u}_i$ and the repair and damage rates. The baseline hazard $h_0(t, \text{sex})$ is modeled with sex-specific splines in Equation 8.10, due to the large disparity in survival by sex. The covariates are $\mathbf{u}_i = (1, \text{sex}, \text{treatment}, f, a_0)$.

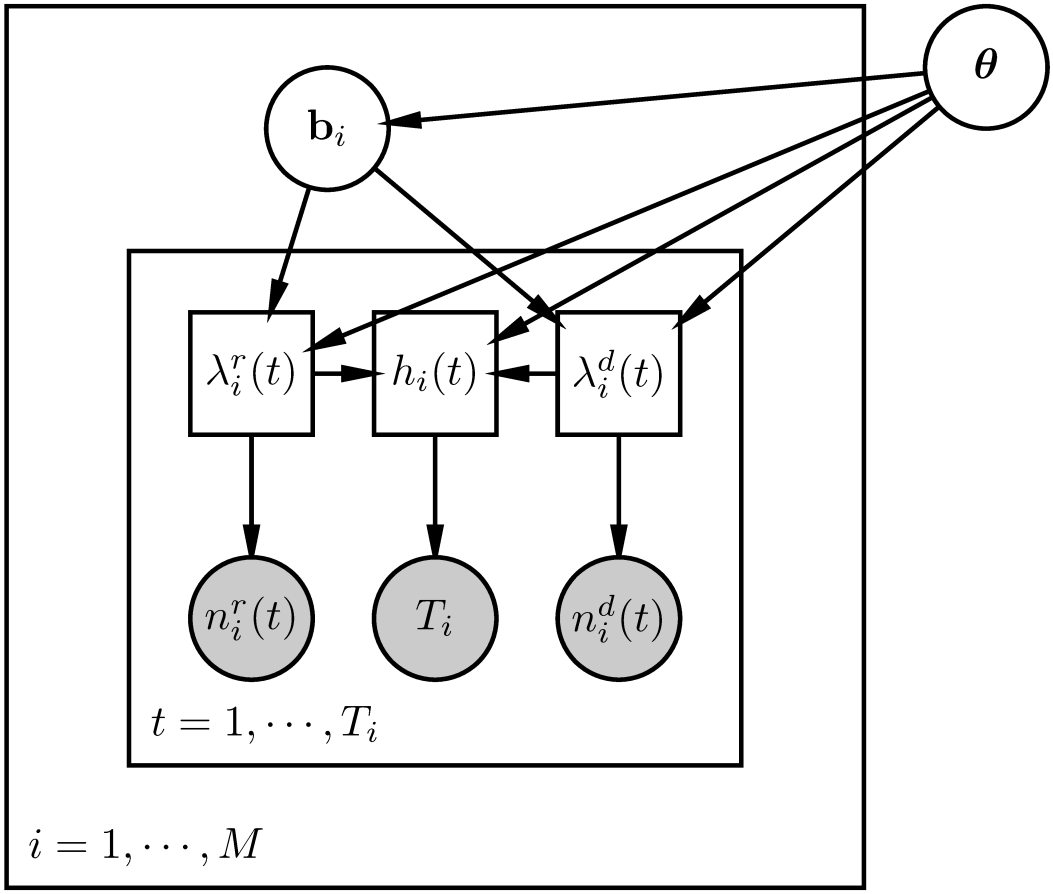


Figure 8.9: **Poisson joint model for damage and repair rates.** This graph summarizes the Poisson joint model used to extract age-dependent damage and repair rates, coupled with survival. The arrows describe conditional dependencies between the variables, describing the factorization of the model. White nodes are unobserved, grey nodes are observed. Indexes i indicate individual subjects, and indexes t represent individual times. The repair counts $n_i^r(t)$ and damage counts $n_i^d(t)$ depend on the repair rate $\lambda_i^r(t)$ and damage rate $\lambda_i^d(t)$. The repair rate, damage rate and hazard rates are not stochastic variables but just intermediate variables, and so are shown as squares. The time of death or time of censoring T_i depends on the hazard rate $h_i(t)$. The parameters $\mathbf{b}_i$ are subject-specific coefficients of the damage and repair rates. The parameters $\boldsymbol{\theta}$ are global parameters of the entire model.

Priors and hyperparameters

We use weakly informative priors to regularize parameters,

$$\begin{aligned}
\beta_0^r,\ \beta_0^d,\ \gamma_0 &\sim \mathcal{N}(0,3), && (8.12)\\
\boldsymbol{\beta}^r,\ \boldsymbol{\beta}^d,\ \boldsymbol{\gamma},\ \gamma^r,\ \gamma^d &\sim \mathcal{N}(0,1), && (8.13)\\
\boldsymbol{\sigma} &\sim \text{HalfCauchy}(0,1), && (8.14)\\
\boldsymbol{\Omega} &\sim \text{LKJ}(2), && (8.15)\\
\boldsymbol{\Sigma} &= \boldsymbol{\sigma}\boldsymbol{\Omega}\boldsymbol{\sigma}, && (8.16)\\
[\mathbf{b}_i^r,\ \mathbf{b}_i^d] &\sim \mathcal{N}(0,\boldsymbol{\Sigma}), && (8.17)\\
\mathbf{a} &\sim \text{Dirichlet}(1.0, L=17), && (8.18)\\
\boldsymbol{\kappa} &= \{\min(\{T_i\}_i), Q_{0.05}(\{T_i\}_i), ..., && (8.19)\\
&\qquad Q_{0.95}(\{T_i\}_i), \max(\{T_i\}_i)\}.
\end{aligned}$$

Broad $\mathcal{N}(0,3)$ priors are used on intercept parameters and narrow $\mathcal{N}(0,1)$ priors on covariate coefficients. The covariance matrix $\boldsymbol{\Sigma}$ for the coupling of the subject-specific parameters $\mathbf{b}_i$ is decomposed in terms of a correlation matrix $\boldsymbol{\Omega}$ with a LKJ prior and standard deviations $\boldsymbol{\sigma}$ with half-Cauchy distributions. Spline coefficients $\mathbf{a}$ use a Dirichlet distribution with concentration 1, representing a uniform prior on the simplex $\sum_{l=1}^{L} a_l = 1,\ a_l \geq 0$. We use $L = 17$ spline knots with knots at the minimum last follow-up age, the maximum, and 15 uniformly spaced quantiles from 0.05 to 0.95 of the last follow-up age.

Integrals of the hazard rate are computed with 5-point Gaussian Quadrature between each observed time interval.

Non-linear modeling for human data

There is much more human data than mice and the data is more complex, where linear effects are not sufficient to capture the combined influence of wealth, baseline age, and time. We use a non-linear Poisson model with non-constant coefficients to include addition degrees of freedom. We parameterize these non-constant coefficients with B-splines. Additionally, the individuals selected from ELSA with wealth data do not have mortality data available, simplifying the model from the joint model used above for mice.

Our model has the form,

$$\lambda_i^r(t) = \text{Softplus}\big(\boldsymbol{\beta}_0^r \cdot \mathbf{x}_i(t) + \beta_1^r(w,a_0) + \beta_2^r(w,a_0) \times \text{sex} + \beta_3^r(w,a_0) \times t \tag{8.20}$$
$$+ \beta_4^r(w,a_0) \times \text{sex} \times t + b_{i,0}^r\big),$$
$$\lambda_i^d(t) = \text{Softplus}\big(\boldsymbol{\beta}_0^d \cdot \mathbf{x}_i(t) + \beta_1^d(w,a_0) + \beta_2^d(w,a_0) \times \text{sex} + \beta_3^d(w,a_0) \times t \tag{8.21}$$
$$+ \beta_4^d(w,a_0) \times \text{sex} \times t + b_{i,0}^d\big),$$
$$n_i^r(t_j)|\lambda_i^r(t_j), n_i(t_j) \sim \text{Poisson}\Big(n_i(t_j)\lambda_i^r(t_j)(t_{j+1} - t_j)\Big), \tag{8.22}$$
$$n_i^d(t_j)|\lambda_i^d(t_j), n_i(t_j) \sim \text{Poisson}\Big(N - n_i(t_j)\lambda_i^d(t_j)(t_{j+1} - t_j)\Big), \tag{8.23}$$

where

$$\mathbf{x}_i(t) = (1,\, t,\, \text{sex},\, w,\, f,\, a_0,\, \text{sex} \times t,\, w \times t,\, a_0 \times t,\, \text{sex} \times f,\, w \times f,\, a_0 \times f). \tag{8.24}$$

The non-constant coefficients $\{\beta_k(w,a_0)\}_k$ are implemented as 2D B-splines for wealth and baseline age with 5 wealth knots and 5 baseline age knots at the minimum, maximum and terciles of these variables. We use smoothing 2D random-walk priors on the spline coefficients,

$$\begin{aligned}
s_{11}, \tau_w, \tau_{b_0} &\sim \mathcal{N}(0,1), && (8.25)\\
p_w, p_{b_0} &\sim \text{Dirichlet}(1.5), && (8.26)\\
s_{ij} &\sim p_w\mathcal{N}(s_{i-1,j}, \tau_w) + p_{b_0}\mathcal{N}(s_{i,j-1}, \tau_{b_0}), && (8.27)\\
\beta_k(w,a_0) &= \sum_{i,j=1}^{5} s_{ij} B_{i,3}(w;\boldsymbol{\kappa}_w) B_{j,3}(a_0;\boldsymbol{\kappa}_{a_0}). && (8.28)
\end{aligned}$$

All other priors are the same as in the mouse modelling.

Note, in the human data we do not include subject-specific time-slopes $b_{i,1}^r$ and $b_{i,1}^d$ as we did in the mouse data, since we have much shorter time-series. When these slopes are included, we see evidence of the model over-fitting to the data by the proportion of residuals with zero within 95% credible intervals being much higher than 0.95, and nearing 0.99 to 1.

Derivatives

We can compute the derivative of the Frailty Index according to the modelled repair and damage rates,

$$\frac{d}{dt} f_i(t) = (1 - f_i)\lambda_i^d(t) - f_i\lambda_i^r(t). \tag{8.29}$$

To understand the effect of interventions, we compute the derivative with respect to time for the repair and damage rates,

$$\begin{aligned}
\frac{d}{dt}\lambda_i^r(t) &= \frac{\partial\lambda_i^r(t)}{\partial t} + \frac{\partial\lambda_i^r(t)}{\partial f}\frac{df_i(t)}{dt}, \\
&= \left(\boldsymbol{\beta}^r \cdot \frac{d\mathbf{x}_i(t)}{dt} + \boldsymbol{\beta}^r \cdot \frac{d\mathbf{x}_i(t)}{df} + \mathbf{b}_i^r \cdot \frac{d\mathbf{z}_i(t)}{dt}\right)\frac{e^{\lambda_i^r(t)}}{e^{\lambda_i^r(t)}+1}, \qquad (8.30) \\
\frac{d}{dt}\lambda_i^d(t) &= \frac{\partial\lambda_i^r(t)}{\partial t} + \frac{\partial\lambda_i^r(t)}{\partial f}\frac{df_i(t)}{dt}, \\
&= \left(\boldsymbol{\beta}^d \cdot \frac{d\mathbf{x}_i(t)}{dt} + \boldsymbol{\beta}^d \cdot \frac{d\mathbf{x}_i(t)}{df} + \mathbf{b}_i^d \cdot \frac{d\mathbf{z}_i(t)}{dt}\right)\frac{e^{\lambda_i^d(t)}}{e^{\lambda_i^d(t)}+1}. \qquad (8.31)
\end{aligned}$$

This is the slope of these rates vs time, with the increase in the Frailty Index $f(t)$ included. While we only include explicit linear effects of time in the model, the increase in Frailty Index with time can influence the derivative to change.

We can compute the curvature as the second derivative of the Frailty Index with age, written in terms of first derivatives of the rates,

$$\frac{d^2}{dt^2}f_i(t) = \left[(1-f_i(t))\frac{d\lambda_i^d(t)}{dt} - \frac{df_i(t)}{dt}\lambda_i^d(t)\right] - \left[f_i(t)\frac{d\lambda_i^r(t)}{dt} + \frac{df_i(t)}{dt}\lambda_i^r(t)\right]. \quad (8.32)$$

The first group of terms are those involving damage rate (robustness) and the second group of terms are those involving repair (resilience). These terms are plotted in Figure 8.7.

Repair and damage timescale mice and human data

We observe the amount of time that has passed between damage and repair events, and vice versa. This can be used to determine the time-scales of these damage and repair processes. However, since a deficit might damage and the individual dies before the deficit is ever repaired, there is right censoring. Additionally, observations are only made at specific time-points and so we cannot determine the exact time at which a deficit damaged or repaired, there is interval censoring. To estimate the distribution of repair and damage times, we treat repair and damage events for each deficit as a mixture of interval and right censored events.

We use a Bayesian survival model with M-splines for the baseline hazard,

$$h(t) = e^{\gamma_0} \sum_{l=1}^{L} a_l M_{l,3}(t|\mathbf{k}), \quad \sum_{l=1}^{L} a_l = 1, \tag{8.33}$$

$$S(t) = \exp\big(-\int_{t_0}^{t} h(s)ds\big). \tag{8.34}$$

This is fit separately for sex and control/intervention groups.

We include both interval censoring and right-censoring in the likelihood,

$$p(T_i^L, T_i^U, T_i, c_i|\{a_l\}_l, \gamma_0) = [S(T_i)]^{c_i}[S(T_i^L) - S(T_i^U)]^{1-c_i}, \tag{8.35}$$

where T^L is the lower interval bound, T^U is the upper interval bound, and T is a time of right censoring. We use 32 knots set at 30 evenly spaced quantiles of the event times from 0.1 to 0.9 together with the minimum and maximum event time. Dirichlet$(1.0, 32)$ prior is used for the spline coefficients and a $\mathcal{N}(0, 10)$ prior for γ_0.

8.5.5 Sampling

We use the STAN no U-turn sampler (NUTS) [113]. We use 4000 warm-up iterations and 6000 sampling iterations on 4 separate chains for the mice joint models. For the human model we use 2 separate chains with 1000 warm-up iterations and 3000 sampling iterations. For the interval-censored Bayesian survival models we use 2000 warm-up iterations and 3000 sampling iterations for 4 separate chains.

Model diagnostics

In the Appendix Figures D.9, D.10, and D.11 we perform a posterior predictive checks for the mice joint models by plotting the residuals of simulated data compared to the observed data, and compute R^2 statistics [286–288]. In Appendix Figures D.12, D.13, and D.14 show residuals for the human model.

Chapter 9

Conclusions and future outlook

The work of this thesis is an attempt at building computational models of aging. These models are not merely predictive models, but can be used to understand aging phenomena. In the introduction to this thesis in Chapter 1, I introduced three fundamental questions that this thesis aims to answer: how can we better understand the mechanisms or causes underlying observed aging phenomena, how can we better predict outcomes at an individual or population level, and, finally, how can we better intervene to decrease mortality and to improve health during aging? I have demonstrated the use of complex computational models to answer questions such as these in aging.

To facilitate the understanding of aging mechanisms, I used the Generic Network Model (GNM) to understand the structure of the different levels of health included in the Frailty Index with complex networks (Chapter 3). I developed the Weighted Network Model to describe the accumulation of binary deficits in aging with a network of average pairwise rates that is inferred from cross-sectional aging data (Chapter 4). I developed the Dynamic Joint Interaction Network (DJIN) model to make quantitative predictions of aging trajectories and infer physiological interaction networks governing the dynamics, and explore the dimensionality of aging (Chapter 5). I developed a latent variable clustering model to infer distinct aging phenotypes (Chapter 7). I developed a novel analysis technique to better understand the processes of damage and repair (robustness and resilience) in aging (Chapter 8).

The WNM model from Chapter 4, the DJIN model from 5, and the clustering model from Chapter 7 are also predictive models, capable of predicting individual health trajectories and mortality. While we were able to use long time-series of aging data in the clustering model for *C. elegans*, this was not available for humans and limited the DJIN model. With *C. elegans*, we are able to condition predictions on long input time-series (50 time points). In contrast, in the DJIN model we only used

a single baseline time-point for predictions. This limitation reflects the availability of longitudinal data of aging.

Figure 9.1 shows examples of the ways in which the DJIN model can be extended in future work. In the DJIN model, we needed to impute the baseline health state. However, we only used data at the baseline age to do this imputation. Figure 9.1a) shows an approach to imputation using multiple time-points, using a recurrent neural network instead of a feed-forward neural network. This approach would require longer time-series than are available in the ELSA data that was used in this thesis. By using multiple input points for imputation, we can utilize the trends in variables for imputation, not just the correlations between values at the same time-point. A similar approach can be used to update the current health state, in prediction scenarios with tracked health data [221].

The DJIN model assumes constant pair-wise interactions between all variables. However, in reality these interactions are likely modulated by other variables. Instead of inferring a set of constant network weights $\{W_{ij}\}_{ij}$, we can fit a complex function (e.g. a neural network) mapping auxiliary variables $\mathbf{u}$ to an inferred network $\{W_{ij}(\mathbf{u})\}_{ij}$, allowing the network interactions to be modified by these input variables, shown in Figure 9.1b).

The DJIN model infers interactions that are useful for predictions, this is not true causality, but "predictive causality" [224–226]. While we do not know if there are unobserved variables confounding the interactions, these connections suggest possible avenues of intervention. The network suggests the effects of increasing or decreasing a variable on a second variable, based on the the sign and direction of interactions. For example in Figure 9.1c), if we intervene to decrease x_4, the model predicts that x_3 would decrease. This suggests possible interventions to modify x_3, and is a method of verifying inferred connections.

Additionally, while our approach interprets the model in terms of interactions between the health variables, we do not attempt to interpret mortality predictions. Such an interpretation would indicate which variables influence mortality rates most strongly. The simplest way to do this would be to replace the complex RNN-based mortality rate with simple proportional hazards [119] or quadratic hazards [289], which describe the mortality rate in terms of simple functions. However, this would

lose the advantages gained with the RNN mortality rates. A better approach would require developing a method of interpreting these rates while retaining the memory of past states.

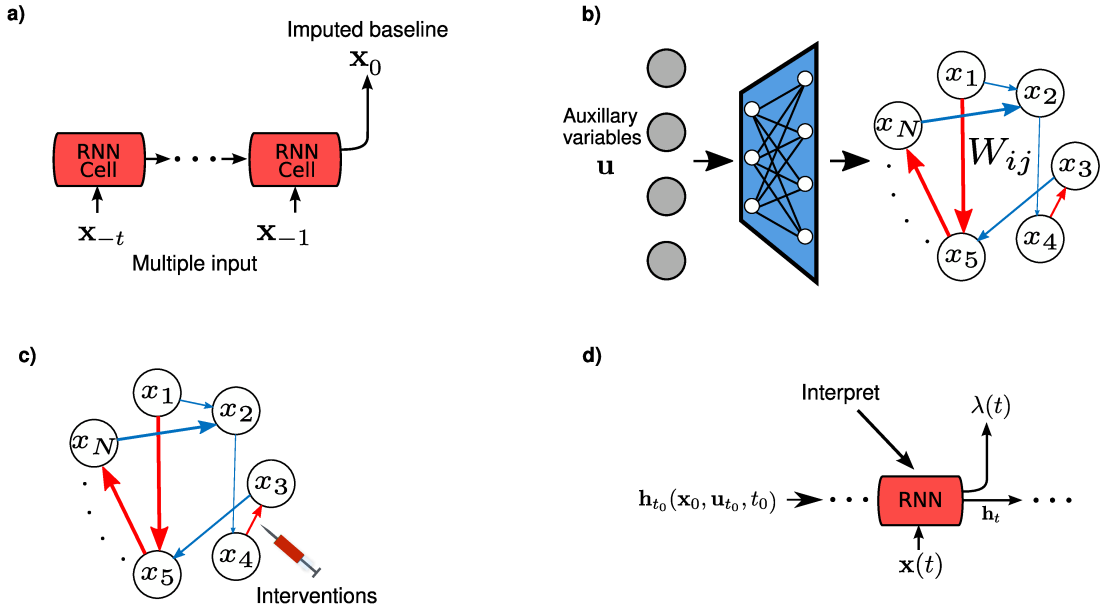


Figure 9.1: **Extensions of the DJIN model. a)** Time-dependent imputation, or updating the current health state with tracked health data. **b)** Instead of constant network interactions, the interactions by be modulated by other variables. In this case, the interaction network can be output by a complex function (blue trapezoid) of the auxiliary variables $\mathbf{u}$ controlling the interactions. **c)** The interactions inferred by the model can be probed with intervention studies, since the model infers both the direction of the interaction and the sign of influence. For a connection between nodes $i \to j$, the model makes predictions about the result on j when i is changed. **d)** While we have interpreted the network interactions between health variables, we have not tried to interpret the influence of variables on mortality.

While this thesis focused on interpretable models to understand aging, purely predictive models also have use in aging research. "Digital twins" are synthetic generated individuals from a machine learning model, used to expand the control arm of clinical trials. This enables trials to be conducted with smaller control arms and bigger treatment arms, leading to more efficient trials. While they have currently not been used explicitly for aging research, these approaches have been developed for Alzheimer's Disease [290, 291], a common disease of aging. Since the DJIN model can generate synthetic individuals, extensions of this model may have applications

in this area. In the long term when we are able to gather large amounts of data per individual (e.g. individual health tracking with wearable devices), "personalized medicine" or "precision medicine" to individualize treatment will involve predictive machine learning approaches.

Chapter 8 studied the effects of aging interventions on resilience and robustness in aging. This work only looked at summary measures of repair and damage, and highlighted the importance of studying both of these processes. A further understanding of the underlying microscopic mechanisms of repair and damage requires detailed biological data.

One of the current limitations to modelling aging is the lack of large longitudinal studies combining data from multiple levels of functioning, from low-level omics data to high-level physical functioning data [48]. This thesis used data on high-level physical functioning such as walking speed, grip strength, or other physical disabilities, phenotypic markers such as blood pressure and pulse, and blood biomarker measures such as blood glucose and hemoglobin. However, this thesis did not use any genetics, epigenetics or other omics data that have recently become important tools in aging biology. This is due to the availability of this data, and longitudinal studies with this data are just starting to emerge [60, 292–295].

This data could be used to develop models integrating multiple physiological systems to understand how observable aging phenotypes result from the underlying biological processes [54–56]. For example in Figure 9.2a), we can combine latent variable summary models with the DJIN interacting network model (Chapter 5). In this approach high-dimensional data for distinct physiological subsystems could be used to form summary measures for these systems, which then interact in a network as in the DJIN model. This approach could have the advantages of latent variable summary models, which compress high-dimensional information, and the advantages of network models, which describe interactions.

Alternatively, this data can be used to develop hierarchical models of aging. Figure 9.2b) shows a method of building such a model. A set of hierarchical latent variables describe the dynamics of aging at different levels, with high-level dynamics describing the high-level observed features, and low-level dynamics describing the low-level observed features. Unobserved aspects of aging can be incorporated by a

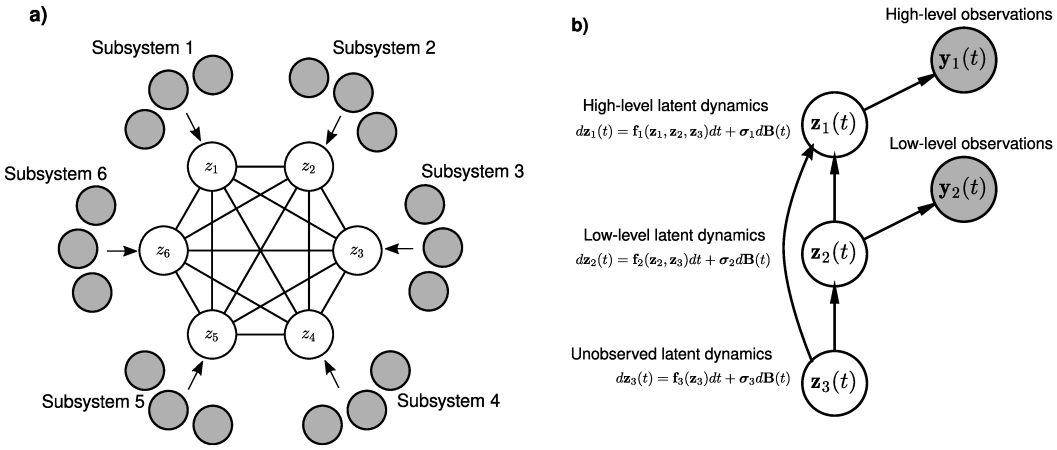


Figure 9.2: **Future latent variable models of aging. a)** A high dimensional network model. Omics data can contain millions of variables; building interaction networks between these could become intractable. However, we can combine the ideas of interaction network models and low-dimensional summary models to simplify the problem by representing individual physiological subsystems by low-dimensional summary measures of the omics variables, that interact in a network. Grey nodes represent observed high-dimensional omics health variables, and white nodes the inferred latent summary variables. **b)** Hierarchical latent dynamics models. Different levels of health can be modelled with an explicit hierarchy of dynamics. Latent variables representing each level of the hierarchy can be modelled, with high-level latent dynamics corresponding to the dynamics of observed high-level health and low-level latent dynamics corresponding to observed low-level health. Additionally, we can include the latent dynamics of unobserved health aspects, which influence both the high-level and low-level dynamics.

set of latent variables at the bottom of the hierarchy, influencing the other dynamics. Such models of hierarchical latent variable dynamics exist for other systems, such as in neuroscience [251, 296].

While the data is currently not available for many of these approaches, models can start to be developed with synthetic data, and then used once the data becomes available. I believe there is a bright future applying these techniques in aging research.

Bibliography

[1] Tamara Tchkonia, Yi Zhu, Jan van Deursen, Judith Campisi, and James L Kirkland. Cellular senescence and the senescent secretory phenotype: therapeutic opportunities. *J. Clin. Invest.*, 123(3):966–972, March 2013.

[2] Luigi Fontana, Brian K Kennedy, Valter D Longo, Douglas Seals, and Simon Melov. Medical research: treat ageing. *Nature*, 511(7510):405–407, July 2014.

[3] Susan E Howlett, Andrew D Rutenberg, and Kenneth Rockwood. The degree of frailty as a translational measure of health in aging. *Nature Aging*, 1(8):651–665, 2021.

[4] Carlos López-Otín, Maria A. Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. The hallmarks of aging. *Cell*, 153:1194–1217, 2013.

[5] Brian K. Kennedy, Shelley L. Berger, Anne Brunet, Judith Campisi, Ana Maria Cuervo, Elissa S. Epel, Claudio Franceschi, Gordon J. Lithgow, Richard I. Morimoto, Jeffrey E. Pessin, Thomas A. Rando, Arlan Richardson, Eric E. Schadt, Tony Wyss-Coray, and Felipe Sierra. Aging: a common driver of chronic diseases and a target for novel interventions. *Cell*, 159(4):709–713, 2014.

[6] Alan A Cohen. Complex systems dynamics in aging: new evidence, continuing questions. *Biogerontology*, 17(1):205–220, 2016.

[7] ICCS workshop on complex systems dynamics in aging biology. Tenth International Conference on Complex Systems, 2020.

[8] Linda P Fried, Alan A Cohen, Qian-Li Xue, Jeremy Walston, Karen Bandeen-Roche, and Ravi Varadhan. The physical frailty syndrome as a transition from homeostatic symphony to cacophony. *Nature Aging*, 1(1):36–46, 2021.

[9] Leonid A. Gavrilov and Natalia S. Gavrilova. The reliability theory of aging and longevity. *Journal of Theoretical Biology*, 213(4):527 – 545, 2001.

[10] Dervis C. Vural, Greg Morrison, and L. Mahadevan. Aging in complex interdependency networks. *Physical Review E*, 89(022811), 2014.

[11] Swadhin Taneja, Arnold B. Mitnitski, Kenneth Rockwood, and Andrew D. Rutenberg. Dynamical network model for age-related health deficits and mortality. *Physical Review E*, 93:022309, 2016.

[12] Nicholas Stroustrup, Winston E. Anthony, Zachary M. Nash, Vivek Gowda, Adam Gomez, Isaac F. López-Moyado, Javier Apfeld, and Walter Fontana. The temporal scaling of Caenorhabditis elegans aging. *Nature*, 530:103–107, 2016.

[13] D. Podolskiy, I. Molodtcov, A. Zenin, V. Kogan, L. I. Menshikov, Vadim N. Gladyshev, Robert J. Shmookler Reis, and P. O. Fedichev. Critical dynamics of gene networks is a mechanism behind ageing and Gompertz law, 2016.

[14] A B Mitnitski, A J Mogilner, and K Rockwood. Accumulation of deficits as a proxy measure of aging. *The Scientific World*, 1:323–36, 2001.

[15] Samuel D Searle, Arnold Mitnitski, Evelyne A Gahbauer, Thomas M Gill, and Kenneth Rockwood. A standard procedure for creating a frailty index. *BMC Geriatrics*, 8:24, 2008.

[16] Linda P. Fried, Catherine M. Tangen, Jeremy Walston, Anne B. Newman, Calvin Hirsch, John Gottdiener, Teresa Seeman, Russell Tracy, Willem J. Kop, Gregory Burke, and Mary Ann McBurnie. Frailty in older adults: Evidence for a phenotype. *The Journals of Gerontology: Series A*, 56(3):M146–M157, 2001.

[17] Petr Klemera and Stanilav Doubal. A new approach to the concept and computation of biological age. *Mechanisms of Ageing and Development*, 127(3):240–248, 2006.

[18] Steve Horvath. DNA methylation age of human tissues and cell types. *Genome Biology*, 14:R115, 2013.

[19] Emmanuel Milot, V Morissette-Thomas, Qing Li, Linda P Fried, Luigi Ferrucci, and Alan A Cohen. Trajectories of physiological dysregulation predicts mortality and health outcomes in a consistent manner across three populations. *Mechanisms of Ageing and Development*, 141-142:56–63, 2014.

[20] Konstantin G Arbeev, Svetlana V Ukraintseva, Olivia Bagley, Ilya Y Zhbannikov, Alan A Cohen, Alexander M Kulminski, and Anatoliy I Yashin. "physiological dysregulation" as a promising measure of robustness and resilience in studies of aging and a new indicator of preclinical disease. *J. Gerontol. A Biol. Sci. Med. Sci.*, 74(4):462–468, 2019.

[21] Konstantin Avchaciov, Marina P. Antoch, Ekaterina L. Andrianova, Andrei E. Tarkhov, Leonid I. Menshikov, Olga Burmistrova, Andrei V. Gudkov, and Peter O. Fedichev. Identification of a blood test-based biomarker of aging through deep learning of aging trajectories in large phenotypic datasets of mice. *bioRxiv*, 2020.

[22] C Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1:206–215, 2019.

[23] Spencer Farrell, Garrett Stubbings, Kenneth Rockwood, Arnold Mitnitski, and Andrew Rutenberg. The potential for complex computational models of aging. *Mechanisms of Ageing and Development*, 193:111403, 2021.

[24] Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv*, 2001.04385, 2020.

[25] Anuj Karpatne, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering*, 29(10):2318–2331, 2017.

[26] Rama K Vasudevan, Maxim Ziatdinov, Lukas Vlcek, and Sergei V Kalinin. Off-the-shelf deep learning is not enough, and requires parsimony, bayesianity, and causality. *npj Computational Materials*, 7(1):1–6, 2021.

[27] George E P Box. Science and statistics. *Journal of the American Statistical Association*, 71:791–799, 1976.

[28] Benjamin Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society*, 115:513–583, 1825.

[29] Thomas B L Kirkwood. Deciphering death: a commentary on Gompertz (1825) 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'. *Philosophical Transactions Of The Royal Society Of London Series B*, 370(1666):20140379–20140379, April 2015.

[30] Dervis C. Vural, Greg Morrison, and L. Mahadevan. Aging in complex interdependency networks. *Phys. Rev. E*, 89:022811, 2014.

[31] Eric D Sun, Thomas C T Michaels, and L Mahadevan. Optimal control of aging in complex networks. *Proceedings of the National Academy of Sciences*, 117(34):20404–20410, August 2020.

[32] Nicholas Stroustrup, Winston E Anthony, Zachary M Nash, Vivek Gowda, Adam Gomez, Isaac F López-Moyado, Javier Apfeld, and Walter Fontana. The temporal scaling of *Caenorhabditis elegans* ageing. *Nature*, 530(7588):103–107, 2016.

[33] Anatoli I. Yashin, Konstantin G. Arbeev, Igor Akushevich, Aliaksandr Kulminski, Lucy Akushevich, and Svetlana V. Ukraintseva. Stochastic model for analysis of longitudinal data on aging and mortality. *Mathematical Biosciences*, 208:538–551, 2007.

[34] A. I. Yashin, K. G. Arbeev, I. Akushevich, A. Kulminski, S. V. Ukraintseva, E. Stallard, and K. C. Land. The quadratic hazard model for analyzing longitudinal data on aging, health, and the life span. *Physics of Life Reviews*, 9:177–188, 2012.

[35] Konstantin G. Arbeev, Igor Akushevich, Alexander M. Kulminski, Svetlana V. Ukraintseva, and Anatoliy I. Yashin. Joint analyses of longitudinal and time-to-event data in research on aging: Implications for predicting health and survival. *Frontiers in Public Health*, 2(228), 2014.

[36] Ilya Y. Zhbannikov, Konstantin Arbeev, Igor Akushevich, Eric Stallard, and Anatoliy I. Yashin. stpm: an R package for stochastic process model. *BMC Bioinformatics*, 18(125), 2017.

[37] E. Pierson, P. W. Koh, T. Hashimoto, D. Koller, and Percy Liang. Inferring multidimensional rates of aging from cross-sectional data. *Proc Mach Learn Res*, 89:97–107, 2019.

[38] Arnold Mitnitski, Le Bao, and Kenneth Rockwood. Going from bad to worse: A stochastic model of transitions in deficit accumulation, in relation to mortality. *Mech. Ageing Dev.*, 127(5):490–493, May 2006.

[39] Arnold Mitnitski, Xiaowei Song, and Kenneth Rockwood. Improvement and decline in health status from late middle age: modeling age-related changes in deficit accumulation. *Exp. Gerontol.*, 42(11):1109–1115, November 2007.

[40] Arnold B Mitnitski, Nader Fallah, Charmaine B Dean, and Kenneth Rockwood. A multi-state model for the analysis of changes in cognitive scores over a fixed time interval. *Stat. Methods Med. Res.*, 23(3):244–256, June 2014.

[41] A Mitnitski, X Song, and K Rockwood. Trajectories of changes over twelve years in the health status of canadians from late middle age. *Exp. Gerontol.*, 47(12):893–899, December 2012.

[42] Natasha Oswal, Olivier M F Martin, Sofia Stroustrup, Monika Anna Matusiak Bruckner, and Nicholas Stroustrup. A hierarchical process model links behavioral aging and lifespan in c. elegans. April 2021.

[43] Omer Karin, Amit Agrawal, Ziv Porat, Valery Krizhanovsky, and Uri Alon. Senescent cell turnover slows with age providing an explanation for the Gompertz law. *Nature Communications*, 10(1):1–9, December 2019.

[44] S. G. Farrell, Arnold B. Mitnitski, Kenneth Rockwood, and Andrew D. Rutenberg. Network model of human aging: frailty limits and information measures. *Physical Review E*, 94:052409, 2016.

[45] S. G. Farrell, Arnold B. Mitnitski, Kenneth Rockwood, and Andrew D. Rutenberg. Probing the network structure of health deficits in human aging. *Physical Review E*, 98:032302, 2018.

[46] A. B. Mitnitski, A. D. Rutenberg, S. Farrell, and K. Rockwood. Aging, frailty and complex networks. *Biogerontology*, 18:433–446, 2017.

[47] Andrew D. Rutenberg, Arnold B. Mitnitski, Spencer G. Farrell, and Kenneth Rockwood. Unifying aging and frailty through complex dynamical networks. *Experimental Gerontology*, 107:126–129, 2018.

[48] Spencer Farrell, Arnold Mitnitski, Kenneth Rockwood, and Andrew Rutenberg. Generating synthetic aging trajectories with a weighted network model using cross-sectional data. *Scientific Reports*, page 19833, 2020.

[49] Garrett Stubbings, Spencer Farrell, Arnold Mitnitski, Kenneth Rockwood, and Andrew Rutenberg. Informative frailty indices from binarized biomarkers. *Biogerontology*, 70(1):1–11, 2020.

[50] Spencer G Farrell and Andrew D Rutenberg. Anomalously slow transport in single-file diffusion with slow binding kinetics. *Phys. Rev. E*, 98(2):022114, 2018.

[51] Spencer Farrell and Andrew D Rutenberg. Non-Fickian single-file pore transport. *Phys. Rev. E*, 104(3):L032102, 2021.

[52] Thomas B. L. Kirkwood. Understanding the odd science of aging. *Cell*, 120:437 – 447, 2005.

[53] Thomas B. L. Kirkwood and Caleb E. Finch. The old worm turns more slowly. *Nature*, 419:794–795, 2002.

[54] Luigi Ferrucci, Morgan E. Levine, Pei-Lun Kuo, and Eleanor M. Simonsick. Time and the metrics of aging. *Circulation Research*, 123:740–744, 2018.

[55] Arnold Mitnitski and Kenneth Rockwood. *The problem of integrating biological and clinical markers of aging*. Springer, 2019.

[56] P-L Kuo, J A Schrack, M D Shardell, M Levine, A Z Moore, Y An, P Elango, A Karikkineth, T Tanaka, R de Cabo, L M Zukley, M AlGhatrif, C W Chia, E M Simonsick, J M Egan, S M Resnick, and L Ferrucci. A roadmap to build a phenotypic metric of ageing: insights from the Baltimore Longitudinal Study of Aging. *J. Intern. Med.*, 287(4):373–394, 2020.

[57] Alice E. Kane and David A. Sinclair. Frailty biomarkers in humans and rodents: Current approaches and future advances. *Mechanisms of Ageing and Development*, 180:117–128, 2019.

[58] Gregory Livshits, Ida Malkin, Ruth C E Bowyer, Serena Verdi, Jordana T Bell, Cristina Menni, Frances M K Williams, and Claire J Steves. Multi-omics analyses of frailty and chronic widespread musculoskeletal pain suggest involvement of shared neurological pathways. *Pain*, 159(12):2565–2572, 2018.

[59] Benoit Lehallier, David Gate, Nicholas Schaum, Tibor Nanasi, Song Eun Lee, Hanadie Yousef, Patricia Moran Losada, Daniela Berdnik, Andreas Keller, Joe Verghese, Sanish Sathyan, Claudio Franceschi, Sofiya Milman, Nir Barzilai, and Tony Wyss-Coray. Undulating changes in human plasma proteome profiles across the lifespan. *Nature Medicine*, 25:1843–1850, 2019.

[60] Sara Ahadi, Wenyu Zhou, Sophia Miryam Schüssler-Fiorenza Rose, M. Reza Sailani, Kévin Contrepois, Monika Avina, Melanie Ashland, Anne Brunet, and Michael Snyder. Personal aging markers and ageotypes revealed by deep longitudinal profiling. *Nature Medicine*, 26:83–90, 2020.

[61] Luigi Ferrucci, Marta Gonzalez-Freire, Elisa Fabbri, Eleanor Simonsick, Toshiko Tanaka, Zenobia Moore, Shabnam Salimi, Felipe Sierra, and Rafael de Cabo. Measuring biological aging in humans: A quest. *Aging Cell*, 19:e13080, 2020.

[62] A B Mitnitski, A J Mogilner, and K Rockwood. Accumulation of deficits as a proxy measure of aging. *The Scientific World*, 1:323–36, 2001.

[63] K Rockwood, R A Fox, P Stolee, D Robertson, and B L Beattie. Frailty in elderly people: an evolving concept. *CMAJ*, 150(4):489–495, 1994.

[64] D. Gu, M. E. Dupre, J. Sautter, Haiyan Zhu, Yuzhi Liu, and Zeng Yi. Frailty and mortality among chinese at advanced ages. *Journal of Gerontology: Social Sci*, 64B(2):279–289, 2009.

[65] K. Rockwood, A. Mogilner, and A. Mitnitski. Changes with age in the distribution of a frailty index. *Mechanisms of Ageing and Development*, 125:517, 2004.

[66] Arnold Mitnitski, Xiaowei Song, and Kenneth Rockwood. Assessing biological aging: the origin of deficit accumulation. *Biogerontology*, 14(6):709–17, 2013.

[67] Ruth E Hubbard, Nancye M Peel, Mayukh Samanta, Leonard C Gray, Brant E Fries, Arnold Mitnitski, and Kenneth Rockwood. Derivation of a frailty index from the interRAI acute care instrument. *BMC Geriatrics*, 15:27, 2015.

[68] Stephanie Bennett, Xiaowei Song, Arnold Mitnitski, and Kenneth Rockwood. A limit to frailty in very old, community-dwelling people: a secondary analysis of the Chinese longitudinal health and longevity study. *Age and Ageing*, 42(3):372–377, 2013.

[69] Alexander M. Kulminski, Svetlana V. Ukraintseva, Igor V. Akushevich, Konstantin G. Arbeev, and Anatoli I. Yashin. Cumulative index of health deficiencies as a characteristic of long life. *Journal of the American Geriatrics Society*, 55:935–940, 2007.

[70] Susan E Howlett, Michael R H Rockwood, Arnold Mitnitski, and Kenneth Rockwood. Standard laboratory tests to identify older adults at increased risk of death. *BMC Medicine*, 12:171, 2014.

[71] Arnold Mitnitski, Joanna Collerton, Carmen Martin-Ruiz, Carol Jagger, Thomas von Zglinicki, Kenneth Rockwood, and Thomas B. L. Kirkwood. Age-related frailty and its association with biological markers of aging. *BMC Medicine*, 13:161, 2015.

[72] Joanna M. Blodgett, Olga Theou, Susan E. Howlett, and Kenneth Rockwood. A frailty index from common clinical and laboratory tests predicts increased risk of death across the life course. *GeroScience*, 39:447, 2017.

[73] Jocelyne C Whitehead, Barbara A Hildebrand, Michael Sun, Michael R Rockwood, Robert A Rose, Kenneth Rockwood, and Susan E Howlett. A clinical frailty index in aging mice: comparisons with frailty index data in humans. *J. Gerontol. A Biol. Sci. Med. Sci.*, 69(6):621–632, 2014.

[74] K Rockwood, X Song, C MacKnight, H Bergman, D B Hogan, I McDowell, and A Mitnitski. A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal*, 173(5):489–495, 2005.

[75] Gregory Hannum, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, SriniVas Sadda, Brandy Klotzle, Marina Bibikova, Jian-Bing Fan, Yuan Gao, Rob Deconde, Menzies Chen, Indika Rajapakse, Stephen Friend, Trey Ideker, and Kang Zhang. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 49(2):359–367, 2013.

[76] Morgan E. Levine, Ake T. Lu, Austin Quach, Brian H. Chen, Themistocles L. Assimes, Stefania Bandinelli, Lifang Hou, Andrea A. Baccarelli, James D. Stewart, Yun Li, Eric A. Whitsel, James G Wilson, Alex P Reiner, Abraham Aviv, Kurt Lohman, Yongmei Liu, Luigi Ferrucci, and Steve Horvath. An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, 10(4):573–591, 2018.

[77] Ake T. Lu, Austin Quach, James G. Wilson, Alex P. Reiner, Abraham Aviv, Kenneth Raj, Lifang Hou, Andrea A. Baccarelli, Yun Li, James D. Stewart, Eric A. Whitsel, Themistocles L. Assimes, Luigi Ferrucci, and Steve Horvath. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging*, 11(2):303–327, 2019.

[78] Daniel W Belsky, Terrie E Moffitt, Alan A Cohen, David L Corcoran, Morgan E Levine, Joseph A Prinz, Jonathan Schaefer, Karen Sugden, Benjamin Williams, Richie Poulton, and Avshalom Caspi. Eleven telomere, epigenetic clock, and Biomarker-Composite quantifications of biological aging: Do they measure the same thing? *Am. J. Epidemiol.*, 187(6):1220–1230, 2018.

[79] Sarah J Mitchell, Morten Scheibye-Knudsen, Dan L Longo, and Rafael de Cabo. Animal models of aging research: implications for human aging and age-related diseases. *Annu Rev Anim Biosci*, 3:283–303, 2015.

[80] Heehwa G Son, Ozlem Altintas, Eun Ji E Kim, Sujeong Kwon, and Seung-Jae V Lee. Age-dependent changes and biomarkers of aging in caenorhabditis elegans. *Aging Cell*, 18(2):e12853, 2019.

[81] Nicholas A Swierczek, Andrew C Giles, Catharine H Rankin, and Rex A Kerr. High-throughput behavioral analysis in *C. elegans*. *Nature Methods*, 8:592–598, 2011.

[82] William B. Zhang, Drew B. Sinha, William E. Pittman, Erik Hvatum, Nicholas Stroustrup, and Zachary Pincus. A C. Elegans mutant that lives twice as long as wild type. *Cell Systems*, 3:333–345, 2016.

[83] Céline N Martineau, André E X Brown, and Patrick Laurent. Multidimensional phenotyping predicts lifespan and quantifies health in Caenorhabditis elegans. *PLoS Comput. Biol.*, 16(7):e1008002, 2020.

[84] Titus Kaletta and Michael O Hengartner. Finding function in novel targets: C. elegans as a model organism. *Nat. Rev. Drug Discov.*, 5(5):387–399, 2006.

[85] Valerie Vanhooren and Claude Libert. The mouse as a model organism in aging research: usefulness, pitfalls and possibilities. *Ageing Res. Rev.*, 12(1):8–21, 2013.

[86] Alice E Kane, Sarah N Hilmer, John Mach, Sarah J Mitchell, Rafael de Cabo, and Susan E Howlett. Animal models of frailty: current applications in clinical research. *Clin. Interv. Aging*, 11:1519–1529, 2016.

[87] Kaitlyn Keller, Alice Kane, Stefan Heinze-Milne, Scott A Grandy, and Susan E Howlett. Chronic treatment with the ACE inhibitor enalapril attenuates the development of frailty and differentially modifies pro- and anti-inflammatory cytokines in aging male and female C57BL/6 mice. *J. Gerontol. A Biol. Sci. Med. Sci.*, 74(8):1149–1157, 2019.

[88] Elise S. Bisset. Voluntary aerobic exercise attenuates frailty in a sex-specific manner in older male and female C57Bl/6 mice. Master's thesis, Dalhousie University, 2021.

[89] Dario Riccardo Valenzano, Aziz Aboobaker, Andrei Seluanov, and Vera Gorbunova. Non-canonical aging model systems and why we need them. *EMBO J.*, 36(8):959–963, April 2017.

[90] James L Kirkland, Michael B Stout, and Felipe Sierra. Resilience in aging mice. *J. Gerontol. A Biol. Sci. Med. Sci.*, 71(11):1407–1414, 2016.

[91] Marten Scheffer, J Elizabeth Bolhuis, Denny Borsboom, Timothy G Buchman, Sanne M W Gijzel, Dave Goulson, Jan E Kammenga, Bas Kemp, Ingrid A van de Leemput, Simon Levin, Carmel Mary Martin, René J F Melis, Egbert H van Nes, L Michael Romero, and Marcel G M Olde Rikkert. Quantifying resilience of humans and other animals. *Proc. Natl. Acad. Sci. U. S. A.*, 115(47):11883–11890, 2018.

[92] Markus Schosserer, Gareth Banks, Soner Dogan, Peter Dungel, Adelaide Fernandes, Darja Marolt Presen, Ander Matheu, Marcin Osuchowski, Paul Potter, Coral Sanfeliu, Bilge Guvenc Tuna, Isabel Varela-Nieto, and Ilaria Bellantuono. Modelling physical resilience in ageing mice. *Mech. Ageing Dev.*, 177:91–102, 2019.

[93] Sanne M W Gijzel, Heather E Whitson, Ingrid A van de Leemput, Marten Scheffer, Dieneke van Asselt, Jerrald L Rector, Marcel G M Olde Rikkert, and René J F Melis. Resilience in clinical care: Getting a grip on the recovery potential of older adults. *J. Am. Geriatr. Soc.*, 67(12):2650–2657, 2019.

[94] Svetlana Ukraintseva, Konstantin Arbeev, Matt Duan, Igor Akushevich, Alexander Kulminski, Eric Stallard, and Anatoliy Yashin. Decline in biological resilience as key manifestation of aging: Potential mechanisms and role in health and longevity. *Mech. Ageing Dev.*, 194:111418, 2021.

[95] Cathleen Colón-Emeric, Carl F Pieper, Kenneth E Schmader, Richard Sloane, Allison Bloom, Micah McClain, Jay Magaziner, Kim M Huffman, Denise Orwig, Donna M Crabtree, and Heather E Whitson. Two approaches to classifying and quantifying physical resilience in longitudinal data. *J. Gerontol. A Biol. Sci. Med. Sci.*, 75(4):731–738, 2020.

[96] Timothy V Pyrkov, Konstantin Avchaciov, Andrei E Tarkhov, Leonid I Menshikov, Andrei V Gudkov, and Peter O Fedichev. Longitudinal analysis of blood markers reveals progressive loss of resilience and predicts human lifespan limit. *Nat. Commun.*, 12(1):1–10, 2021.

[97] Daniel C Parker, Cathleen Coln-Emeric, Janet L Huebner, Ching-Heng Chou, Virginia Byers Kraus, Carl F Pieper, Richard Sloane, Heather E Whitson, Denise Orwig, Donna M Crabtree, Jay Magaziner, James R Bain, Michael Muehlbauer, Olga R Ilkayeva, and Kim M Huffman. Biomarkers associated with physical resilience after hip fracture. *J. Gerontol. A Biol. Sci. Med. Sci.*, 75(10):e166–e172, 2020.

[98] Albert-Laszlo Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[99] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Physical Review E*, 63:066123, 2001.

[100] Babak Fotouhi and Michael G. Rabbat. Degree correlation in scale-free graphs. *The European Physical Journal B*, 86:510, 2013.

[101] Alain Barrat and Romualdo Pastor-Satorras. Rate equation approach for correlations in growing network models. *Physical Review E*, 71:036127, 2005.

[102] R. Xulvi-Brunet and I. M. Sokolov. Reshuffling scale-free networks: From random to assortative. *Physical Review E*, 70:066102, 2004.

[103] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 69(6 Pt 2):066138, 2004.

[104] O. Vasicek. A test for normality based on sample entropy. *Journal of the Royal Statistical Society. Series B*, 38(1):54–59, 1976.

[105] E. J. Dudewicz and E. C. van der Meulen. Entropy-based tests of uniformity. *Journal of the American Statistical Association*, 76(376):967–974, 1981.

[106] B. van Es. Estimating functionals related to a density by a class of statistics based on spacing. *Scandinavian Journal of Statistics*, 19(1):61–72, 1992.

[107] E. G. Learned-Miller and J. W. Fisher-III. Ica using spacing estimates of entropy. *Journal of Machine Learning Research*, 4:1271–1295, 2003.

[108] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[109] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Netw.*, 2(5):359–366, January 1989.

[110] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[111] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations*, 2015.

[112] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar, 2014.

[113] Stan Development Team. The Stan core library version 2.25.0, 2020.

[114] Matthew D Hoffman and Andrew Gelman. The No-U-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, January 2014.

[115] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G R Day, Clint Richardson, Charles K Fisher, and David J Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Phys. Rep.*, 810:1–124, 2019.

[116] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *Proceedings of the International Conference on Learning Representations*, 2014.

[117] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on International Conference on Machine Learning*, 2014.

[118] Yarin Gal. *Uncertainty in deep learning*. PhD thesis, University of Cambridge, 2016.

[119] D R Cox. Regression models and life-tables. *Journal Of The Royal Statistical Society Series B-Methodological*, 34(2):187–220, 1972.

[120] Norman E Breslow. Discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society: B*, 34:216–217, 1972.

[121] Samuel L. Brilleman, Eren M. Elci, Jacqueline Buros Novik, and Rory Wolfe. Bayesian survival analysis using the rstanarm r package. *arXiv*, 2002.09633, 2020.

[122] Thomas B. L. Kirkwood. Understanding the odd science of aging. *Cell*, 120:437 – 447, 2005.

[123] Jan Vijg and Judith Campisi. Puzzles, promises and a cure for ageing. *Nature*, 454(7208):1065–1071, 2008.

[124] David Gems and Linda Partridge. Genetics of longevity in model organisms: debates and paradigm shifts. *Annu. Rev. Physiol.*, 75:621–644, 2013.

[125] Spencer G Farrell, Arnold B Mitnitski, Kenneth Rockwood, and Andrew D Rutenberg. Network model of human aging: Frailty limits and information measures. *Physical Review E*, 94(5):052409, 2016.

[126] Susan E Howlett, Michael Rockwood, Arnold Mitnitski, and Kenneth Rockwood. Standard laboratory tests to identify older adults at increased risk of death. *BMC Medicine*, 12(1), 2014.

[127] Joanna M. Blodgett, Olga Theou, Susan E. Howlett, and Kenneth Rockwood. A frailty index from common clinical and laboratory tests predicts increased risk of death across the life course. *GeroScience*, 39(4):447–455, 2017.

[128] Spencer G Farrell, Arnold B Mitnitski, Olga Theou, Kenneth Rockwood, and Andrew D Rutenberg. Probing the network structure of health deficits in human aging. *Physical Review E*, 98(3):187, 2018.

[129] Laura A. Herndon, Peter J. Schmeissner, Justyna M. Dudaronek, Paula A. Brown, Kristin M. Listner, Yuko Sakano, Marie C. Paupard, David H. Hall, and Monica Driscoll. Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. *Nature*, 419:808 – 814, 2002.

[130] K. Rockwood, J. M. Blodgett, O. Theou, M. H. Sun, H. A. Feridooni, A. Mitnitski, R. A. Rose, J. Godin, E. Gregson, and S. E. Howlett. A frailty index based on deficit accumulation quantifies mortality risk in humans and in mice. *Scientific Reports*, 7:43068, 2017.

[131] A. E. Kane, O. Ayaz, A. Ghimire, H. A. Feridooni, and S. E. Howlett. Implementation of the mouse frailty index. *Can. J. Physiol. Pharmacol.*, 95:1149, 2017.

[132] Stephen J. Evans, Margaret Sayers, Arnold Mitnitski, and Kenneth Rockwood. The risk of adverse outcomes in hospitalized older patients in relation to a frailty index based on a comprehensive geriatric assessment. *Age and Ageing*, 43:127–132, 2014.

[133] Arnold Mitnitski, Susan E Howlett, and Kenneth Rockwood. Heterogeneity of human aging and its assessment. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 72:877, May 2016.

[134] Gotaro Kojima, Steve Iliffe, and Kate Walters. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age and Ageing*, 47(2):193–200, 2018.

[135] Thomas B. L. Kirkwood. Deciphering death: a commentary on Gompertz (1825) 'on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'. *Philosophical Transactions of the Royal Society B*, 370:20140379, 2015.

[136] Albert-Laszlo Barabási. Scale-free networks: A decade and beyond. *Science*, 325:412–413, 2009.

[137] Henrik Hedegaard Klausen, Janne Petersen, Thomas Bandholm, Helle Gybel Juul-Larsen, Juliette Tavenier, Jesper Eugen-Olsen, and Ove Andersen. Association between routine laboratory tests and long-term mortality among acutely admitted older medical patients: a cohort study. *BMC Geriatrics*, 17:62, 2017.

[138] Katherine E. King, Gerda G. Fillenbaum, and Harvey J. Cohen. A cumulative deficit laboratory test-based frailty index: Personal and neighborhood associations. *Journal of the American Geriatric Society*, 65:1981, 2017.

[139] Richard A. McPherson and Matthew R. Pincus. *Henry's: Clinical diagnosis and management by laboratory methods*. Elsevier, 23 edition, 2017.

[140] Joanna M. Blodgett, Olga Theou, Susan E. Howlett, Frederick C. W. Wu, and Kenneth Rockwood. A frailty index based on laboratory deficits in community-dwelling men predicted their risk of adverse outcomes. *Age and Ageing*, 45:463 – 468, 2016.

[141] Kenneth Rockwood, Miranda McMillan, Arnold Mitnitski, and Susan E Howlett. A frailty index based on common laboratory tests in comparison with a clinical frailty index for older adults in long-term care facilities. *JAMDA*, 16:842 – 847, 2015.

[142] Albert-Laszlo Barabási. *Network Science*. Cambridge University Press, 2016.

[143] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89:208701, 2002.

[144] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81:25, 1977.

[145] Michael A. Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A*, 104:1876 – 1889, 2000.

[146] Andrew Clegg, Chris Bates, John Young, Ronan Ryan, Linda Nichols, Elizabeth Ann Teale, Mohammed A. Mohammed, John Parry, and Tom Marshall. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age and Ageing*, 45:353 – 360, 2016.

[147] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory, Second Edition*. Wiley, New Jersey, 2006.

[148] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379, 1948.

[149] David Blokh and Ilia Stambler. The application of information theory for the research of aging and aging-related diseases. *Progress in Neurobiology*, 157:158–173, 2017.

[150] Spotswood L. Spruance, Julia E. Reid, Michael Grace, and Matthew Samore. Hazard ratio in clinical trials. *Antimicrobial Agents and Chemotherapy*, 48:2787, 2004.

[151] A. J. Butte and I. S. Kohane. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pacific Symposium on Biocomputing*, 5:415–426, 2000.

[152] Adam A. Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7:S7, 2006.

[153] Arnold B. Mitnitski, Janice E. Graham, Alexander J. Mogilner, and Kenneth Rockwood. Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC Geriatrics*, 2:1, 2002.

[154] Bin Zhang and Steve Horvath. A general framework for weighted gene coexpression network analysis. *Statistical Applications in Genetics and Molecular Biology*, 4:1, 2005.

[155] A. Barrat, M. Bathélemy, R. Pastor-Satorras, and A. Vespignani. The architechture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752, 2004.

[156] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87:925, 2015.

[157] James P. Gleeson. High-accuracy approximation of binary-state dynamics on networks. *Physical Review Letters*, 107:068701, 2011.

[158] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Advances in Physics*, 51:1079, 2002.

[159] Centers for Disease Control and Prevention National Center for Health Statistics. National health and nutrition examination survey data, Updated 2014.

[160] Canadian Study of Health and Aging Working Group. Canadian study of health and aging: study methods and prevalence of dementia. *Canadian Medical Association Journal*, 150(6):899, 1994.

[161] David Jones, Xiaowei Song, Arnold Mitnitski, and Kenneth Rockwood. Evaluation of a frailty index based on a comprehensive geriatric assessment in a population based study of elderly Canadians. *Aging Clinical and Experimental Research*, 17:465, 2005.

[162] Oliver L. Hatheway, Arnold Mitnitski, and Kenneth Rockwood. Frailty affects the initial treatment response and time to recovery of mobility in acutely ill older adults admitted to hospital. *Age and Ageing*, 46:920, 2017.

[163] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[164] E. Arias. United States life tables, 2010. *National Vital Statistics Reports*, 63(7), 2014.

[165] A. Barrat and M. Weigt. On the properties of small-world networks. *European Physics Journal B*, 13:547–560, 2000.

[166] Yamir Moreno, Javier B. Gómez, and Amalio F. Pacheco. Epidemic incidence in correlated complex networks. *Physical Review E*, 68:035103(R), 2003.

[167] Jeffrey D. Allen, Yang Xie, Min Chen, Luc Girard, and Guanghua Xiao. Comparing statistical methods for constructing large scale gene networks. *PLOS ONE*, 7:e29348, 2012.

[168] Garrett Stubbings, Spencer Farrell, Arnold Mitnitski, Kenneth Rockwood, and Andrew Rutenberg. Informative frailty indices from binarized biomarkers. *Biogerontology*, 70(1):1–11, 2020.

[169] Andrew Clegg, John Young, Steve Iliffe, Marcel Olde Rikkert, and Kenneth Rockwood. Frailty in elderly people. *The Lancet*, 381(9868):752–762, 2013.

[170] Xia Li, Alexander Ploner, Yunzhang Wang, Patrik KE Magnusson, Chandra Reynolds, Deborah Finkel, Nancy L Pedersen, Juulia Jylhävä, and Sara Hägg. Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *eLife*, 9(e51507), 2020.

[171] Arnold B. Mitnitski, Alexander J. Mogilner, Chris MacKnight, and Kenneth Rockwood. The mortality rate as a function of accumulated deficits in a frailty index. *Mechanisms of Ageing and Development*, 123:1457–1460, 2002.

[172] Kenneth Rockwood and Arnold Mitnitski. Frailty in relation to the accumulation of deficits. *Journal of Gerontology: Medical Sciences*, 62A(7):722–727, 2007.

[173] E. Pierson, P. W. Koh, T. Hashimoto, D. Koller, and Percy Liang. Inferring multidimensional rates of aging from cross-sectional data. *Proc Mach Learn Res*, 89:97–107, 2019.

[174] Joshua J Armstrong, Arnold Mitnitski, Lenore J Launer, Lon R White, and Kenneth Rockwood. Frailty in the Honolulu-Asia Aging Study: deficit accumulation in a male cohort followed to 90% mortality. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 70(1):125–131, 2015.

[175] Edward L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

[176] Frank E. Harrell Jr, Robert M. Califf, David B. Pryor, Kerry L. Lee, and Robert A. Rosati. Evaluating the yield of medical tests. *The Journal of the American Medical Association*, 247(18):2543–2546, 1982.

[177] Laura Antolini, Patrizia Boracchi, and Elia Biganzoli. A time-dependent discrimination index for survival data. *Statistics in Medicine*, 24:3927 – 3944, 2005.

[178] Erika Graf, Claudia Schmoor, Will Sauerbrei, and Martin Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18:2529–2545, 1999.

[179] Baohua Zhou, David Hofmann, Itai Pinkoviezky, Samuel J. Sober, and Ilya Nemenman. Change, long tails, and inference in a non-gaussian, bayesian theory of vocal learning in songbirds. *Proceedings of the National Academy of Sciences*, 115(38):E8358, 2018.

[180] Daniel Stow, Fiona E. Matthews, and Barbara Hanratty. Frailty trajectories to identify end of life: a longitudinal population study. *BMC Medicine*, 16(171), 2018.

[181] Jiska Cohen-Mansfield, Michal Skornick-Bouchbinder, and Shai Brill. Trajectories of end of life: A systematic review. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, 73(4):564–572, 2018.

[182] Martine T. E. Puts, Paul Lips, and Dorly J. H. Deeg. Sex differences in the risk of frailty for mortality independent of disability and chronic diseases. *Journal of the American Geriatrics Society*, 53:40–47, 2005.

[183] Arnold Mitnitski, Xiaowei Song, Ingmar Skoog, G. A. Broe, Jafna L. Cox, Eva Grunfeld, and Kenneth Rockwood. Relative fitness and frailty of elderly men and women in developed countries and their relationship with mortality. *Journal of the American Geriatrics Society*, 53:2184–2189, 2005.

[184] E. H. Gordon, N. M. Peel, O. Theou, S. E. Howlett, and R. E. Hubbard. Sex differences in frailty: A systematic review and meta-analysis. *Experimental Gerontology*, 89:30–40, 2017.

[185] Melissa K. Andrew, Arnold B. Mitnitski, and Kenneth Rockwood. Social vulnerability, frailty and mortality in elderly people. *PLOS One*, 3(5):e2232, 2008.

[186] Linda P. Fried. Interventions for human frailty: Physical activity as a model. *Cold Spring Harbor Perspectives in Medicine*, 6:a025916, 2016.

[187] Nina T. Rogers, Alan Marshall, Chrissy H. Roberts, Panayotes Demakakos, Andrew Steptoe, and Shaun Scholes. Physical activity and trajectories of frailty among older adults: Evidence from the english longitudinal study of ageing. *PLOS One*, 12(2):e0170878, 2017.

[188] Zhesi Shen, Wen-Xu Wang, Ying Fan, Zengru Di, and Ying-Cheng Lai. Reconstructing propagation networks with natural diversity and identifying hidden sources. *Nature Communications*, 5:4323, 2014.

[189] Joseph L. Natale, David Hofmann, Hernandez Damian G, and Ilya Nemenman. Reverse-engineering biological networks from large data sets. *Quantitative Biology Theory, Computational Methods, and Models: Chapter 10*, 2017.

[190] Arnold B. Mitnitski, Janice E. Graham, Alexander J. Mogilner, and Kenneth Rockwood. Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC Geriatrics*, 2:1, 2002.

[191] Carmen García-Peña, Ricardo Ramírez-Aldana, Lorena Parra-Rodriguez, Juan Carlos Gómez-Verján, Mario Ulises Pérez-Zepeda, and Luis Miguel Gutiérrez-Robledo. Network analysis of frailty and aging: Empirical data from the Mexican Health and Aging Study. *Experimental Gerontology*, 128:110747, December 2019.

[192] Ryan N. Gutenkunst, Joshua J. Waterfall, Fergal P. Casey, Kevin S. Brown, Christopher R. Myers, and James P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3:10, 2007.

[193] Bryan Lim and Mihaela van der Schaar. Disease-atlas: Navigating disease trajectories using deep learning. *Proceeding of Machine Learning Research*, 85:137–160, 2018.

[194] Peter Schulam and Saria Suchi. A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure. *Advances in Neural Information Processing Systems 28*, 2015.

[195] Ahmed M. Alaa and Mihaela van der Schaar. Forecasting individualized disease trajectories using interpretable deep learning. 2018.

[196] Charles K Fisher, Aaron M Smith, and Jonathan R Walsh. Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Scientific Reports*, 9(1):1–14, September 2019.

[197] Leo Breiman. Hinging hyperplanes for regression, classification, and function approximation. *Biogerontology*, 39(3):999–1013, 1993.

[198] James Kennedy and Russell Eberheart. Particle swarm optimization. *Proceedings of IEEE International Conference on Neural Networks*, IV(1942):1942–1948, 1995.

[199] Peter J Diggle and Richard J Gratton. Monte carlo methods of inference for implicit statistical models. *J. R. Stat. Soc. Series B Stat. Methodol.*, 46(2):193–227, 1984.

[200] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proc. Natl. Acad. Sci. U. S. A.*, 117(48):30055–30062, 2020.

[201] S. Clemens, A. Phelps, Z. Oldfield, M. Blake, A. Oskala, M. Marmot, N. Rogers, J. Banks, A. Steptoe, and J. Nazroo. English longitudinal study of ageing: Waves 0-8 1998-2017. *UK Data Service. 30th Edition*, 5050, 2019.

[202] Morgan E. Levine. Modeling the rate of senescence: Can estimated biological age predict mortality more accurately than chronological age? *The Journals of Gerontology: Series A*, 68(6):667–674, 12 2012.

[203] Konstantin Avchaciov, Marina P. Antoch, Ekaterina L. Andrianova, Andrei E. Tarkhov, Leonid I. Menshikov, Olga Burmistrova, Andrei V. Gudkov, and Peter O. Fedichev. Identification of a blood test-based biomarker of aging through deep learning of aging trajectories in large phenotypic datasets of mice. *bioRxiv*, 2020.

[204] Yu-Ying Liu, Shuang Li, Fuxin Li, Le Song, and James M Rehg. Efficient learning of continuous-time hidden markov models for disease progression. *Advances in Neural Information Processing Systems*, pages 3600–3608, 2015.

[205] Jonathon R. Walsh, Aaron M. Smith, Yannick Pouliot, David Li-Bland, Anton Loukianov, and Charles K. Fisher. Generating digital twins with multiple sclerosis using probabilistic neural networks. 2020.

[206] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, May 2019.

[207] Yeping Lina Qiu, Hong Zheng, and Olivier Gevaert. Genomic data imputation with variational auto-encoders. *GigaScience*, 9(8), 08 2020.

[208] Yu Gong, Hossein Hajimirsadeghi, Jiawei He, Megha Nawhal, Thibaut Durand, and Greg Mori. Variational selective autoencoder. In Cheng Zhang, Francisco Ruiz, Thang Bui, Adji Bousso Dieng, and Dawen Liang, editors, *Proceedings of The 2nd Symposium on Advances in Approximate Bayesian Inference*, volume 118 of *Proceedings of Machine Learning Research*, pages 1–17. PMLR, 08 Dec 2020.

[209] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1530–1538, Lille, France, 07–09 Jul 2015. PMLR.

[210] A. I. Yashin and Kenneth G. Manton. Effects of unobserved and partially observed covariate processes on system failure: A review of models and estimation strategies. *Statistical Science*, 12:20–34, 1997.

[211] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

[212] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, Articles*, 45(3):1–67, 2011.

[213] Daniel J. Stekhoven and Peter Bühlmann. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118, 10 2011.

[214] Humza Haider, Bret Hoehn, Sarah Davis, and Russel Greiner. Effective ways to build and evaluate individual survival distributions. *Journal of Machine Learning Research*, 21:1–63, 2020.

[215] Steve Horvath. DNA methylation age of human tissues and cell types. *Genome Biology*, 14:R115, 2013.

[216] Timothy V Pyrkov, Evgeny Getmantsev, Boris Zhurov, Konstantin Avchaciov, Mikhail Pyatnitskiy, Leonid Menshikov, Kristina Khodova, Andrei V Gudkov, and Peter O Fedichev. Quantitative characterization of biological age and frailty based on locomotor activity records. *Aging*, 10(10):2973–2990, October 2018.

[217] David Lopez-Paz and Maxime Oquab. Revisiting classifier two-sample tests. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

[218] Daniele Bertolini, Anton D. Loukianov, Aaron M. Smith, David Li-Bland, Yannick Pouliot, Jonathan R. Walsh, and Charles K. Fisher. Modeling disease progression in mild cognitive impairment and Alzheimer's disease with digital twins. 2020.

[219] Yu-Zhong Chen and Ying-Cheng Lai. Sparse dynamical Boltzmann machine for reconstructing complex networks with binary dynamics. *Physical Review E*, 97:032317, 2018.

[220] Yulia Rubanova, Tian Qi Chen, and David Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. *NeurIPS*, 2019.

[221] Edward De Brouwer, Jaak Simm, Adam Arany, and Yves Moreau. GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series. *NeurIPS*, pages 7377–7388, 2019.

[222] James Jordon, Alan Wilson, and Mihaela van der Schaar. Synthetic data: Opening the data floodgates to enable faster, more directed development of machine learning methods. 2020.

[223] A synthetic population of nearly 10^7 individuals with 20 years of annually sampled trajectories from baseline is available at https://zenodo.org/record/4733386.

[224] C. W. J. Granger. Economic processes involving feedback. *Information and Control*, 6:28–48, 1963.

[225] K. J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *NeuroImage*, 19(4):1273–1302, 2003.

[226] K. J. Friston, Katrin H. Preller, Chris Mathys, Hayriye Cagnan, Jakob Heinzle, Adeel Razi, and Peter Zeidman. Dynamic causal modelling revisited. *NeuroImage*, 199:730–744, 2019.

[227] Shuai Xiao, Junchi Yan, Xiaokang Yang, Hongyuan Zha, and Stephen M. Chu. Modeling the intensity function of point process via recurrent neural networkss. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, page 1597–1603, 2017.

[228] Shuai Xiao, Junchi Yan, Mehrdad Farajtabar, Le Song, Xiaokang Yang, and Hongyuan Zha. Learning time series associated event sequences with recurrent point process networks. *IEEE Transactions on Neural Networks and Learning Systems*, 30(10):3124–3136, 2019.

[229] Z. Qian, A. Alaa, Alexis Bellot, J. Rashbass, and M. van der Schaar. Learning dynamic and personalized comorbidity networks from event data using deep diffusion processes. In *AISTATS*, 2020.

[230] Laurie E Davies, Gemma Spiers, Andrew Kingston, Adam Todd, Joy Adamson, and Barbara Hanratty. Adverse outcomes of polypharmacy in older people: Systematic review of reviews. *Journal of the American Medical Directors Association*, 21(2):181–187, 2020.

[231] Amanda J Miller, Olga Theou, Miranda McMillan, Susan E Howlett, Karthik K Tennankore, and Kenneth Rockwood. Dysnatremia in relation to frailty and age in community-dwelling adults in the National Health and Nutrition Examination Survey. *Journals of Gerontology A*, 72(3):376–381, March 2017.

[232] A. Golightly and D.j. Wilkinson. Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics Data Analysis*, 52(3):1674–1693, 2008.

[233] Gavin A. Whitaker, Andrew Golightly, Richard J. Boys, and Chris Sherlock. Bayesian inference for diffusion-driven mixed-effects models. *Bayesian Analysis*, 12(2):435–463, 2017.

[234] Cédric Archambeau, Manfred Opper, Yuan Shen, Dan Cornford, and John S. Shawe-Taylor. Variational inference for diffusion processes. *Advances in Neural Information Processing Systems*, 20:17–24, 2008.

[235] Manfred Opper. Variational inference for stochastic differential equations. *Annalen der Physik*, 531(3):1800233, 2019.

[236] Xuechen Li, Ting-Kam Leonard Wong, Ricky T. Q. Chen, and David Duvenaud. Scalable gradients for stochastic differential equations. *arXiv*, 2001.01328, 2020.

[237] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

[238] Kan Ren, Jiarui Qin, Lei Zheng, Zhengyu Yang, Weinan Zhang, Lin Qiu, and Yong Yu. Deep recurrent survival analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4798–4805, 2019.

[239] Andreas Rößler. Runge–Kutta methods for the strong approximation of solutions of stochastic differential equations. *SIAM J. Numer. Anal.*, 48(3):922–952, January 2010.

[240] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc, New York, NY, USA, 2001.

[241] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[242] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: B*, 67(2):301–320, 2005.

[243] Swadhin Taneja, Arnold B Mitnitski, Kenneth Rockwood, and Andrew D Rutenberg. Dynamical network model for age-related health deficits and mortality. *Physical Review E*, 93(2):022309–11, 2016.

[244] Garrett Stubbings. Laboratory health measures and optimal structures for aging. Master's thesis, Dalhousie University, 2021.

[245] Ryan N. Gutenkunst, Joshua J. Waterfall, Fergal P. Casey, Kevin S. Brown, Christopher R. Myers, and James P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3:10, 2007.

[246] Yuanjun Gao, Evan Archer, Liam Paninski, and John P Cunningham. Linear dynamical neural population models through nonlinear embeddings. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 163–171, Red Hook, NY, USA, December 2016.

[247] Çağatay Yıldız, Markus Heinonen, and Harri Lahdesmäki. ODE^{2}VAE: deep generative second order ODEs with bayesian neural networks. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 13434–13443. Red Hook, NY, USA, 2019.

[248] Alexander B Wiltschko, Matthew J Johnson, Giuliano Iurilli, Ralph E Peterson, Jesse M Katon, Stan L Pashkovski, Victoria E Abraira, Ryan P Adams, and Sandeep Robert Datta. Mapping Sub-Second structure in mouse behavior. *Neuron*, 88(6):1121–1135, 2015.

[249] Christos Louizos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6449–6459, Red Hook, NY, USA, 2017.

[250] Matthew J Johnson, David K Duvenaud, Alex Wiltschko, Ryan P Adams, and Sandeep R Datta. Composing graphical models with neural networks for structured representations and fast inference. In D Lee, M Sugiyama, U Luxburg, I Guyon, and R Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29, 2016.

[251] Luke Y Prince, Shahab Bakhtiari, Colleen J Gillon, and Blake A Richards. Parallel inference of hierarchical latent dynamics in two-photon calcium imaging of neuronal populations. 2021.

[252] Evgeny Putin, Polina Mamoshina, Alexander Aliper, Mikhail Korzinkin, Alexey Moskalev, Alexey Kolosov, Alexander Ostrovskiy, Charles Cantor, Jan Vijg, and Alex Zhavoronkov. Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging*, 8(5):1021–1033, 2016.

[253] Michael B Schultz, Alice E Kane, Sarah J Mitchell, Michael R MacArthur, Elisa Warner, David S Vogel, James R Mitchell, Susan E Howlett, Michael S Bonkowski, and David A Sinclair. Age and life expectancy clocks based on machine learning analysis of mouse frailty. *Nat. Commun.*, 11(1):1–12, 2020.

[254] Alex Zhavoronkov and Polina Mamoshina. Deep aging clocks: The emergence of AI-based biomarkers of aging and longevity. *Trends in Pharmacological Sciences*, 40(8):546–549, 2019.

[255] Konstantin G Arbeev, Olivia Bagley, Svetlana V Ukraintseva, Deqing Wu, Hongzhe Duan, Alexander M Kulminski, Eric Stallard, Kaare Christensen, Joseph H Lee, Bharat Thyagarajan, Joseph M Zmuda, and Anatoliy I Yashin. Genetics of physiological dysregulation: findings from the long life family study using joint models. *Aging*, 12(7):5920–5947, April 2020.

[256] William B Zhang, Drew B Sinha, William E Pittman, Erik Hvatum, Nicholas Stroustrup, and Zachary Pincus. Extended twilight among isogenic *C. elegans* causes a disproportionate scaling between lifespan and health. *Cell Systems*, 3(4):333–345.e4, October 2016.

[257] Céline N Martineau, Bora Baskaner, Renée I Seinstra, William R Schafer, André E X Brown, Ellen A A Nollen, and Patrick Laurent. Deep behavioural phenotyping reveals divergent trajectories of ageing and quantifies health state in c. elegans. 2019.

[258] Zachary Pincus, Thalyana Smith-Vikos, and Frank J Slack. MicroRNA predictors of longevity in caenorhabditis elegans. *PLoS Genet.*, 7(9):e1002306, 2011.

[259] Zachary Pincus, Travis C Mazer, and Frank J Slack. Autofluorescence as a measure of senescence in c. elegans: look to red, not blue or green. *Aging*, 8(5):889–898, 2016.

[260] Lior Shamir, Catherine A Wolkow, and Ilya G Goldberg. Quantitative measurement of aging using image texture entropy. *Bioinformatics*, 25(23):3060–3063, 2009.

[261] J Martin Bland and Douglas G Altman. The logrank test. *BMJ*, 328(7447):1073, 2004.

[262] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *J. Mach. Learn. Res.*, 9(86):2579–2605, 2008.

[263] Tomoharu Iwata, David Duvenaud, and Zoubin Ghahramani. Warped mixtures for nonparametric cluster shapes. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 311–320, Arlington, Virginia, USA, 2013.

[264] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, 2009.

[265] Morgan E. Levine, Ake T. Lu, Austin Quach, Brian H. Chen, Themistocles L. Assimes, Stefania Bandinelli, Lifang Hou, Andrea A. Baccarelli, James D. Stewart, Yun Li, Eric A. Whitsel, James G Wilson, Alex P Reiner, Abraham Aviv, Kurt Lohman, Yongmei Liu, Luigi Ferrucci, and Steve Horvath. An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, 10(4):573–591, 2018.

[266] Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Stat. Sin.*, 4(2):639–650, 1994.

[267] Shengjia Zhao, Jiaming Song, and Stefano Ermon. Learning hierarchical features from deep generative models. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 4091–4099. PMLR, 2017.

[268] David M Blei and Michael I Jordan. Variational methods for the dirichlet process. In *Proceedings of the twenty-first international conference on Machine learning*, ICML '04, page 12, New York, NY, USA, 2004.

[269] Zachary Zimmer, Yasuhiko Saito, Olga Theou, Clove Haviva, and Kenneth Rockwood. Education, wealth, and duration of life expected in various degrees of frailty. *Eur. J. Ageing*, 18(3):393–404, September 2021.

[270] Emiel O Hoogendijk, Jonathan Afilalo, Kristine E Ensrud, Paul Kowal, Graziano Onder, and Linda P Fried. Frailty: implications for clinical practice and public health. *Lancet*, 394(10206):1365–1375, 2019.

[271] Gotaro Kojima, Steve Iliffe, and Kate Walters. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age Ageing*, 47(2):193–200, 2018.

[272] Evan C Hadley, George A Kuchel, Anne B Newman, and Workshop Speakers and Participants. Report: NIA workshop on measures of physiologic resiliencies in human aging. *J. Gerontol. A Biol. Sci. Med. Sci.*, 72(7):980–990, 2017.

[273] Jerrald L Rector, Sanne M W Gijzel, Ingrid A van de Leemput, Fokke B van Meulen, Marcel G M Olde Rikkert, and René J F Melis. Dynamical indicators of resilience from physiological time series in geriatric inpatients: Lessons learned. *Exp. Gerontol.*, 149:111341, 2021.

[274] Andres Kriete. Robustness and aging–a systems-level perspective. *Biosystems.*, 112(1):37–48, April 2013.

[275] Felipe Sierra. The emergence of geroscience as an interdisciplinary approach to the enhancement of health span and life span. *Cold Spring Harb. Perspect. Med.*, 6(4):a025163, April 2016.

[276] Felipe Sierra, Avshalom Caspi, Richard H Fortinsky, Laura Haynes, Gordon J Lithgow, Terrie E Moffitt, S Jay Olshansky, Daniel Perry, Eric Verdin, and George A Kuchel. Moving geroscience from the bench to clinical care and health policy. *J. Am. Geriatr. Soc.*, 2021.

[277] A. Phelps, M. Marmot, A. Oskala, S. Clemens, J. Banks, N. Rogers, A. Steptoe, M. Blake, J. Nazroo, and Z. Oldfield. English longitudinal study of ageing: Waves 0-9, 1998-2019. [data collection] 31st edition. *UK Data Service*, 2020.

[278] E H Gordon and R E Hubbard. Physiological basis for sex differences in frailty. *Current Opinion in Physiology*, 6:10–15, December 2018.

[279] Xingqiu Zhao, Qiang Zhao, Jainguo Sun, and Jong S Kim. Generalized log-rank tests for partly interval-censored failure time data. *Biom. J.*, 50(3):375–385, June 2008.

[280] Erwin Stolz, Hannes Mayerl, Emiel O Hoogendijk, Joshua J Armstrong, Regina Roller-Wirnsberger, and Wolfgang Freidl. Acceleration of health deficit accumulation in late-life: evidence of terminal decline in frailty index three years before death in the US health and retirement study. *Ann. Epidemiol.*, 58:156–161, 2021.

[281] Arnold Mitnitski, Nader Fallah, Yougui Wu, Kenneth Rockwood, and Amy R Borenstein. Changes in cognition during the course of eight years in elderly japanese americans: a multistate transition model. *Ann. Epidemiol.*, 20(6):480–486, June 2010.

[282] Sanne M W Gijzel, Ingrid A van de Leemput, Marten Scheffer, Mattia Roppolo, Marcel G M Olde Rikkert, and René J F Melis. Dynamical resilience indicators in time series of self-rated health correspond to frailty levels in older adults. *J. Gerontol. A Biol. Sci. Med. Sci.*, 72(7):991–996, July 2017.

[283] Jocelyne C Whitehead, Barbara A Hildebrand, Michael Sun, Michael R Rockwood, Robert A Rose, Kenneth Rockwood, and Susan E Howlett. A clinical frailty index in aging mice: comparisons with frailty index data in humans. *J. Gerontol. A Biol. Sci. Med. Sci.*, 69(6):621–632, 2014.

[284] Graeme L. Hickey, Pete Philipson, Andrea Jorgenson, and Kolamunnage-Dona. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, 16(117), 2016.

[285] J. O. Ramsay. Monotone regression splines in action. *Statist. Sci.*, 3(4):425–441, 11 1988.

[286] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, 2020.

[287] Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. Visualization in bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):389–402, 2019.

[288] Andrew Gelman, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. Bayesian workflow, 2020.

[289] Anatoli I. Yashin, Konstantin G. Arbeev, Igor Akushevich, Aliaksandr Kulminski, Lucy Akushevich, and Svetlana V. Ukraintseva. Stochastic model for analysis of longitudinal data on aging and mortality. *Mathematical Biosciences*, 208:538–551, 2007.

[290] Charles K Fisher, Aaron M Smith, and Jonathan R Walsh. Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Scientific Reports*, 9(1):1–14, 2019.

[291] Jonathon R. Walsh, Aaron M. Smith, Yannick Pouliot, David Li-Bland, Anton Loukianov, and Charles K. Fisher. Generating digital twins with multiple sclerosis using probabilistic neural networks. *arXiv*, 2002:02779v1, 2020.

[292] Rui Chen, George I Mias, Jennifer Li-Pook-Than, Lihua Jiang, Hugo Y K Lam, Rong Chen, Elana Miriami, Konrad J Karczewski, Manoj Hariharan, Frederick E Dewey, Yong Cheng, Michael J Clark, Hogune Im, Lukas Habegger, Suganthi Balasubramanian, Maeve O'Huallachain, Joel T Dudley, Sara Hillenmeyer, Rajini Haraksingh, Donald Sharon, Ghia Euskirchen, Phil Lacroute, Keith Bettinger, Alan P Boyle, Maya Kasowski, Fabian Grubert, Scott Seki, Marco Garcia, Michelle Whirl-Carrillo, Mercedes Gallardo, Maria A Blasco, Peter L Greenberg, Phyllis Snyder, Teri E Klein, Russ B Altman, Atul J Butte, Euan A Ashley, Mark Gerstein, Kari C Nadeau, Hua Tang, and Michael Snyder. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*, 148(6):1293–1307, March 2012.

[293] Wenyu Zhou, M Reza Sailani, Kévin Contrepois, Yanjiao Zhou, Sara Ahadi, Shana R Leopold, Martin J Zhang, Varsha Rao, Monika Avina, Tejaswini Mishra, Jethro Johnson, Brittany Lee-McMullen, Songjie Chen, Ahmed A Metwally, Thi Dong Binh Tran, Hoan Nguyen, Xin Zhou, Brandon Albright, Bo-Young Hong, Lauren Petersen, Eddy Bautista, Blake Hanson, Lei Chen, Daniel Spakowicz, Amir Bahmani, Denis Salins, Benjamin Leopold, Melanie Ashland, Orit Dagan-Rosenfeld, Shannon Rego, Patricia Limcaoco, Elizabeth Colbert, Candice Allister, Dalia Perelman, Colleen Craig, Eric Wei, Hassan Chaib, Daniel Hornburg, Jessilyn Dunn, Liang Liang, Sophia Miryam Schüssler-Fiorenza Rose, Kim Kukurba, Brian Piening, Hannes Rost, David Tse, Tracey McLaughlin, Erica Sodergren, George M Weinstock, and Michael Snyder. Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature*, 569(7758):663–671, May 2019.

[294] Sophia Miryam Schüssler-Fiorenza Rose, Kévin Contrepois, Kegan J Moneghetti, Wenyu Zhou, Tejaswini Mishra, Samson Mataraso, Orit Dagan-Rosenfeld, Ariel B Ganz, Jessilyn Dunn, Daniel Hornburg, Shannon Rego, Dalia Perelman, Sara Ahadi, M Reza Sailani, Yanjiao Zhou, Shana R Leopold, Jieming Chen, Melanie Ashland, Jeffrey W Christle, Monika Avina, Patricia Limcaoco, Camilo Ruiz, Marilyn Tan, Atul J Butte, George M Weinstock, George M Slavich, Erica Sodergren, Tracey L McLaughlin, Francois Haddad, and Michael P Snyder. A longitudinal big data approach for precision health. *Nat. Med.*, 25(5):792–804, May 2019.

[295] Abdellah Tebani, Anders Gummesson, Wen Zhong, Ina Schuppe Koistinen, Tadepally Lakshmikanth, Lisa M Olsson, Fredrik Boulund, Maja Neiman, Hans Stenlund, Cecilia Hellström, Max J Karlsson, Muhammad Arif, Tea Dodig-Crnković, Adil Mardinoglu, Sunjae Lee, Cheng Zhang, Yang Chen, Axel Olin, Jaromir Mikes, Hanna Danielsson, Kalle von Feilitzen, Per-Anders Jansson, Oskar Angerås, Mikael Huss, Sanela Kjellqvist, Jacob Odeberg, Fredrik Edfors, Valentina Tremaroli, Björn Forsström, Jochen M Schwenk, Peter Nilsson, Thomas Moritz, Fredrik Bäckhed, Lars Engstrand, Petter Brodin, Göran Bergström, Mathias Uhlen, and Linn Fagerberg. Integration of molecular profiles in a longitudinal wellness profiling cohort. *Nat. Commun.*, 11(1):1–14, September 2020.

[296] Scott Linderman, Annika Nichols, David Blei, Manuel Zimmer, and Liam Paninski. Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in c. elegans. April 2019.

[297] V. Holubec, P. Chvosta, M. Einax, and P. Maass. Attempt time monte carlo: An alternative for simulation of stochastic jump processes with time-dependent transition rates. *Europhysics Letters*, 93:40003, 2011.

[298] Jeff Racine and Qi Li. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119:99–130, 2004.

[299] Chi-Yang Chu, Daniel J. Henderson, and Christopher F. Parmeter. Plug-in bandwidth selection for kernel density estimation with discrete data. *Econometrics*, 3:199–214, 2015.

[300] Elre T. Oldewage, Andries P. Engelbrecht, and Christopher W. Cleghorn. Boundary constraint handling techniques for particle swarm optimization in high dimensional problem spaces. *ANTS 2018: Swarm Intelligence*, pages 333–341, 2018.

Appendix A

Chapter 4 supplemental information

A.1 Simulation details

A.1.1 Stochastic simulation

We exactly simulate the model for each individual using a standard stochastic simulation algorithm [144]. Two separate rate processes are simulated for damage and mortality. The first process for damage uses the rate in Equation 4.5 in the main text. The second process uses the mortality rate in Equation 4.6.

Since our rates are time-dependent we use an exact rejection sampling method, called the "attempt time algorithm" [297]. We set the window for possible events as $T = 5$ years, and re-update this maximum horizon as it is passed. Since we have constrained all rates to be monotonic in time (see below for details), we can easily determine an upper bound on the rates, necessary to implement the attempt time algorithm: $\Gamma^{+}_{\max} = \sum_i (1 - d_i)\Gamma^{+}_i(t + T, \{d_j\})$ bounds the total damage rate of all undamaged nodes, while $\Gamma^{D}_{\max} = \Gamma_D(t + T, \{d_j\})$ bounds the mortality rate.

At each step of the algorithm, the event is chosen from the process with the smallest time-to-event, i.e., damage occurs until the time to death is smaller than the time to next damage event. Within the damage process, we choose which event using a computationally efficient tree-based approach [145]. When the mortality event occurs, the algorithm terminates for that individual.

A.1.2 Estimating likelihood

To calculate the likelihood in Equation 4.8, we run $\mathbb{S}$ simulations with the same set of parameters (with $\mathbb{S} > 10^6$ individuals). We use a discrete kernel density estimate of the health term of the likelihood,

$$p(\{d_i\}^{(m)}|t^{(m)};\vec{\theta}) = \frac{1}{\mathbb{S}(t^{(m)})} \sum_{s=1}^{\mathbb{S}(t^{(m)})} K(\{d_i\}^{(s)}, \{d_i\}^{(m)}; \lambda^{(m)}), \tag{A.1}$$

where $\mathbb{S}(t^{(m)})$ is the number of simulated individuals alive at age $t^{(m)}$. The superscript (s) indicates a simulated individual and (m) an observed individual from the data. We use the Racine and Li kernel [298],

$$K(\{d_i\}^{(s)}, \{d_i\}^{(m)}; \lambda^{(m)}) = \begin{cases} \frac{1}{1+\lambda^{(m)}}, & \sum_i |d_i^{(s)} - d_i^{(m)}| = 0,\ t^{(s)} = t^{(m)} \\ \frac{\lambda^{(m)}}{1+\lambda^{(m)}}, & \sum_i |d_i^{(s)} - d_i^{(m)}| = 1,\ t^{(s)} = t^{(m)} \\ 0, & \text{otherwise.} \end{cases} \tag{A.2}$$

This allows individuals that differ in damage by 1 deficit to contribute to the likelihood, as determined by an individual bandwidth $\lambda^{(m)}$.

The bandwidth $\lambda^{(m)}$ is selected by minimizing the mean-squared error [299],

$$\lambda^{(m)} = \left[1 + \frac{\mathbb{S}(t^{(m)})((p^{(m)})^2 + (1-p^{(m)})^2)}{2p^{(m)}(1-p^{(m)})}\right]^{-1}, \tag{A.3}$$

where $p^{(m)}$ is the empirical frequency estimate of the distribution (obtained using bandwidth $\lambda = 0$). The bandwidth $\lambda^{(m)}$ is individual-dependent, so that the well sampled individuals will have a small bandwidth, and poorly sampled individuals have a larger bandwidth to reduce noise.

Both the censored and uncensored mortality terms of the likelihood are calculated by binning the death ages of simulated individuals that match the observed data ($\sum_i |d_i^{(s)} - d_i^{(m)}| = 0,\ t^{(s)} = t^{(m)}$) with 1 year bins.

A.1.3 Regularization

To have an increase in mortality rate with decreasing health, the deficit mortality contributions β_j in Equation 4.7 should all be positive. Without such a bound, we have observed that the optimization sometimes converges to positive and negative β_j values scattered around zero, which leads to the deficit contribution to the mortality rate being negligible, and gives uniform mortality for all individuals. However in contrast, setting a strict bound $\beta_j \geq 0$ causes the optimization to converge to parameters which perform poorly with respect to the health trajectories. Instead, we have found that a soft penalty behaves well when added to the log-likelihood to penalize negative values,

$$+C \sum_j \min(0, \beta_j). \tag{A.4}$$

Values of $\beta_j \sim 1$ give a significant mortality contribution, and log-likelihood values are estimated with a stochastic noise of ± 10, so we choose $C = 100$ by hand to achieve a moderate regularizing effect – and we find that the optimization performs well with this choice.

A.1.4 Parameter bounds

To ensure Equations 4.4, 4.5, 4.6, and 4.7 are monotonically increasing for all $t \geq 0$, we bound the parameters so that their time-derivatives have only negative real or complex roots with respect to t or f_i. For hyperparameters $n_+ = 4$, and $n_f = n_{D_1} = n_{D_2} = 3$, the required bounds are shown in Table A.1.

Table A.1: Parameter bounds during optimization.

Equation	**Parameter bounds**
Equation 4.4: $f_i(t, \{d_i\})$	$\mu_{i1} \geq 0$, $\mu_{i3} \geq 0$, $\mu_{i2} \geq -\sqrt{3\mu_{i1}\mu_{i3}}$
Equation 4.5: $\Gamma_i^+(t, \{d_i\})$	$\gamma_{i1} \geq 0$, $\gamma_{i3} \geq 0$, $\gamma_{i2} \geq -\sqrt{3\gamma_{i1}\gamma_{i3}}$, $\gamma_{i4} \geq 0$
Equation 4.6: $\Gamma_D(t, \{d_i\})$	$\alpha_1 \geq 0$, $\alpha_3 \geq 0$, $\alpha_2 \geq -\sqrt{3\alpha_1\alpha_3}$
Equation 4.7: $x(t, \{d_i\})$	$\eta_1 \geq 0$, $\eta_3 \geq 0$, $\eta_2 \geq -\sqrt{3\eta_1\eta_3}$

A.1.5 Parameter optimization

Optimization is done with particle swarm optimization (PSO) [198]. PSO is derivative-free and highly parallelizable. We use the standard version:

$$\mathbf{v}_{i,t+1} = \omega \mathbf{v}_{i,t} + c_p u_p(\vec{\theta}^p_{i,t} - \vec{\theta}_{i,t}) + c_g u_g(\vec{\theta}^g_t - \vec{\theta}_{i,t}) \quad \text{(A.5)}$$

$$\vec{\theta}_{i,t+1} = \vec{\theta}_{i,t} + \mathbf{v}_{i,t+1}. \quad \text{(A.6)}$$

The parameter values for the ith particle at an iteration t are represented as $\theta_{i,t}$ and the velocity for this particle as $\mathbf{v}_{i,t}$. The current best set of parameters found by particle i is $\theta^p_{i,t}$, and the current global best set of parameters is θ^g_t. We randomly sample u_p, u_g uniformly from $[0, 1]$. We set $\omega = 0.7$, $c_p = c_g = 2$ and use $100 - 200$ particles, depending on compute resources available. The optimization typically converges in 100-200 iterations.

Parameter bounds from section A.1.4 are implemented by rescaling the velocity components $v_{i,k,t}$ with a hyperbolic method so that parameters never exceed their upper bounds U_k and lower bounds L_k [300],

$$v_{i,k,t+1} = \frac{v_{i,k,t+1}}{1 + \frac{v_{i,k,t+1}}{U_k - \theta_{i,k,t}}}, v_{i,k,t+1} > 0, \tag{A.7}$$
$$v_{i,k,t+1} = \frac{v_{i,k,t+1}}{1 - \frac{v_{i,k,t+1}}{\theta_{i,k,t} - L_k}}, v_{i,k,t+1} < 0.$$

Our objective function is stochastic, so to avoid optimizing to the noise, every 10 PSO iterations we recalculate the global maximum likelihood for the global best set of parameters θ_t^g, and the individual particle maximum likelihood values for each particle's best set of parameters $\theta_{i,t}^p$.

A.2 Parameter robustness

To examine parameter robustness, we select new parameter sets by randomly selecting 25% of parameters to randomly perturb from their maximum likelihood values $\hat{\theta}_d$ within the range $\theta_d \in [0.1\hat{\theta}_d, 10\hat{\theta}_d]$.

Broadly following the approach of [179], we weight these perturbed parameters by calculating an approximate log-likelihood by using only the χ^2 of first and second order joint distributions of health and population survival. This amounts to a Gaussian approximation to the log-likelihood,

$$\begin{aligned}
\mathcal{L}(\theta) \simeq & -\frac{1}{N(T_{\max} - T_{\min})} \sum_{i=1}^{N} \int_{T_{\min}}^{T_{\max}} \Big(\frac{\hat{p}(d_i = 1|t;\theta) - p(d_i = 1|t)}{\sigma_{\hat{p}}(t)} \Big)^2 dt \qquad \text{(A.8)} \\
& - \frac{2}{N(N-1)(T_{\max} - T_{\min})} \sum_{i,j \neq i} \int_{T_{\min}}^{T_{\max}} \Big(\frac{\hat{p}(d_i = 1, d_j = 1|t;\theta) - p(d_i = 1, d_j = 1|t)}{\sigma_{\hat{p}}(t)} \Big)^2 dt \\
& - \frac{1}{N(N-1)(T_{\max} - T_{\min})} \sum_{i,j} \int_{T_{\min}}^{T_{\max}} \Big(\frac{\hat{p}(d_i = 1, d_j = 0|t;\theta) - p(d_i = 1, d_j = 0|t)}{\sigma_{\hat{p}}(t)} \Big)^2 dt \\
& - \frac{1}{T_{\max} - T_{\min}} \int_{T_{\min}}^{T_{\max}} \Big(\frac{\langle \hat{S}(a|\{d_i\}, t;\theta) \rangle - S(a|t)}{\sigma_{\hat{S}}(t)} \Big)^2 dt.
\end{aligned}$$

We use this estimate of $\mathcal{L}$ to efficiently estimate any $f(\theta)$:

$$\langle f(\theta) \rangle = \frac{\sum_{\theta_i} e^{\mathcal{L}(\theta)} f(\theta)}{\sum_{\theta_i} e^{\mathcal{L}(\theta)}}, \tag{A.9}$$
$$\sigma_f^2 = \langle f(\theta)^2 \rangle - \langle f(\theta) \rangle^2. \tag{A.10}$$

In Figure A.1 we show deficit prevalence and survival averaged over likelihood-weighted parameterizations via Equations A.9 and A.10. These average predictions are quite close to the maximum likelihood estimates, which indicate that prevalences and mortality are robust predictions of our modelling approach.

In contrast, simple measures of the network structure do not appear to be robust. In Figure A.2 we perform the optimization 13 times, and the compare network degrees for the 10 nodes of our network. We find a broad ranges of degrees for each node, with significant overlap of these ranges between nodes. This indicates that though the behavior of the networks is similar (see e.g. Figure 4.7), the network structures themselves are not robustly predicted by the available data.

A.3 Supplemental figures

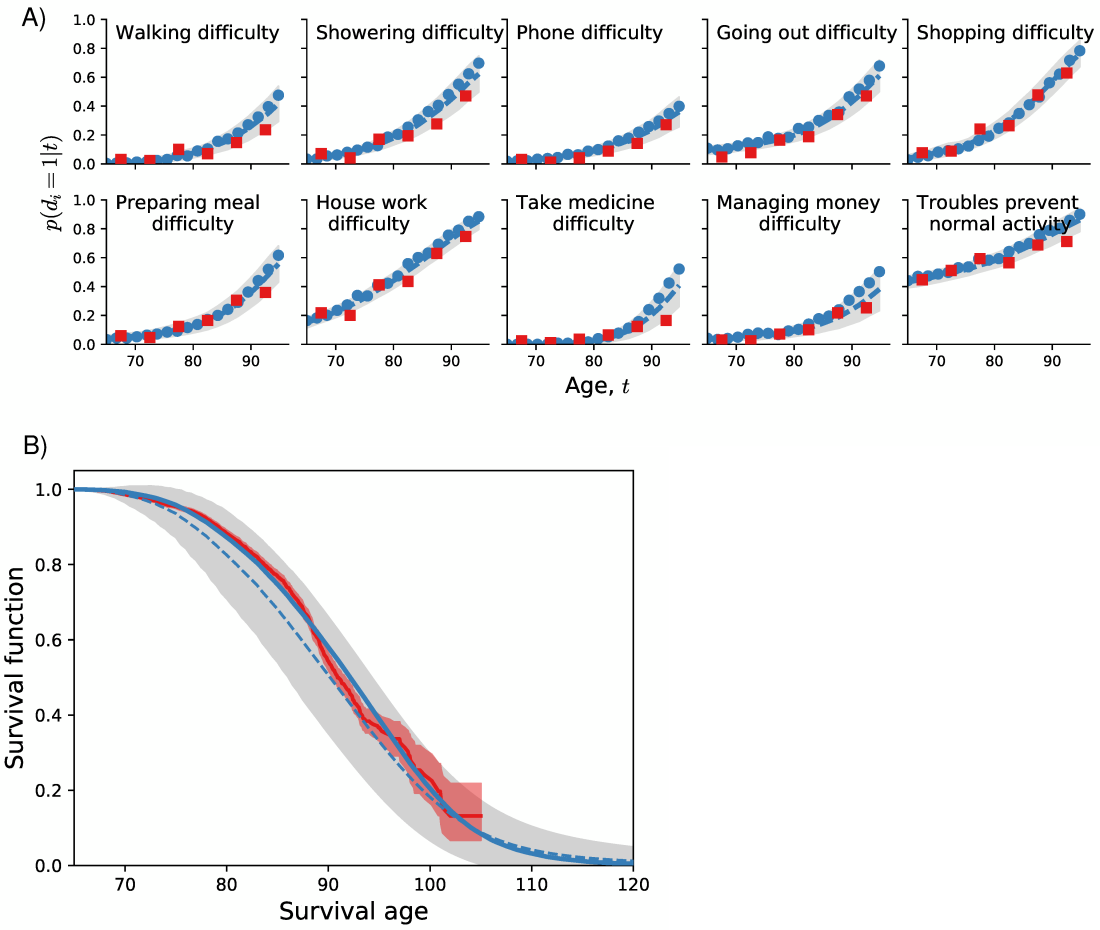


Figure A.1: **Prediction robustness.** A) Average predicted trajectories of ten deficit prevalences for individuals from the test data aged $65 - 70$ surviving past 70 (blue points). Observed CSHA prevalence is shown in red squares. B) Model population average survival function (blue line) and CSHA Kaplan-Meier population survival function (red line and shaded region). For both A) and B), the dashed blue lines are averages of the prevalence or individual survival curves over different parameter sets (Eqn. A.9), the shaded grey region shows one standard deviation (Eqn. A.10) from this average. Agreement of the shaded grey regions with our average results indicates that prevalences and survival are robustly predicted by our approach.

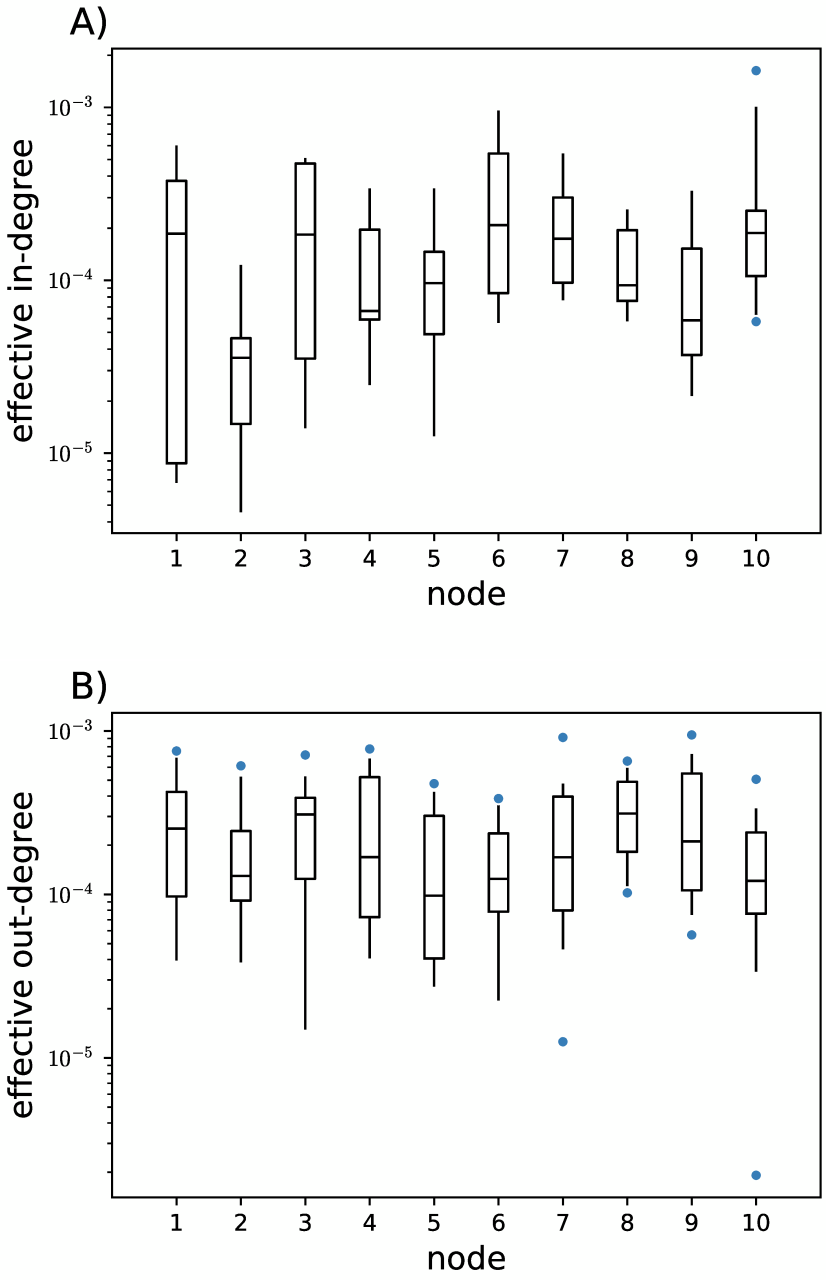


Figure A.2: **Network sloppiness.** Maximum likelihood effective network degrees for 13 numerical optimizations starting from different random seeds. Boxes show the interquartile range and whiskers show the 2.5% and 97.5% values. The network structure does not appear to be robust (note the log-scale). A) Effective in-degree $k_i^{\text{in}} = \sum_j w_{ij}\gamma_{i1}^+$ for each node. B) Effective out-degree $k_j^{\text{out}} = \sum_i w_{ij}\gamma_{i1}^+$ for each node.

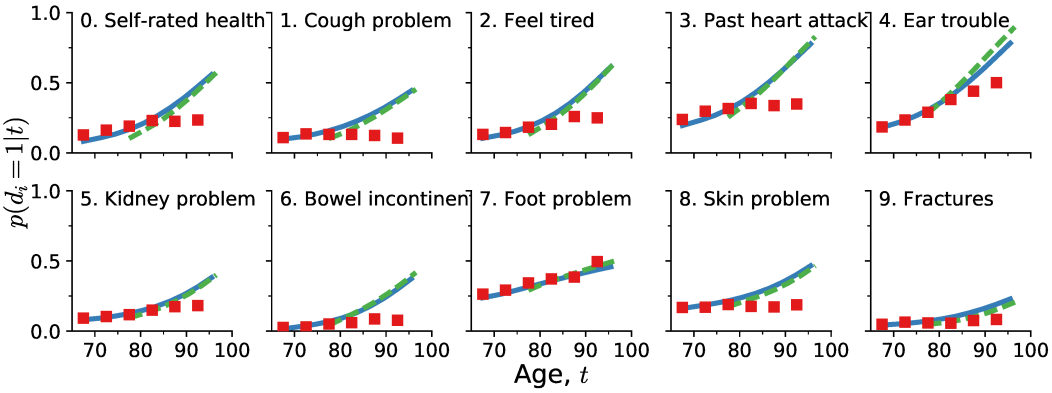


Figure A.3: **Alternate deficit predicted trajectories.** Average predicted trajectories of deficit prevalence for individuals from the test data aged $65-70$ surviving past 70 (solid blue lines) and aged $75-80$ surviving past 80 (dashed green lines). An alternate set of deficits are used as compared to Figure 4.2, in the main text. CSHA prevalence is shown in with red squares, standard errors are smaller than the point size.

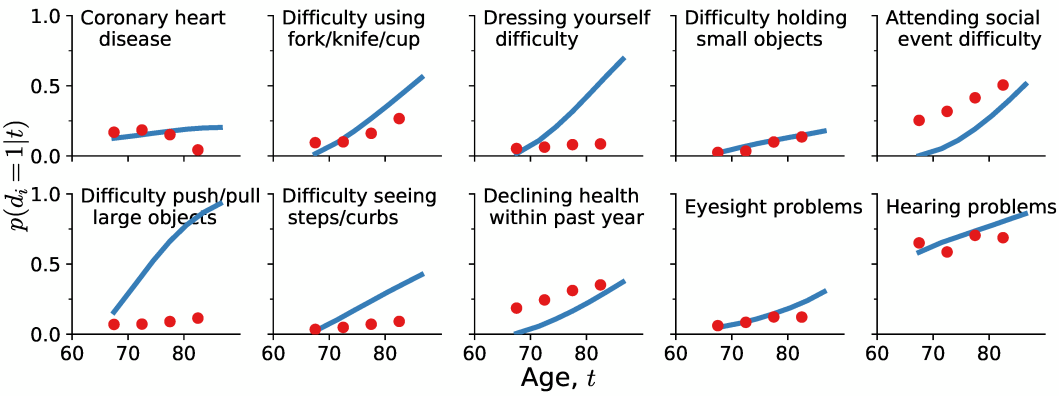


Figure A.4: **NHANES predicted trajectories.** Average predicted trajectories of deficit prevalence for individuals from the test data aged $65-70$ surviving past 70 (solid blue lines). Observed NHANES prevalence is shown with red circles; standard errors are smaller than the point size. Here predictions do not always start at value from the observational data (comparing the agreement for the first red point with Figure A.3) because there is missing data that we impute by simulating from birth.

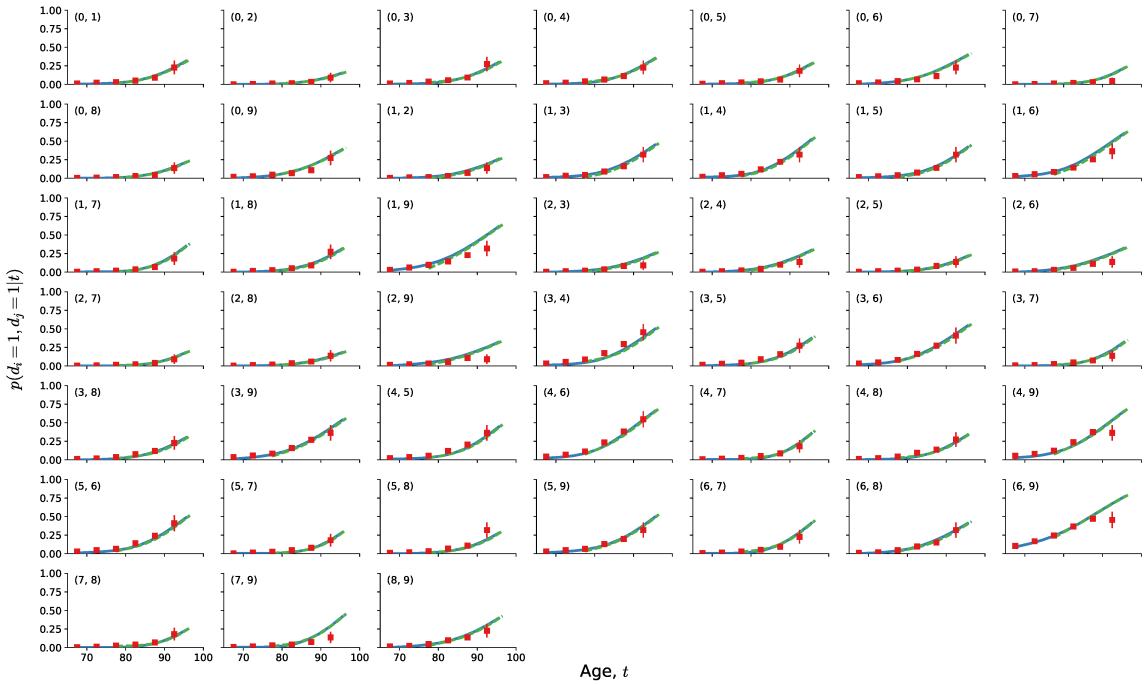


Figure A.5: **Full pair trajectories.** Average predicted trajectories of pairwise deficit prevalence for individuals from the test data aged $65-70$ surviving past 70 (solid blue lines) and aged $75-80$ surviving past 80 (dashed green lines). Subplot titles indicate the two deficits included in the pairwise prevalence, with numbers corresponding to titles in Figure 4.2, in the main text. All pairs are shown. CSHA data is shown by red squares; errorbars represent standard errors of the pairwise prevalence.

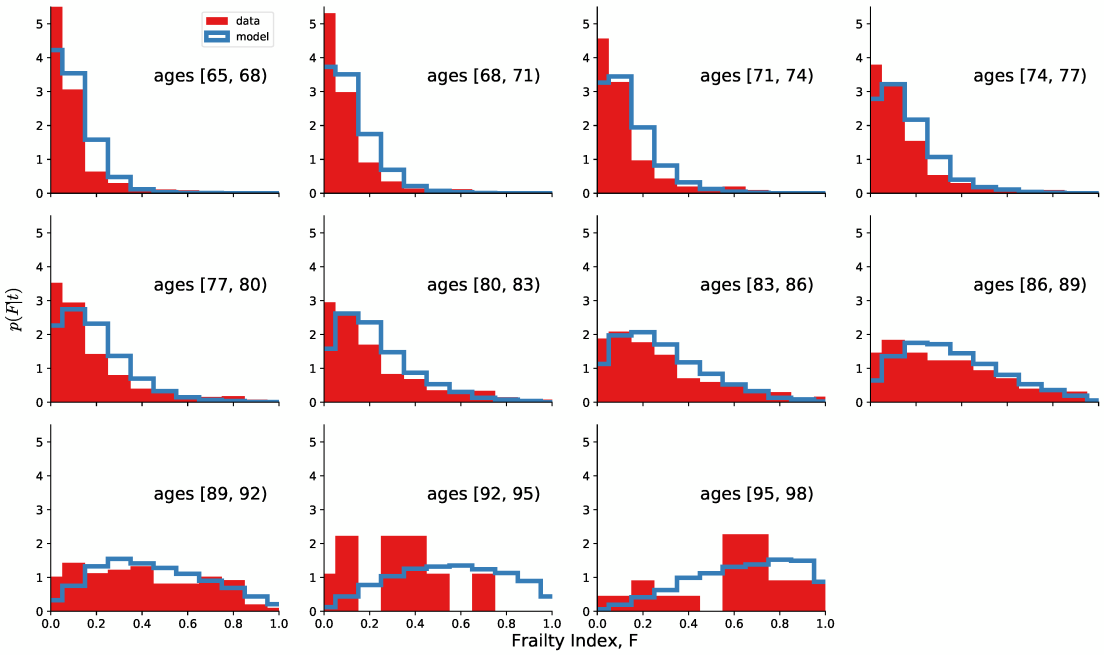


Figure A.6: **Population Frailty Index distributions.** Distribution of the Frailty Index, $F = \sum_{i=1}^{N} d_i/N$ for different age ranges for the model (unfilled blue histograms) and the CSHA data (filled red histograms). Note that the maximum FI for small numbers of deficits (in this case $N = 10$) is 1, as is reached in the CSHA data at later ages. Recovering this FI limit with model data, as well as broader consistency with the CSHA distributions, confirms that we capture the heterogeneity of the population as individuals age.

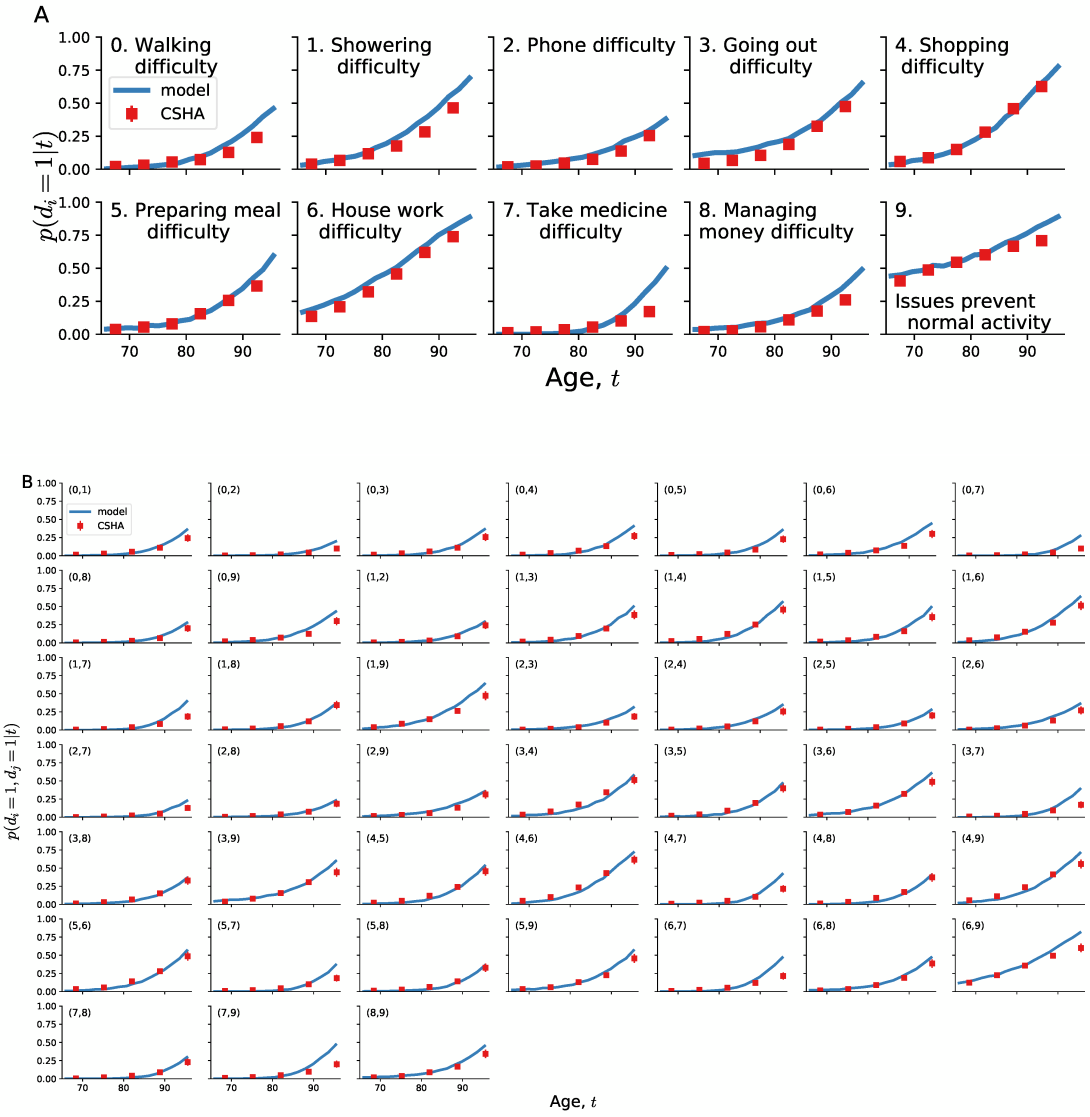


Figure A.7: **Simulated population trajectories from birth.** A) Predicted deficit prevalence $\hat{p}(d_i = 1|t)$ vs age t in our model simulated from zero damage at birth (blue lines) and binned prevalence from the CSHA (red squares). B) Predicted pair prevalence $\hat{p}(d_i = 1, d_j = 1|t)$ vs age in our model simulated from zero damage at birth (blue lines) is compared with binned pair prevalence from the CSHA data (red squares). Standard errors of the data are smaller than the point size.

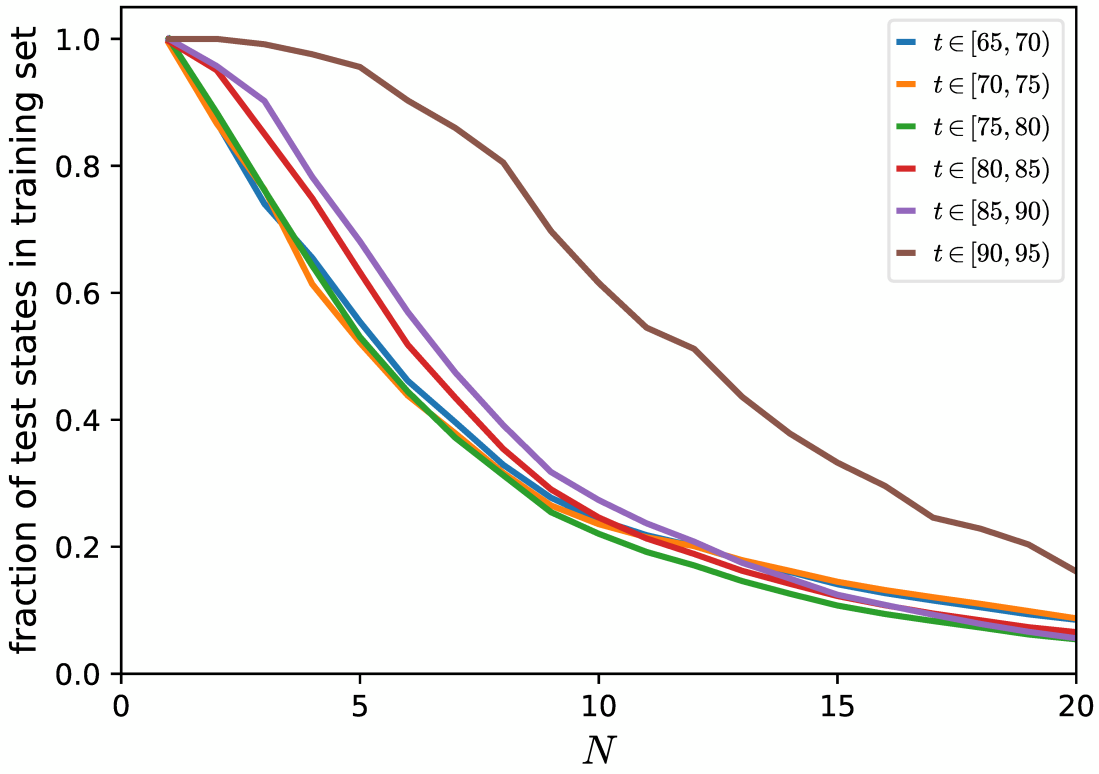


Figure A.8: **CSHA state overlap** Fraction of health states $\{d_i\}_{i=1}^{N}$ in the test set for different ranges of ages (as indicated by the legend) that are also in the training set at any age vs. N, the total number of health attributes (nodes) in our model. We average over permutations of the order that the deficits are chosen. We show that a significant fraction of the states in the test set are not in training set for $N = 10$, except for the oldest individuals aged $90 - 95$.

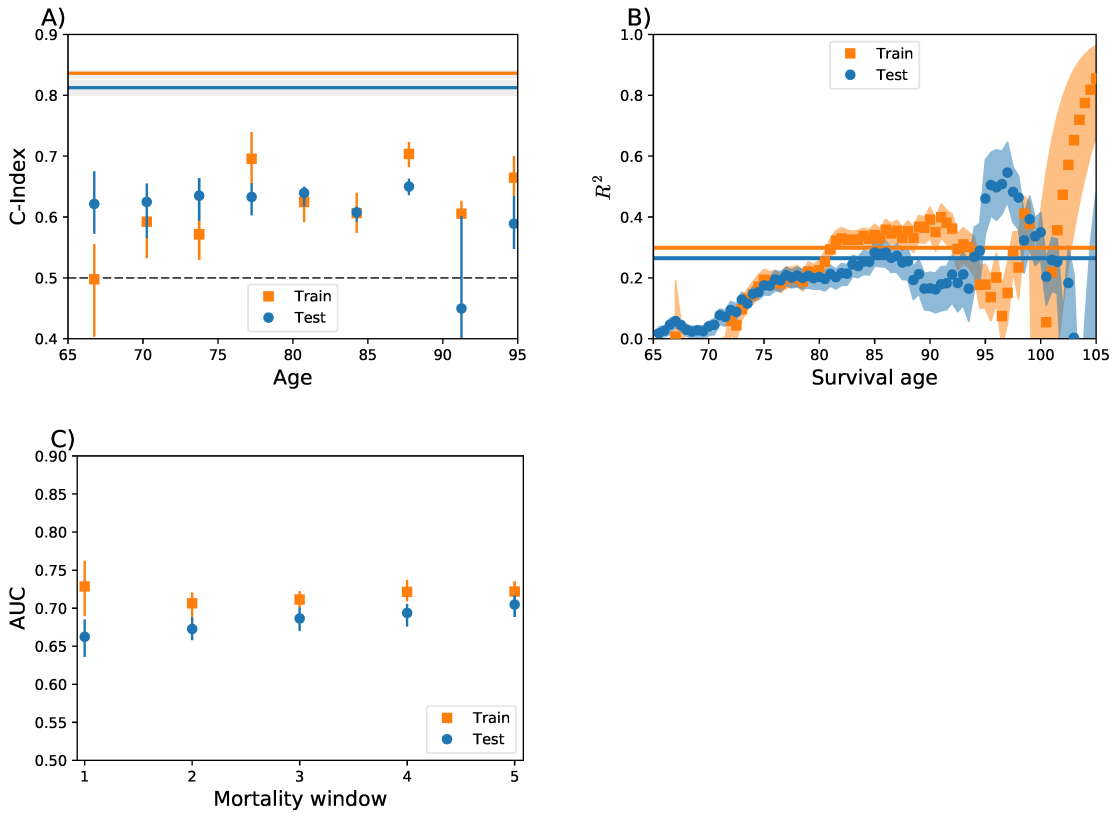


Figure A.9: **Training set mortality predictions.** We observe similar performance between train (orange squares and lines) and test (blue circles and lines) CSHA data, indicating a lack of overfitting. All errorbars and shaded regions show 95% confidence intervals. A) Survival C-Index stratified by age (points) and unstratified (solid lines). The baseline for random predictions is at $C = 0.5$ (dashed black line). B) Explained variance of survival $R^2(a) = 1 - \mathrm{BS}(a)/\mathrm{BS}_0(a)$ from the Brier score [178] is shown as points, with uncertainties indicated with corresponding shaded regions. Baseline Brier Score is calculated with the health-independent Kaplan-Meier estimate of the survival function, $S(a)$. Integrated R^2 over all ages is shown as solid lines. Values above zero show improvement above baseline. C) ROC AUC for the prediction of mortality within a window of time (in years). The baseline for random predictions is at $AUC = 0.5$.

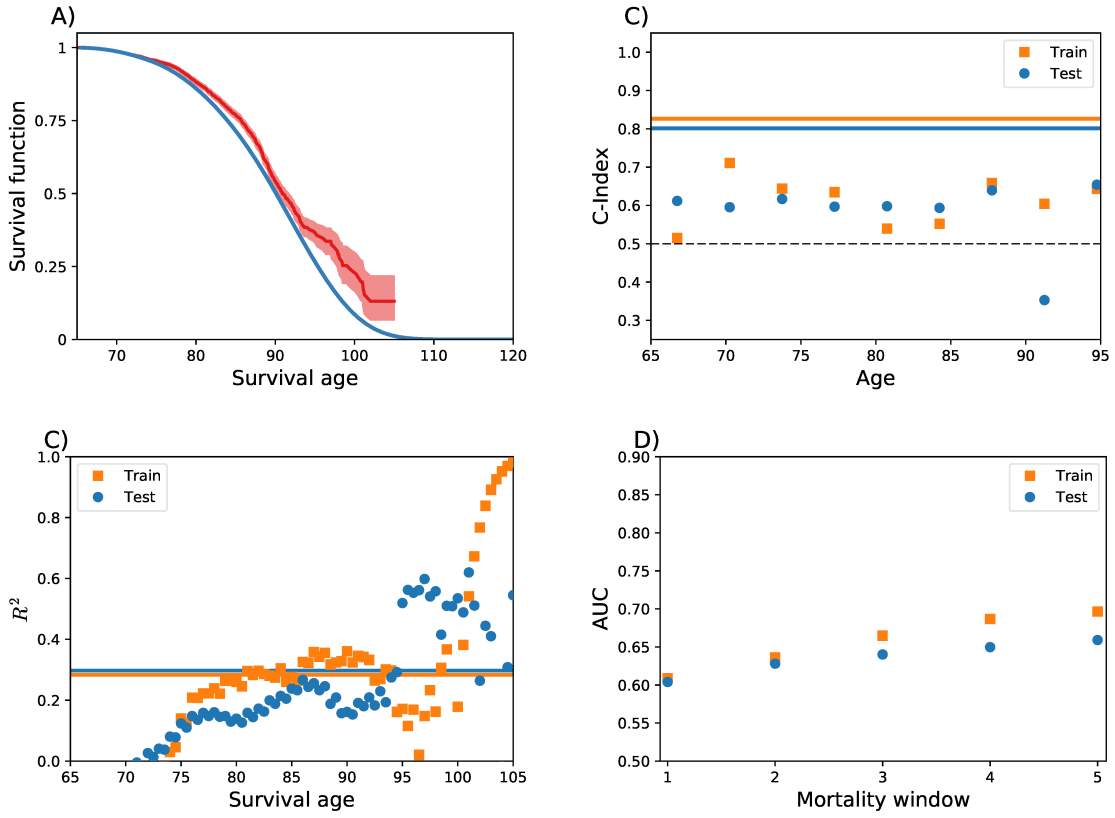


Figure A.10: **Alternate deficits mortality predictions.** Survival predictions with alternative deficits used in Figure A.3 with CSHA data. Compare with Figure A.9. Model predictions work similarly well with these alternate deficits. A) Model population average survival function (blue) and Kaplan-Meier population survival function (red) with a 95% confidence interval. B) Survival C-Index stratified by age (points) and unstratified (solid lines). The baseline for random predictions is $C = 0.5$ (black dashed line). C) Explained variance R^2 from the Brier score. Integrated R^2 is shown as solid lines. Higher values greater than zero mean better predicted survival curves. D) ROC AUC for the prediction of mortality within a window of time (in years) for CSHA data.

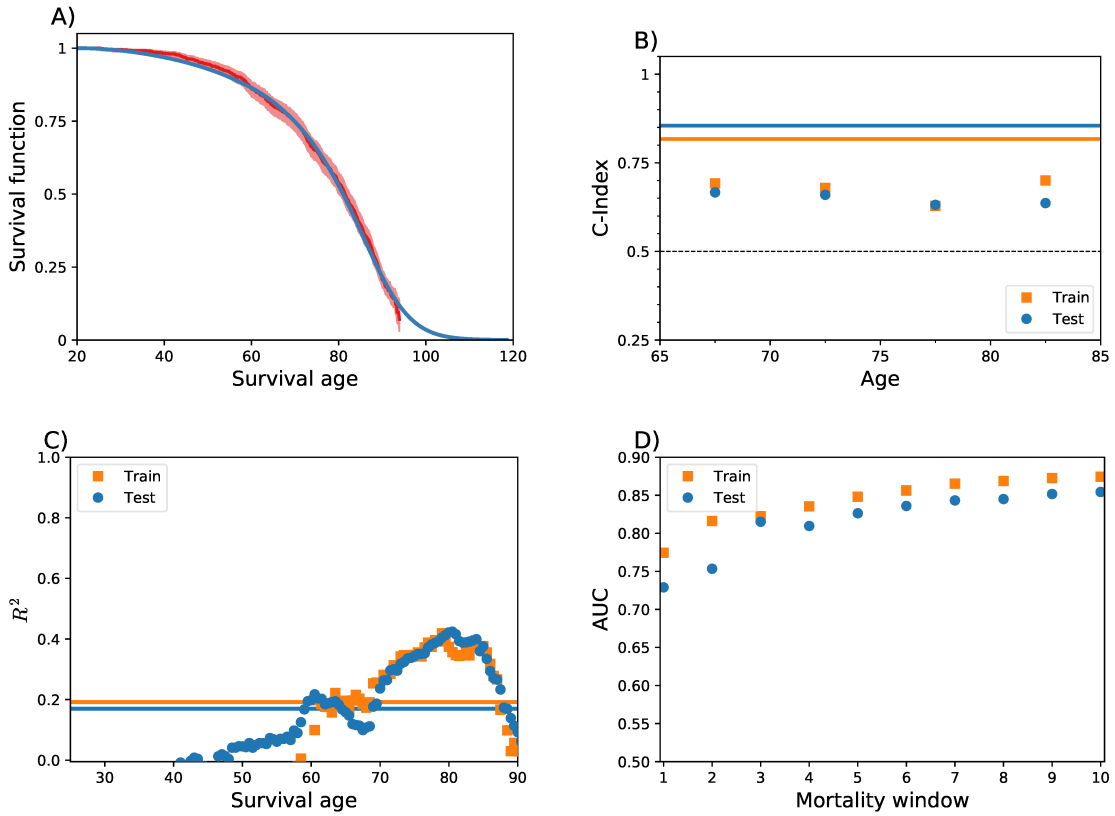


Figure A.11: **NHANES survival predictions** Compare with Figs. A.9 and A.10. A) Model population average survival function (blue) and NHANES Kaplan-Meier population survival function (red) with a 95% confidence interval. B) Survival C-Index stratified by age (orange squares training set, blue circles test set) and unstratified (solid lines). The baseline for random predictions is $C = 0.5$ (dashed black line). C) Explained variance R^2 from the Brier score. Integrated R^2 is shown as solid lines. Higher values greater than zero mean better predicted survival curves. D) ROC AUC for the prediction of mortality within a window of time (in years) for NHANES data.

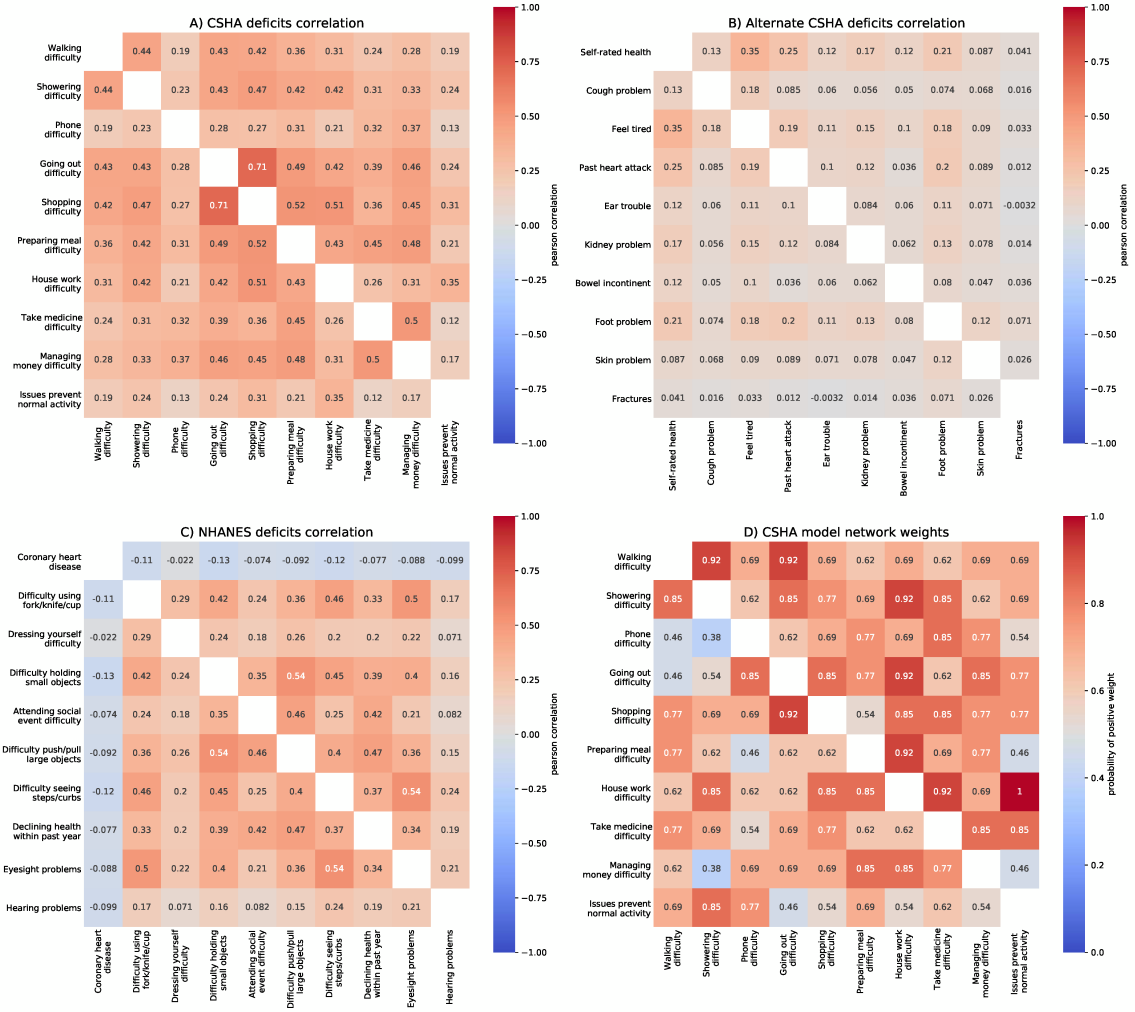


Figure A.12: **Deficits correlation and network weights** A), B) and C) show pairwise correlation for the different sets of deficits used. We see that the correlation structure in all three sets of deficits is quite different. We see stronger positive correlations in the CSHA deficits used for the main plots than the alternative CSHA deficits. We generally see positive correlations for the NHANES deficits, but see a negative correlation with all deficits for the deficit "Coronary heart disease". This deficit also has slightly decreasing prevalence with age, so this is likely due to those individuals with coronary heart disease dying early. In D), we show the probability of an inferred weight in the model being positive, for 13 different model fits for the same deficits as A). The sign of the weights is quite different from the correlations. Probabilities near 0.5 mean that the inferred connection is not very robust.

Appendix B

Chapter 5 supplemental information

B.1 Deriving the variational loss

We denote health variables observed at age t_k by $\mathbf{y}_{t_k}$, the background information at baseline by $\mathbf{u}_{t_0}$, the model health variable predictions by $\mathbf{x}(t_k)$, the latent variables for imputation/generation by $\mathbf{z}$, the age of death or last censoring age by a, the censoring indicator by c, parameters by $\boldsymbol{\theta}$, and variational parameters by $\boldsymbol{\phi}$.

To fit the model, we minimize the KL-divergence between the approximate posterior and the true posterior. This is equivalent to maximizing a lower bound to the model evidence. Starting with the model evidence,

$$
\begin{aligned}
& \log p(\{\mathbf{y}_{t_k}\}_k|\mathbf{u}_{t_0}, \{\mathbf{o}_{t_k}\}_k, a, c) \qquad \text{(B.1)} \\
= \; & \log \int d\boldsymbol{\theta} d\mathbf{z} d\mathbf{x}_0 dt \, p(\boldsymbol{\theta}) p(\mathbf{z}) p(\mathbf{x}_0|\mathbf{z}, \mathbf{u}_{t_0}, t_0) p(\mathbf{x}(t)|\mathbf{x}_0, \mathbf{u}_{t_0}, t) p(a, c|\mathbf{x}(t), \mathbf{u}_{t_0}, t) \\
& \times \prod_{k=0}^{K} p(\mathbf{y}_{t_k}|\mathbf{x}(t_k), \mathbf{o}_{t_k}) \\
= \; & \log \int d\boldsymbol{\theta} d\mathbf{z} \, p(\boldsymbol{\theta}) p(\mathbf{z}) \int d\mathbf{x}_0 \, p(\mathbf{x}_0|\mathbf{z}, \mathbf{u}_{t_0}, t_0) \int dt \, p(\mathbf{x}(t)|\mathbf{x}_0, \mathbf{u}_{t_0}, t) p(a, c|\mathbf{x}(t), \mathbf{u}_{t_0}, t) \\
& \times \prod_{k=0}^{K} p(\mathbf{y}_{t_k}|\mathbf{x}(t_k), \mathbf{o}_{t_k}) \\
= \; & \log \int d\boldsymbol{\theta} d\mathbf{z} \, p(\boldsymbol{\theta}) p(\mathbf{z}) \frac{q(\mathbf{z}|\mathbf{y}_{t_0}, \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, t_0) q(\boldsymbol{\theta})}{q(\mathbf{z}|\mathbf{y}_{t_0}, \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, t_0) q(\boldsymbol{\theta})} \int d\mathbf{x}_0 \, p(\mathbf{x}_0|\mathbf{z}, \mathbf{u}_{t_0}, t_0) \int dt \, p(\mathbf{x}(t)|\mathbf{x}_0, \mathbf{u}_{t_0}, t) \\
& \times \frac{q(\mathbf{x}(t)|\mathbf{x}_0, \mathbf{u}_{t_0}, t)}{q(\mathbf{x}(t)|\mathbf{x}_0, \mathbf{u}_{t_0}, t)} p(a, c|\mathbf{x}(t), \mathbf{u}_{t_0}, t) \prod_{k=0}^{K} p(\mathbf{y}_{t_k}|\mathbf{x}(t_k), \mathbf{o}_{t_k}) \\
= \; & \log \mathbb{E}_{\mathbf{z}, \boldsymbol{\theta} \sim q} \Big[\frac{p(\mathbf{z})}{q(\mathbf{z}|\mathbf{y}_{t_0}, \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, t_0)} \mathbb{E}_{\mathbf{x}_0|\mathbf{z} \sim p, \mathbf{x}(t)|\mathbf{x}_0 \sim q} \Big[\int_{t_0}^{a} \frac{p(\mathbf{x}(t)|\mathbf{x}_0, \mathbf{u}_{t_0}, t)}{q(\mathbf{x}(t)|\mathbf{x}_0, \mathbf{u}_{t_0}, t)} p(a, c|\mathbf{x}(t), \mathbf{u}_{t_0}, t) dt \\
& \times \prod_{k=0}^{K} p(\mathbf{y}_{t_k}|\mathbf{x}(t_k), \mathbf{o}_{t_k}) \Big] \Big],
\end{aligned}
$$

where we have introduced the approximate posteriors q. Using Jensen's Inequality we move the logarithm into the expectations, and define this lower bound as the

objective function,

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\phi}) = \mathbb{E}_{\boldsymbol{\theta},\mathbf{z}\sim q,\ \mathbf{x}_0|\mathbf{z}\sim p,\ \mathbf{x}(t)|\mathbf{x}_0\sim q}\Bigg[\\
&\int_{t_0}^{a} \log \frac{p(\boldsymbol{\theta})p(\mathbf{z})p(\mathbf{x}(t)|\mathbf{x}_0,\mathbf{u}_{t_0},t)p(\mathbf{y}_t|\mathbf{x}(t),\mathbf{o}_{t_k})p(a,c|\mathbf{x}(t),\mathbf{u}_{t_0},t_0)}{q(\mathbf{z}|\mathbf{y}_{t_0},\mathbf{u}_{t_0},\mathbf{o}_{t_0},t_0)q(\boldsymbol{\theta})q(\mathbf{x}(t)|\mathbf{x}_0,\mathbf{u}_{t_0},t)} dt\Bigg] \\
= \ & \mathbb{E}\Bigg[\sum_k \log p(\mathbf{y}_{t_k}|\mathbf{x}(t_k),\mathbf{o}_{t_k}) \\
& + \int_{t_0}^{a} \Big\{ \log p(a,c|\mathbf{x}(t),\mathbf{u}_{t_0},t) + \log p(\mathbf{x}(t)|\mathbf{x}_0,\mathbf{u}_{t_0},t) - \log q(\mathbf{x}(t)|\mathbf{x}_0,\mathbf{u}_{t_0},t)\Big\} dt\Bigg] \\
& - KL(q(\boldsymbol{\theta})||p(\boldsymbol{\theta})) - KL(q(\mathbf{z}|\mathbf{y}_{t_0},\mathbf{u}_{t_0},\mathbf{o}_{t_0},t_0)||p(\mathbf{z})) \\
= \ & \mathbb{E}\Bigg[\sum_k \mathbf{o}_{t_k} \odot \log \mathcal{N}(\mathbf{y}_{t_k}|\mathbf{x}(t_k),\boldsymbol{\sigma}_{\mathbf{y}}) \\
& + (1-c)\big[\log \lambda(a|\mathbf{x}(t),\mathbf{u}_{t_0},t_0) + \log S(a|\mathbf{x}(t),\mathbf{u}_{t_0},t_0)\big] \\
+ \ & \int_{t_0}^{a} c \log S(t|\mathbf{x}(t),\mathbf{u}_{t_0},t_0) dt \\
+ \ & \int_{a}^{a_{\max}} (1-c)\log\left(1 - S(t|\mathbf{x}(t),\mathbf{u}_{t_0},t_0)\right) dt - \\
& \frac{1}{2}\int_{t_0}^{a} \Big|\Big|\boldsymbol{\sigma}_x^{-1} \odot \big(\mathbf{W}\mathbf{x} + \mathbf{f}(\mathbf{x}(t),\mathbf{u}_{t_0},t) - \mathbf{g}(\mathbf{x}(t),\mathbf{u}_{t_0},t)\big)\Big|\Big|^2 dt\Bigg] \\
- \ & KL(q(\boldsymbol{\theta})||p(\boldsymbol{\theta})) - KL(q(\mathbf{z}|\mathbf{y}_{t_0},\mathbf{u}_{t_0},\mathbf{o}_{t_0},t_0)||p(\mathbf{z})).
\end{aligned} \tag{B.2}$$

Plugging in the normalizing flows for the posterior of $\mathbf{z}$,

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\phi}) \ = \ & \mathbb{E}\Bigg[\sum_k \mathbf{o}_{t_k} \odot \log \mathcal{N}(\mathbf{y}_{t_k}|\mathbf{x}(t_k),\boldsymbol{\sigma}_{\mathbf{y}}) \\
& + (1-c)\big[\log \lambda(a|\mathbf{x}(t),\mathbf{u}_{t_0},t_0) + \log S(a|\mathbf{x}(t),\mathbf{u}_{t_0},t_0)\big] \\
+ \ & \int_{t_0}^{a} c \log S(t|\mathbf{x}(t),\mathbf{u}_{t_0},t_0) dt + \int_{a}^{a_{\max}} (1-c)\log\left(1 - S(t|\mathbf{x}(t),\mathbf{u}_{t_0},t_0)\right) dt \\
- \ & \frac{1}{2}\int_{t_0}^{a} \Big|\Big|\boldsymbol{\sigma}_x^{-1} \odot \big(\mathbf{W}\mathbf{x} + \mathbf{f}(\mathbf{x}(t),\mathbf{u}_{t_0},t) - \mathbf{g}(\mathbf{x}(t),\mathbf{u}_{t_0},t)\big)\Big|\Big|^2 dt\Bigg] \\
- \ & KL(q(\boldsymbol{\theta})||p(\boldsymbol{\theta})) - KL(q(\mathbf{z}^{(0)}|\mathbf{y}_{t_0},\mathbf{u}_{t_0},\mathbf{o}_{t_0},t_0)||p(\mathbf{z}^{(0)})) \\
+ \ & \sum_{l=1}^{L} \log\left|\det \frac{\partial a^{(l)}(\mathbf{z}^{(l)},\boldsymbol{\gamma}_z,\phi_z)}{\partial \mathbf{z}^{(l)}}\right|.
\end{aligned} \tag{B.3}$$

Here we do not show the variational parameters $\boldsymbol{\phi}$ in the notation for the approximate posteriors q and the parameters $\boldsymbol{\theta}$ from the conditional distributions for simplicity.

Additionally, we have averaged over the imputed or generated $\mathbf{x}_0$. This is the objective function used in the methods.

B.2 Alternate models

B.2.1 Full neural network drift function and one-dimensional summary models

We compare our pair-wise interactions network model with two alternate models, where we directly incorporate dynamics for the latent state $\mathbf{z}(t)$ and apply the decoder to estimate the health variables $\mathbf{x}(t)$ at specific ages. With this approach we do not need to impute the baseline state of health variables, or to directly include dynamics for the baseline health. Rather a decoder directly maps the latent states $\mathbf{z}(t)$ to the predicted output of the health variables $\mathbf{y}_t$. This approach has the form,

$$\mathbf{z}_0, \boldsymbol{\theta} \sim p(\mathbf{z}_0)p(\boldsymbol{\theta}) \qquad \text{(Prior)}$$

$$d\mathbf{z}(t) = \mathbf{f}(\mathbf{z}(t), \mathbf{u}_{t_0}, t; \boldsymbol{\theta}_{\boldsymbol{f}}) + \boldsymbol{\sigma}_{\mathbf{z}}(\mathbf{z}(t))d\mathbf{B}(t),\ \mathbf{z}(t_0) = \mathbf{z}_0, \qquad \text{(Dynamics)}$$

$$S(t) = \exp\big(-\int_{t_0}^{t} \lambda(\{\mathbf{z}(\tau)\}_{\tau \le t'}, \mathbf{u}_{t_0}, t'; \boldsymbol{\theta}_\lambda)dt'\big), \qquad \text{(Survival)}$$

$$\mathbf{y}_t \sim \mathcal{N}\Big(\boldsymbol{\psi}^{-1}(\boldsymbol{\mu}(\mathbf{z}(t), \mathbf{u}_{t_0}, \mathbf{o}_{t_0}, \boldsymbol{\theta}_p)), \mathrm{diag}({\sigma_{\mathbf{y}}}^2)\Big), \qquad \text{(Health observation)}$$

$$a \sim \lambda(\{\mathbf{z}(\tau)\}_{\tau \le a}, \mathbf{u}_{t_0}, a; \boldsymbol{\theta}_\lambda)S(a), \qquad \text{(Survival observation)}$$

$$p(\{\mathbf{z}(t)\}_t, \boldsymbol{\theta} | \{\mathbf{y}_{t_k}\}_k, \mathbf{u}_{t_0}, t_0, a, c) \propto p(\boldsymbol{\theta})p(\mathbf{z}_0)p(\{\mathbf{z}(t)\}_t | \mathbf{z}_0, \mathbf{u}_{t_0}, t, \boldsymbol{\theta}) \times \qquad \text{(Inference)}$$
$$p(a, c | \{\mathbf{z}(t)\}_t, \mathbf{u}_{t_0}, t, \boldsymbol{\theta}) \prod_k p(\mathbf{y}_{t_k} | \{\mathbf{z}(t_k)\}_k, \boldsymbol{\theta}),$$

$$\boldsymbol{\theta} = \big\{\mathbf{W}, \sigma_{\mathbf{y}}, \sigma_{\mathbf{x}}, \boldsymbol{\theta}_\lambda, \boldsymbol{\theta}_p, \boldsymbol{\theta}_f\big\}, \qquad \text{(Parameters)}$$

where instead of the variable-wise neural networks in the pair-wise network model, the function $\mathbf{f}$ is now a full neural network including the interactions of all variables. The function $\boldsymbol{\mu}$ is a decoder neural network which outputs the mean of a Gaussian distribution for the health variables $\mathbf{y}_t$, from the latent state at that age.

For the full neural network model, we set the dimension of $\mathbf{z}$ to be $N + 1$. For a second alternative model we use a dimension of 1 – which leads to one-dimensional dynamics for a summary health-variable. For the one-dimensional model, we only use

the sex, ethnicity, and country of birth components of $\mathbf{u}_{t_0}$ for the drift functions $\mathbf{f}$ and $\mathbf{g}$ and the mortality rate λ. This is because the other components contain additional health information that should be included in the one-dimensional summary health measure.

All hyperparameters remain the same from these alternate models and the pairwise network model. In particular, the form of the loss function remains the same, except that the priors for $\mathbf{W}$ are removed, and the form of the drift function in the SDE is adjusted. The neural network parameters for these alternative models are trained with the loss function using the same approach as our primary DJIN model.

B.2.2 Non-recurrent neural network mortality rate

In our network model presented in the main results, we model the mortality rate with a recurrent neural network (RNN). This allows the use of a history of health to compute the mortality rate. We have also tested a model where we instead use a feed-forward neural network taking $\mathbf{x}(t), \mathbf{u}_{t_0}, t$ as input – this allows no memory of previous states to determine mortality. We use the same layer sizes as the recurrent neural network model, and use ELU activations.

B.3 Generated synthetic population

We have made a synthetic population available at `https://zenodo.org/record/4733386`. This population includes 3 million individuals for each baseline age of 65, 75, and 85 years old, for a total of 9 million individuals. The background health state has been generated by sampling based on the age and sex-dependent ELSA population. For binary variables we sample a 0 or 1 based on the observed sex and age-dependent prevalence, for continuous variables we sample from a Gaussian distribution with mean and standard deviation from the observed sex and age-dependent population mean and standard deviation. We set all individuals with no medications.

Using this input, we sample a baseline state for each synthetic individual and simulate their health trajectories for 20 years.

Table B.1: Health variables used from the ELSA dataset. Background variables are only used at the first time-step, as $\mathbf{u}_{t_0}$. Longitudinal variables are predicted in $\mathbf{y}_t$. All variables are z-scored; additional transformations before z-scoring are indicated.

Health variable	**Category**	**Wave type**	**Transformation**
Gait speed (average of 3 measurements, speed over 8 feet, age 60+)	Longitudinal	Self-report	
Dominant hand grip strength (average of 3 measurements)	Longitudinal	Nurse	
Non-dominant hand grip strength (average of 3 measurements)	Longitudinal	Nurse	
ADL score (count from 0-10, see Table B.3)	Longitudinal	Self-report	
IADL score (count from 0-13, see Table B.3)	Longitudinal	Self-report	
Time to rise from a chair 5x	Longitudinal	Nurse	
Time held leg raise (eyes open, maximum 30 secs)	Longitudinal	Nurse	Log-scaled
Time held full tandem stance (maximum 30 secs)	Longitudinal	Nurse	Log-scaled
Self-rated health (scored 0=excellent to 1=poor, 5 levels)	Longitudinal	Self-report	
Eyesight (with aids) (scored 0=excellent 1=legally blind, levels=6)	Longitudinal	Self-report	
Hearing (with aids) (scored 0=excellent to 1=poor, 5 levels)	Longitudinal	Self-report	
Walking ability score (unaided ability to walk 1/4 mile) (scored 0=no difficulty to 1=unable to do this, 4 levels)	Longitudinal	Self-report	
Diastolic blood pressure (average of 3 measurements)	Longitudinal	Nurse	
Systolic blood pressure (average of 3 measurements)	Longitudinal	Nurse	
Pulse (average of 3 measurements)	Longitudinal	Nurse	
Triglycerides	Longitudinal	Nurse	Log-scaled
C-reactive protein	Longitudinal	Nurse	Log-scaled
HDL cholesterol	Longitudinal	Nurse	
LDL cholesterol	Longitudinal	Nurse	
Glucose (fasting)	Longitudinal	Nurse	
Insulin-like growth factor 1	Longitudinal	Nurse	
Hemoglobin	Longitudinal	Nurse	
Fibrinogen	Longitudinal	Nurse	
Ferritin	Longitudinal	Nurse	Log-scaled
Total cholesterol	Longitudinal	Nurse	
White blood cell count	Longitudinal	Nurse	Log-scaled
Mean corpuscular haemoglobin	Longitudinal	Nurse	Log-scaled
Glycated hemoglobin (HgbA1c) (%)	Longitudinal	Nurse	
Vitamin-D	Longitudinal	Nurse	Log-scaled

Table B.2: Background variables used from the ELSA dataset. Background variables are only used at the first time-step, as $\mathbf{u}_{t_0}$. Longitudinal variables are predicted in $\mathbf{y}_t$. All variables are z-scored; additional transformations before z-scoring are indicated.

Background variable	**Category**	**Wave type**
Long-standing illness (yes/no)	Background	Self-report
Long-standing illness limits activities (yes/no)	Background	Self-report
Everything is an effort lately (yes/no)	Background	Self-report
Ever smoked (yes/no)	Background	Self-report
Currently smoke (yes/no)	Background	Self-report
Height	Background	Nurse
Body mass index (weight/height^2)	Background	Nurse
Mobility status (1=walking without help/support, 0=walking requires help/support, bed bound, wheelchair, uncertain impairment)	Background	Nurse
Country of birth (UK/outside UK)	Background	Self-report
Drink alcohol (last 12 months, scored 1=almost every day to 6=every couple of months)	Background	Self-report
Ever had a joint replacement (yes/no)	Background	Self-report
Ever had bone fractures (yes/no)	Background	Self-report
Sex	Background	Self-report
Ethnicity (white/non-white)	Background	Self-report
Hypertension medication (yes/no)	Background	Self-report
Anticoagulant medication (yes/no)	Background	Self-report
Cholesterol medication (yes/no)	Background	Self-report
Hip/knee treatment (medication or exercise, yes/no)	Background	Self-report
Lung/asthma medication (yes/no)	Background	Self-report

Table B.3: Activities of daily living (ADL) and Instrumental activities of daily living (IADL) from the ELSA dataset, for a total of 10 ADL and 13 IADL.

Activities of daily living (ADL)
Walking 100 yards
Sitting for about two hours
Getting up from a chair after sitting for long periods
Climbing several flights of stairs without resting
Climbing one flight of stairs without resting
Stooping, kneeling, or crouching
Reaching or extending arms above shoulder level
Pulling/pushing large objects like a living room chair
Lifting/carrying over 10 lbs, like a heavy bag of groceries
Picking up a 5p coin from a table

Instrumental activities of daily living (IADL)
Dressing, including putting on shoes and socks
Walking across a room
Bathing or showering
Eating, such as cutting up your food
Getting in or out of bed
Using the toilet, including getting up or down
Using a map to get around a strange place
Preparing a hot meal
Shopping for groceries
Making telephone calls
Taking medications
Doing work around the house or garden
Managing money, eg paying bills and keeping track of expenses

B.4 Supplemental figures

In Figure B.1 we show the variables used in the ELSA data set, and the number of individuals for which each variable is observed at each year from the time of entrance to the study. The shaded fills indicate the proportion of observed variables (with respect to the maximum of that variable), with the darkest fill indicating almost 100%. Most variables are unobserved at any given time – which reinforces the need for effective baseline imputation. The full names of these variables are provided in Table B.1.

In Fig. B.2, we show the relative RMSE of our model predictions and the elastic net linear model predictions for each health variable between 1 and 6 years – plotted against the proportion of observations for which the variable is missing in the full dataset. Our model predictions are generally worse for the variables with a higher proportion missing, with observable degradation for proportions of missing around 0.95 where accuracy goes above the population mean predictions, although our model is always better than the elastic net linear model.

In Fig. B.3 we show 3 different example individuals from the test set and the model predicted trajectories. We choose the 6 of the best predicted variables to show. These predictions show the estimated uncertainty for these individual trajectories, and the variety in behaviour in the data for different individuals. The relative RMSE for these individuals averaged over each time point is shown, for comparison with Fig. 2 in the main results.

In Fig. B.4 we show the generated synthetic population Kaplan-Meier survival curve (red line and shading) and the observed population Kaplan-Meier survival curve (blue line and shading) with 95% confidence intervals indicated by the shading. The same censoring distribution seen in the observed population is applied to the synthetic population by sampling censoring ages above the baseline age from the test data with replacement. Agreement is good until $\sim$ 90 years, after which the number of individuals observed in the dataset is very low.

In Fig. B.5 we show the one-dimensional marginal distributions for each health variable for the generated synthetic population and observed population at baseline. We see the synthetic population agrees with the observed population, but often has a slightly lower variance. In Fig. B.6 we show the mean and standard deviation of

generated synthetic population trajectories (red lines and shading) and the observed population trajectories (blue lines and shading). The synthetic trajectories have somewhat lower variance but reasonable agreement with the means.

In Fig. B.7, we contrast our network model's weight matrix with a pair-wise Pearson correlation network, where weights have been pruned with a p-value above 0.01 to match the 99% credible intervals used in our approach. We see many differences. Our weight matrix is much more sparse, including only the links useful for prediction. Our network is also directed and asymmetric, and one-way links between variables are observed, as well as distinct strengths of links in the different directions. However, the sign of the links in the weight matrix is typically the same as in the correlation network.

In Fig. B.8 we show, for each network weight, the posterior mean of the weight vs. the proportion of the approximate posterior distribution that is above zero for posterior weights, or below zero for negative weights. We exclude weights when the probability of the weight being in the opposite direction of the mean is above 1%. This approach only accepts connections with a large probability of having a definite sign. We see that large weights only have a small proportion of the posterior with the opposite sign; showing that the strong connections inferred by the model are robust.

Several alternative models were explored. In Fig. B.9 we summarize predictions for the one-dimensional summary model, in which dynamics are built on one latent summary health variable. This model performs worse than our DJIN model for both health and survival, and is often even worse than a static baseline prediction model (blue squares) for health. In Fig. B.10 we show model results with a full neural network drift function that includes all interactions, in contrast to the linear pair-wise network in our main results with the DJIN model. This shows that the full NN model only does slightly better than the pair-wise network model for health, and is slightly worse for survival. This indicates that the pair-wise network assumptions made by our DJIN model do not sacrifice much accuracy. In Fig. B.11 we show the model results with a feed-forward neural network for the mortality rate instead of a recurrent neural network (GRU). Our recurrent neural network (RNN) model achieves slightly better C-index and Brier scores, particularly for older ages. The models are nearly equivalent for longitudinal prediction.

In Fig. B.12 we show the D-calibration histogram comparison between the DJIN model and the elastic net Cox model. The histograms reflect the χ^2 and p-values given in the main results, showing that the both models have calibrated probabilities.

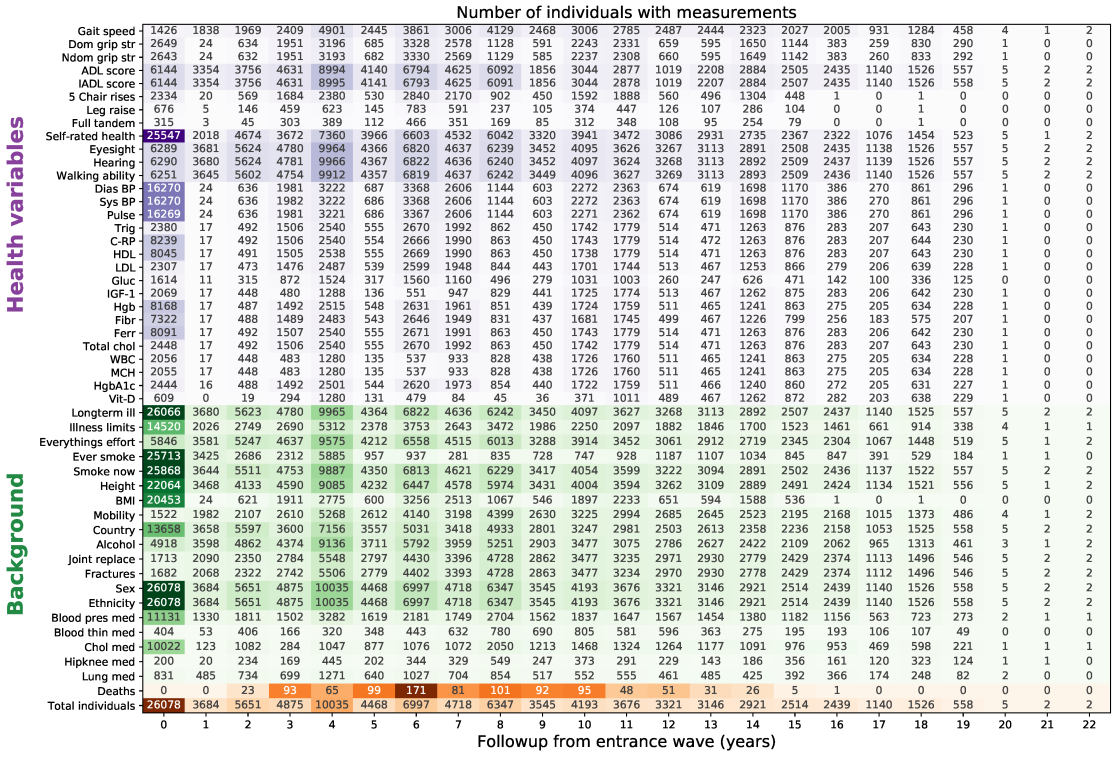


Figure B.1: **Coverage of ELSA dataset.** Number of individuals with measurements vs number of years after entrance to the study. Although ELSA study design has each wave 2 years apart, the age of individuals can change between 1 and 3 years between visits. Health variables (purple shading) are included in $\mathbf{y}_t$. Background variables (green shading) are included in $\mathbf{u}_{t_0}$. Indicated at the bottom (orange shading) are the number of deaths reported, the number of individuals, and the average coverage percentage for those individuals. The darker shading indicates more measurements, relative to the maximum for that variable.

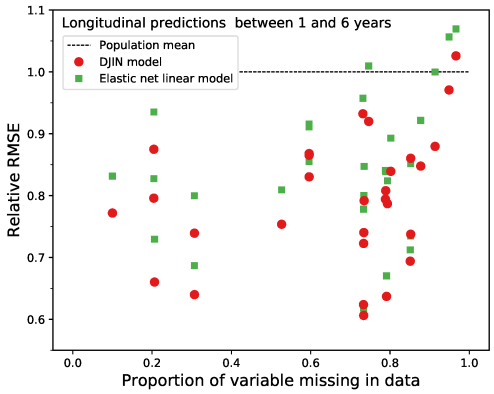


Figure B.2: **Longitudinal predictions vs proportion missing.** Relative RMSE up to six-years past baseline for longitudinal predictions of each health variable plotted against the proportion of observations where that variable was missing. Red circles show our network DJIN model, while green squares show the elastic net linear model. Predictions degrade only at high missingness.

Table B.4: Neural network architectures used in the DJIN model. The health variables $\mathbf{y}_{t_0}$ are size $N = 29$, the health variable observed mask $\mathbf{o}_{t_0}$ is of size $N = 29$, and the background health variables $\mathbf{u}_{t_0}$ with appended missing mask are of size $B + 17 = 36$.

Encoder (VAE)	
Layer #	Description
1	Input $(\mathbf{y}_{t_0}, \mathbf{o}_{t_0}, \mathbf{u}_{t_0}, t_0)$
2	(2N+B+17+1)x95 Fully connected layer
3	Batchnorm
4	ELU
5	95x70 Fully connected layer
6	Batchnorm
7	ELU
8	70x50 Fully connected layer

Decoder (VAE)	
Layer #	Description
1	Input $(\mathbf{z}, \mathbf{u}_{t_0}, t_0)$
2	(20+B+17+1)x65 Fully connected layer
3	Batchnorm
4	ELU
5	65xN Fully connected layer

Diagonal dynamics f_i	
Layer #	Layer description
1	Input $(x_i(t), t, \mathbf{u}_{t_0})$
2	(2+B+17)x12 Fully connected layer
3	ELU
4	12x1 Fully connected layer

Mortality rate λ	
Layer #	Layer description
1	Input $(\mathbf{x}(t), t)$
2	(N+1)x25 GRU
3	25x10 GRU
4	ELU
5	10x1 Linear layer

Posterior drift function g	
Layer #	Layer description
1	Input $(\mathbf{x}(t), \mathbf{u}_{t_0}, t)$
2	(N+B+17+1)x8 Fully connected layer
3	ELU
4	8xN Fully connected layer

Inferring $\mathbf{h}_{t_0}$	
Layer #	Layer description
1	Input $(\mathbf{x}(t_0), \mathbf{u}_{t_0}, t_0)$
2	(N+B+17+1)x75 Fully connected layer
3	ELU
4	75x40 Fully connected layer

Normalizing flow a	
Layer #	Layer description
1	Input $(\mathbf{z}^{(0)}, \boldsymbol{\gamma})$
2	30x24 Fully connected layer
3	BatchNorm
4	Tanh
5	24x20 Fully connected layer

Dynamical noise strength $\boldsymbol{\sigma}_{\mathbf{x}}$	
Layer #	Layer description
1	Input $(\mathbf{x}(t))$
2	NxN Fully connected layer
3	ELU
4	NxN Fully connected layer
5	Sigmoid

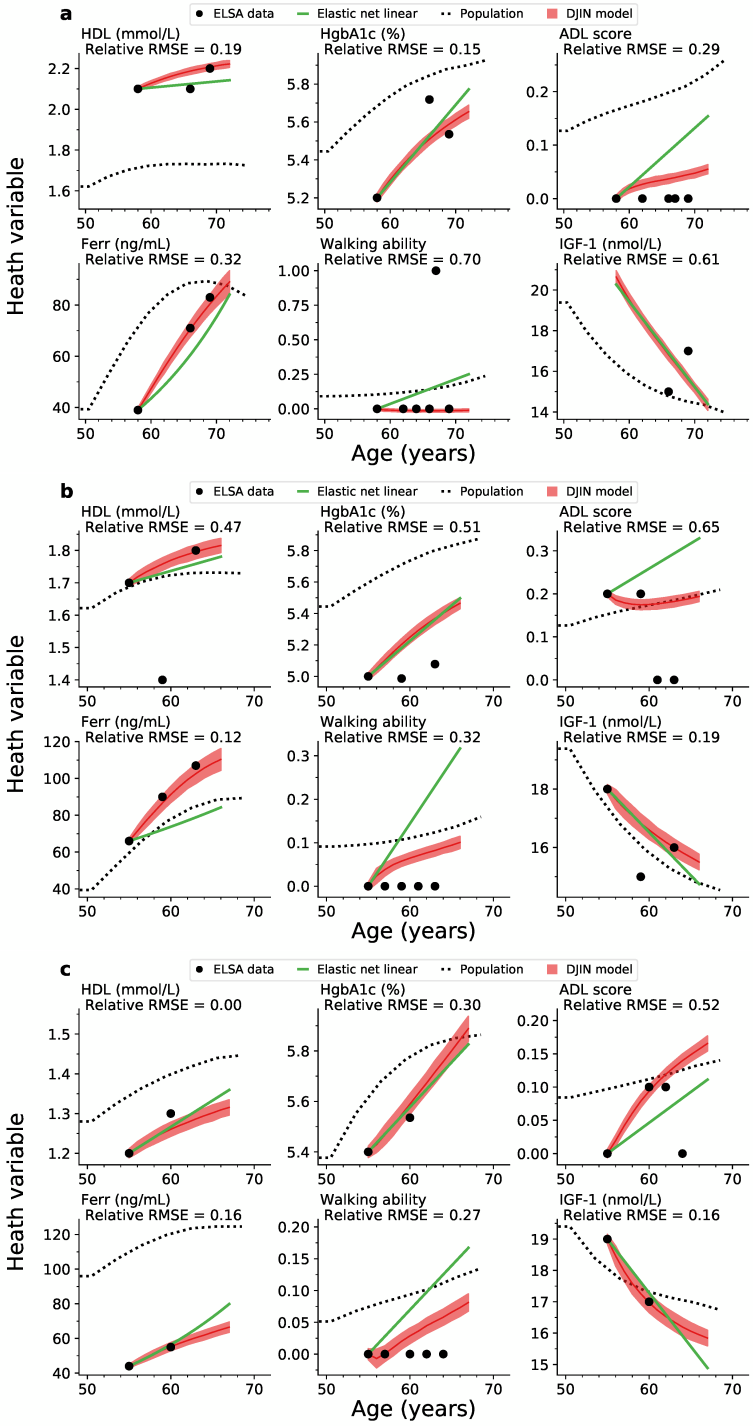


Figure B.3: **Model example trajectories.** We show example predictions for 3 test individuals (**a**, **b**, and **c**). For each individual we show the top 6 best predicted health variables from Figure 2 in the main results. Black circles show the observed ELSA data. Red lines indicate the mean predicted $\mathbf{x}(t)$ and the red shaded region is one standard deviation from the predicted mean trajectory. Green lines indicate the linear model prediction (which appear curved for log-scaled variables such as Ferritin). The average relative RMSE for each variable for each individual is shown.

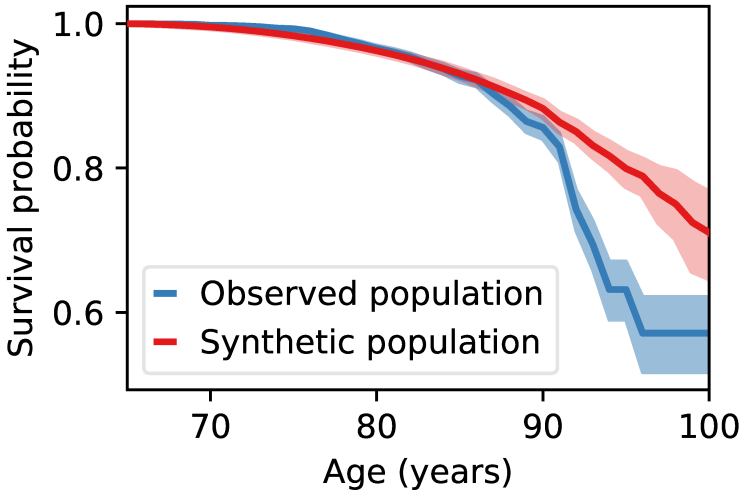


Figure B.4: **Synthetic survival distribution.** Survival curve for synthetic and observed populations, as indicated. The shaded regions show the 95% confidence intervals for Kaplan-Meier curves. The observed population censoring distribution is applied to the synthetic population. The survival probability is approximately the same until 90 years, indicating that the mortality of the synthetic population is representative until older ages.

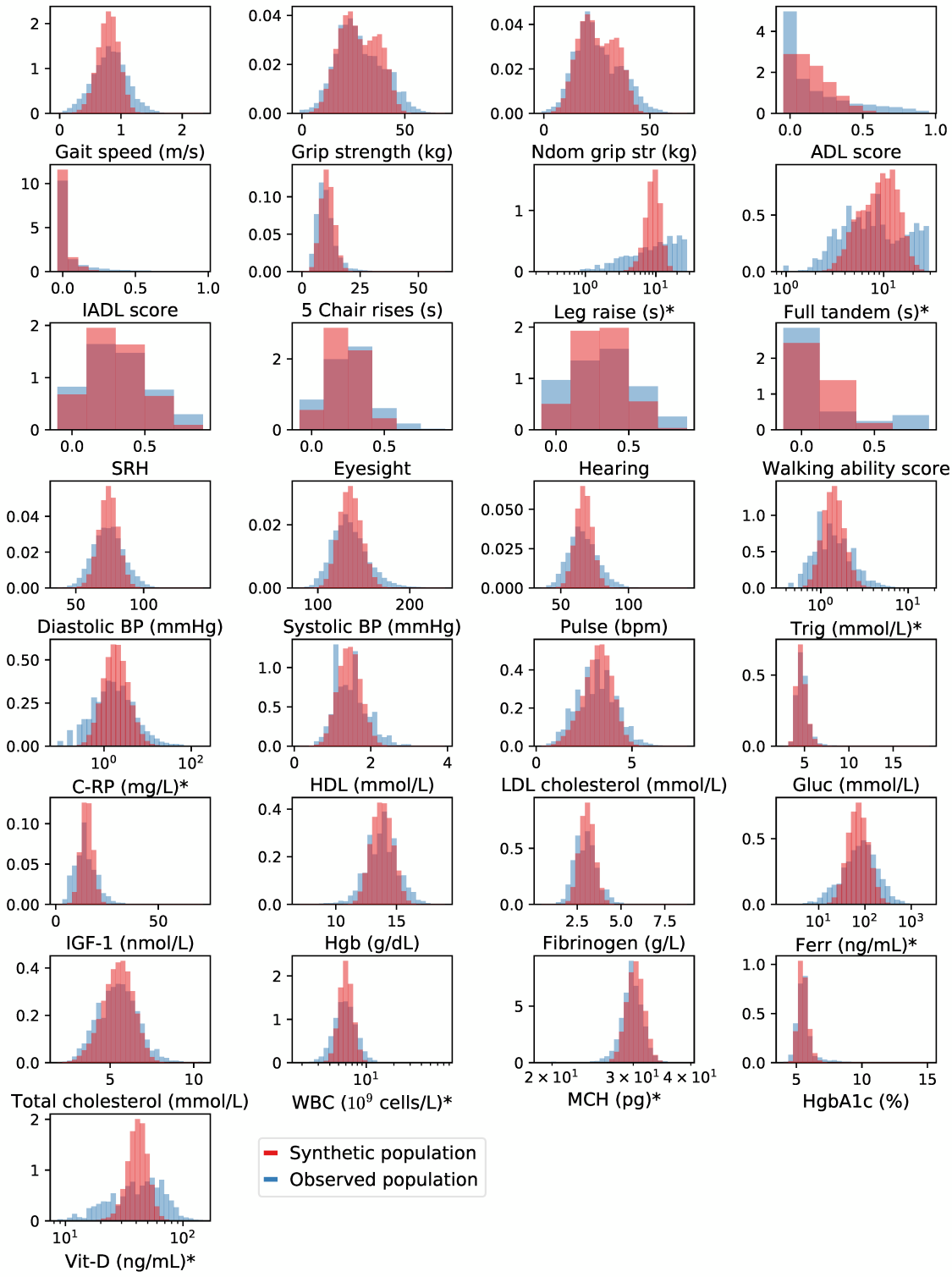


Figure B.5: **Synthetic population baseline distributions.** Each plot shows a synthetic baseline marginal distribution (red shading) for each variable. The synthetic baseline is generated given the background variables $\mathbf{u}_{t_0}$ for the test set. Also shown is the observed population distribution (blue shading). Log-scaled variables are shown with a logarithmic x-axis, and are indicated with an $*$.

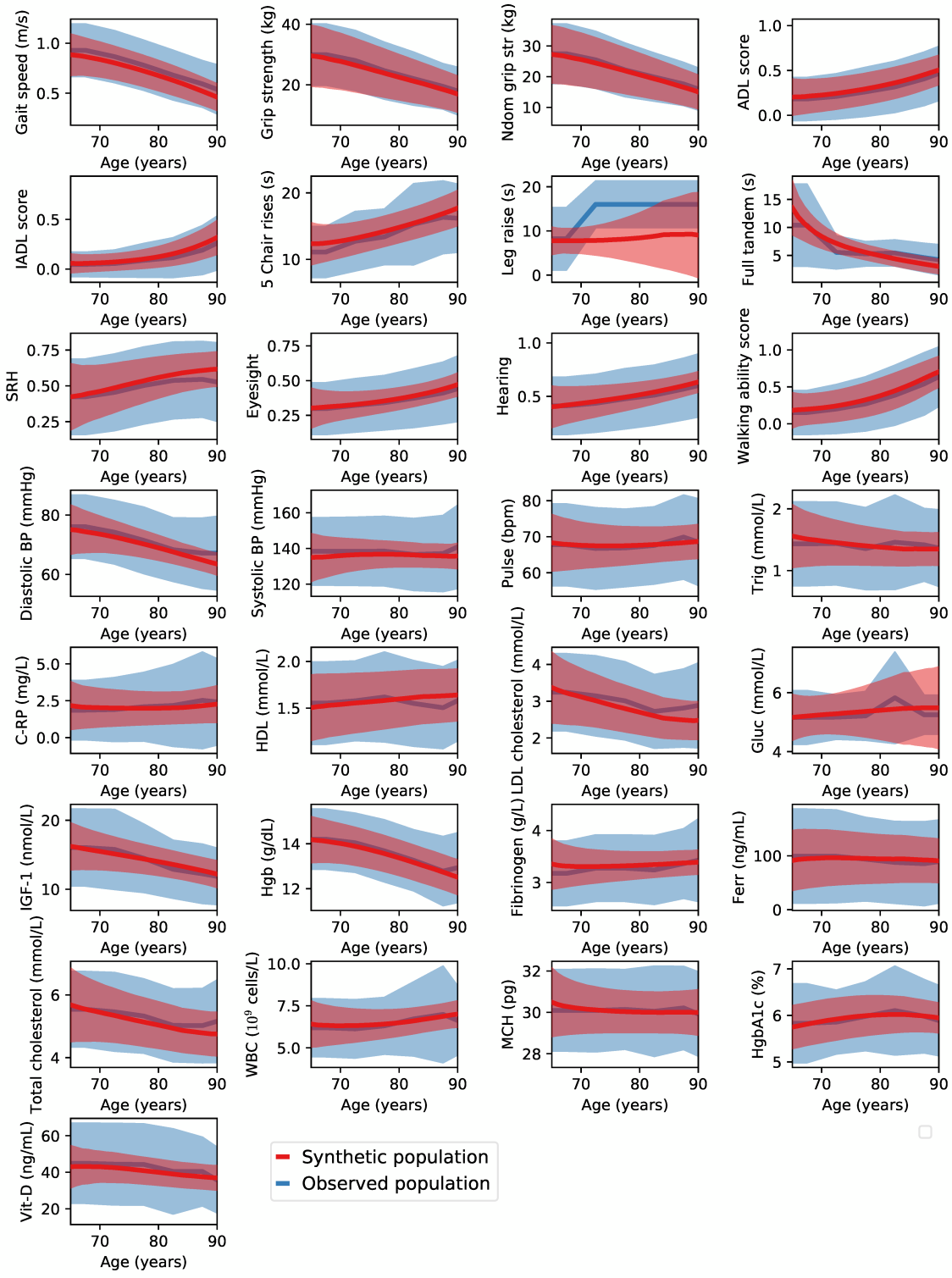


Figure B.6: **Synthetic population trajectories.** Red lines show the synthetic population trajectory marginal distribution means for each variable. Red shaded regions indicate 1 standard deviation away from the mean. Synthetic trajectories are generated from the baseline states shown in Fig. **??**. Blue lines and shaded regions indicate the corresponding means and 1 standard deviation away for the observed population.

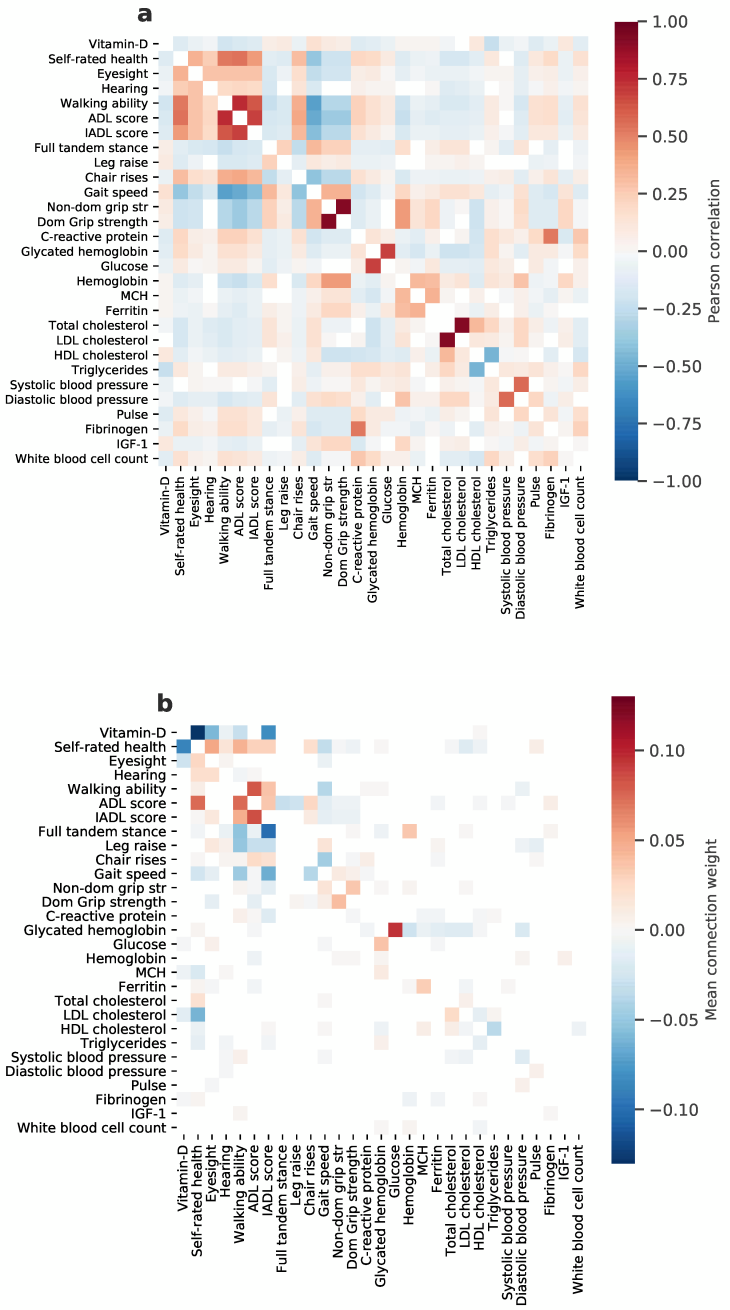


Figure B.7: **Comparison with correlation network. a)** Pearson correlation network between the health variables for all individuals at all time-points, values are pruned for p-values above 0.01. **b)** Our model interaction network, for comparison. Weights are pruned when the 99% posterior credible interval includes zero.

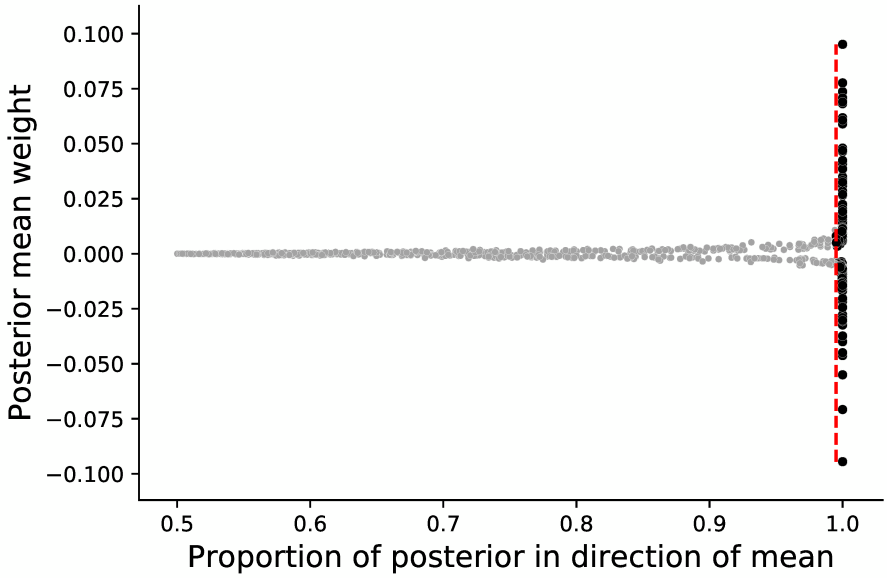


Figure B.8: **Network criterion.** Criteria for determining robust connections. We show the posterior mean of the network weights $\{W_{ij}\}$ vs. the proportion of the posterior above zero for weights with a positive mean, and below zero for weights with a negative mean. The vertical dashed red line shows the criteria for robust connections (shown in Fig. **??**b), which is a 99% credible interval around the mean not containing zero. We see that larger weights are all credible, while many smaller weights are not.

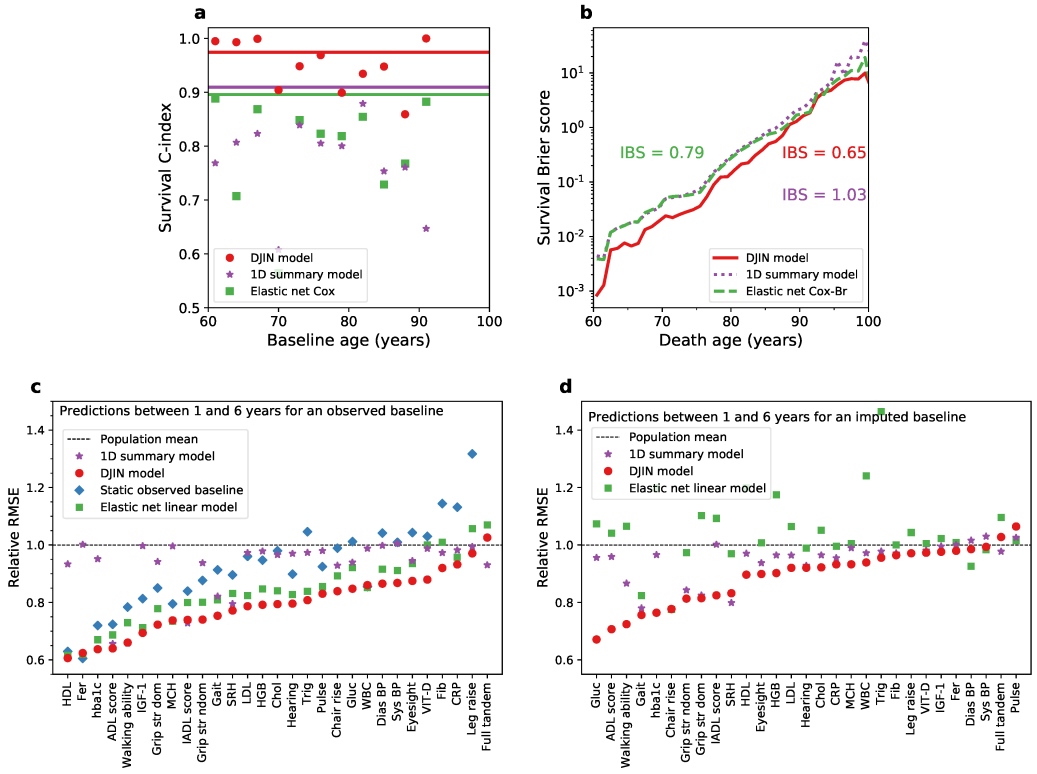


Figure B.9: **One-dimensional summary model.** **a)** Time-dependent C-index stratified vs age (points) and for all ages (line). Results are shown for the 1D summary model (purple), the DJIN network model shown in the main results (red) and a Elastic net Cox model (green). (Higher scores are better). **b)** Brier scores for the survival function vs death age. Integrated Brier scores (IBS) over the full range of death ages is also shown. The Breslow estimator is used for the baseline hazard in the Cox model (Cox-Br). (Lower scores are better). **c)** RMSE scores when the baseline value is observed for each health variable for predictions at least 5 years from baseline, scaled by the RMSE score from the age and sex dependent population mean (relative RMSE scores). We show the predictions from the 1D summary model starting from the baseline value (purple stars), our network model (red circles), predictions assuming a static baseline value (blue diamonds), an elastic-net linear model (green squares). (Lower is better). **d)** Relative RMSE scores when the when the baseline value for each health variable is imputed for predictions past 5 years from baseline. We show the predictions from the 1D summary model starting from the imputed value (purple stars), our network model (red circles), and predictions with an elastic-net linear model (green squares).

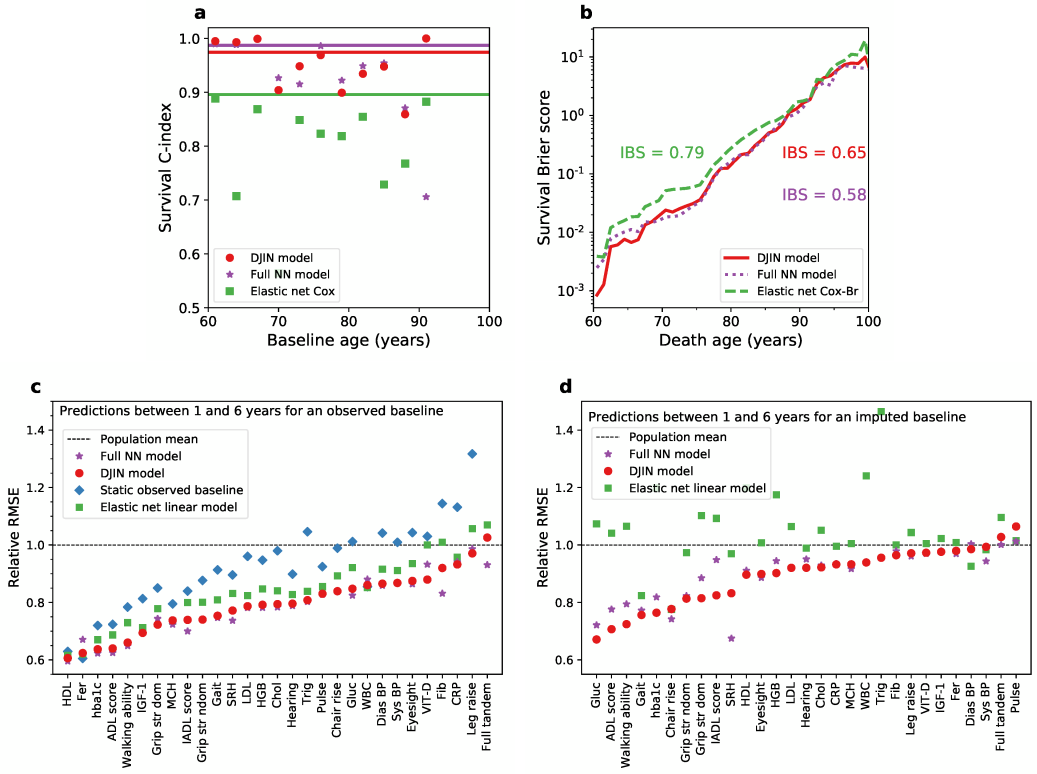


Figure B.10: **Full neural network drift model. a)** Time-dependent C-index stratified vs age (points) and for all ages (line). Results are shown for the full neural network model with a 30-dimensional latent state (purple), the DJIN network model shown in the main results (red) and a Elastic net Cox model (green). (Higher scores are better). **b)** Brier scores for the survival function vs death age. Integrated Brier scores (IBS) over the full range of death ages is also shown. The Breslow estimator is used for the baseline hazard in the Cox model (Cox-Br). (Lower scores are better). **c)** RMSE scores when the baseline value is observed for each health variable for predictions at least 5 years from baseline, scaled by the RMSE score from the age and sex dependent population mean (relative RMSE scores). We show the predictions from the full neural network model starting from the baseline value (purple stars), our network model (red circles), predictions with a static baseline value (blue diamonds), an elastic-net linear model (green squares). (Lower is better). **d)** Relative RMSE scores when the when the baseline value for each health variable is imputed for predictions past 5 years from baseline. We show the predictions from the full neural network model starting from the imputed value (purple stars), our network model (red circles), and predictions with an elastic-net linear model (green squares).

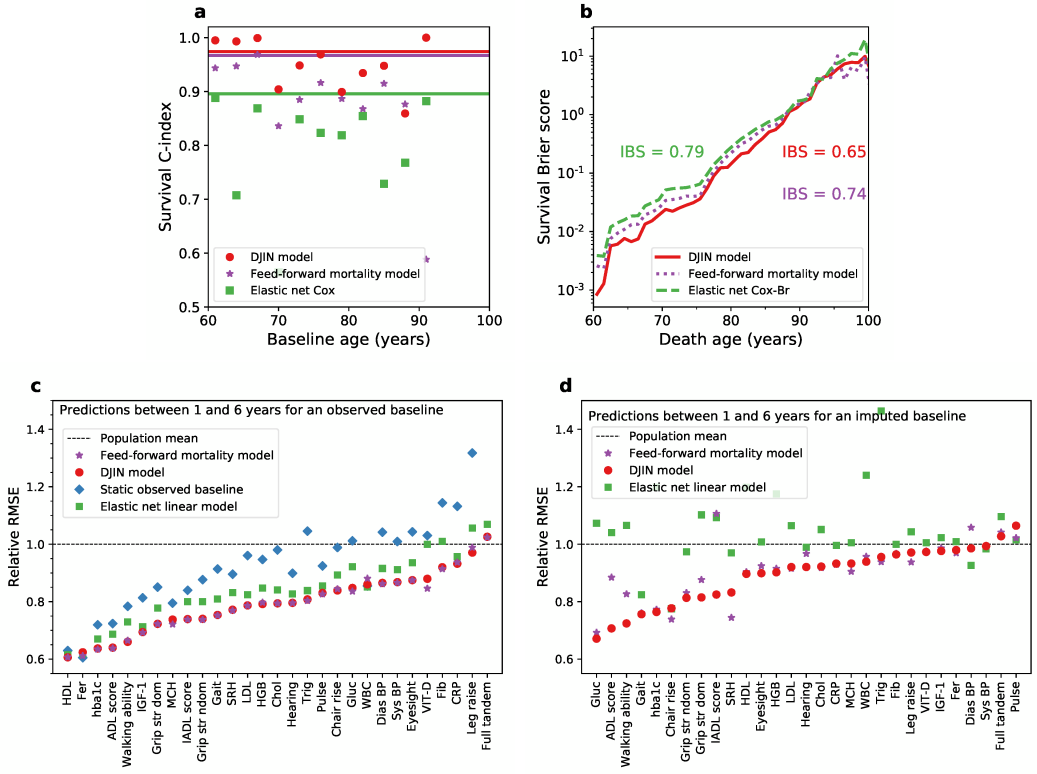


Figure B.11: **Feed-forward mortality rate model.** **a)** Time-dependent C-index stratified vs age (points) and for all ages (line). Results are shown for the feed-mortality mortality rate model (purple), the DJIN network model with a recurrent neural network mortality rate shown in the main results (red) and a Elastic net Cox model (green). (Higher scores are better). **b)** Brier scores for the survival function vs death age. Integrated Brier scores (IBS) over the full range of death ages is also shown. The Breslow estimator is used for the baseline hazard in the Cox model (Cox-Br). (Lower scores are better). Our DJIN model performs better than the feed-forward mortality model. **c)** RMSE scores when the baseline value is observed for each health variable for predictions at least 5 years from baseline, scaled by the RMSE score from the age and sex dependent population mean (relative RMSE scores). We show the predictions from the feed-forward model starting from the baseline value (purple stars), our DJIN model (red circles), predictions assuming a static baseline value (blue diamonds), an elastic-net linear model (green squares). (Lower is better). **d)** Relative RMSE scores when the when the baseline value for each health variable is imputed for predictions past 5 years from baseline. We show the predictions from the feed-forward mortality model starting from the imputed value (purple stars), our DJIN model (red circles), and predictions with an elastic-net linear model (green squares). For longitudinal predictions, the DJIN model is almost equivalent to the feed-forward mortality model.

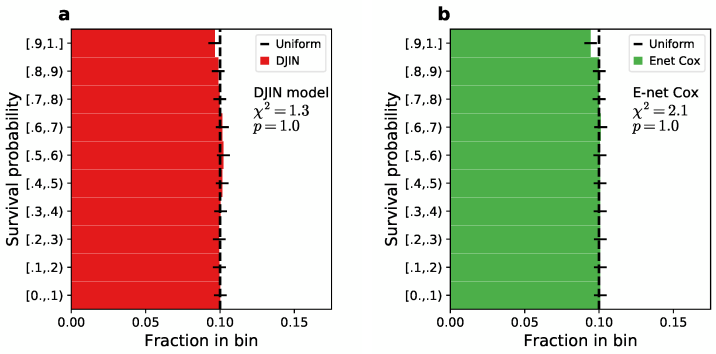


Figure B.12: **D-calibration comparison with elastic-net Cox model. a)** D-calibration of survival predictions for the DJIN model. Estimated survival probabilities are expected to be uniformly distributed (dashed black line). We use Pearson's χ^2 test to assess the distribution of survival probabilities finding $\chi^2 = 1.3$ and $p = 1.0$ and an elastic net Cox model. (Higher p-values and smaller χ^2 statistics are better). **b)** D-calibration of survival predictions for the elastic-net Cox model. Estimated survival probabilities are expected to be uniformly distributed (dashed black line). We use Pearson's χ^2 test to assess the distribution of survival probabilities finding $\chi^2 = 2.1$ and $p = 1.0$. Error bars show the standard deviation.

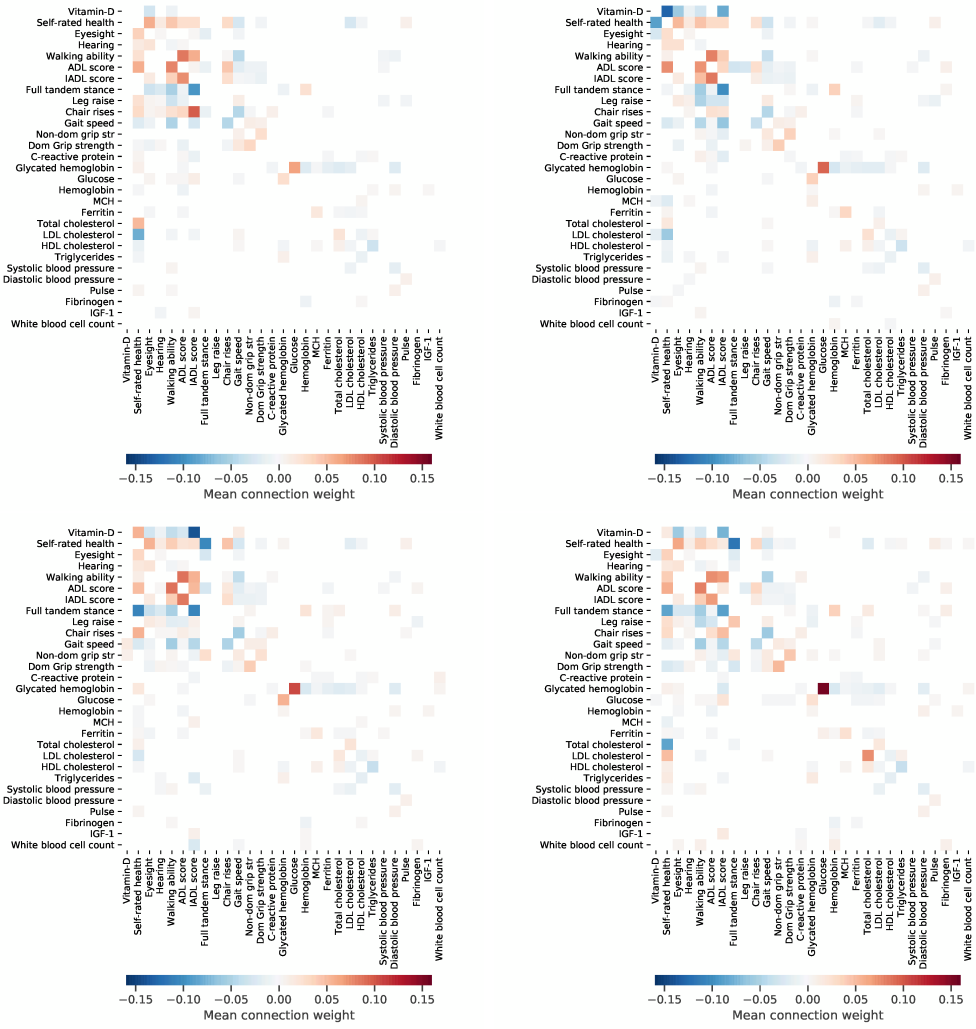


Figure B.13: **Different DJIN network fits.** Inferred network for 4 different fits of the model. These networks visually look very similar, however there are some differences in magnitudes of the connections.

Appendix C

Chapter 7 supplemental information

C.1 Clustering with simulated data

Data is simulated from a toy model where the exact clusters are known to verify that the clustering model works.

In this model, a latent space of size 4 is considered, with 3 clusters. Data is simulated according to the following generative model,

$$
\begin{aligned}
\{\boldsymbol{\mu}\}_c &\sim \mathcal{N}(0,1), && \text{(C.1)}\\
\{\mathbf{A}_c\} &\sim \mathcal{N}(0,0.1^2), && \text{(C.2)}\\
\{\boldsymbol{\Sigma}_c\}_c &= \{\mathbf{A}_c^T\mathbf{A}_c\}_c, && \text{(C.3)}\\
r &\sim \text{Bernoulli}(0.6), && \text{(C.4)}\\
\{\mathbf{W}_c\}_c &\sim r\mathcal{N}(\mathbf{W}_c|0.1,0.1^2)+(1-r)\delta(\mathbf{W}_c), && \text{(C.5)}\\
c &\sim \text{Categorical}(1/3,1/3,1/3), && \text{(C.6)}\\
\mathbf{z}_0|c &\sim \mathcal{N}(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c), && \text{(C.7)}\\
d\mathbf{z}(t) &= \mathbf{W}_c\mathbf{z}dt+0.1d\mathbf{B}(t),\ \mathbf{z}(0)=\mathbf{z}_0, && \text{(C.8)}\\
\boldsymbol{\gamma} &\sim \mathcal{N}(0,1), && \text{(C.9)}\\
h(\mathbf{z}(t),t) &= 0.03e^{0.4t}\exp\left(0.9\boldsymbol{\gamma}\cdot\mathbf{z}(t)\right), && \text{(C.10)}\\
S(\mathbf{z}(t),t) &= \exp\left(-\int_{t_0}^{t}h(\mathbf{z}(t'),t')dt'\right) && \text{(C.11)}\\
\mathbf{w}_1 &\sim \mathcal{N}(0,0.1^2), && \text{(C.12)}\\
\mathbf{w}_2 &\sim \mathcal{N}(0,0.1^2), && \text{(C.13)}\\
\mathbf{b}_1 &\sim \mathcal{N}(-0.25,1), && \text{(C.14)}\\
\mathbf{b}_2 &\sim \mathcal{N}(0,0.1^2), && \text{(C.15)}\\
\boldsymbol{\mu}_x(\mathbf{z}) &= \mathbf{w}_2\text{ReLU}(\mathbf{w}_1\mathbf{z}+\mathbf{b}_1)+\mathbf{b}_2, && \text{(C.16)}\\
\mathbf{x_t}|\mathbf{z}(t) &\sim \mathcal{N}(\boldsymbol{\mu}_x(\mathbf{z}),0.01^2). && \text{(C.17)}
\end{aligned}
$$

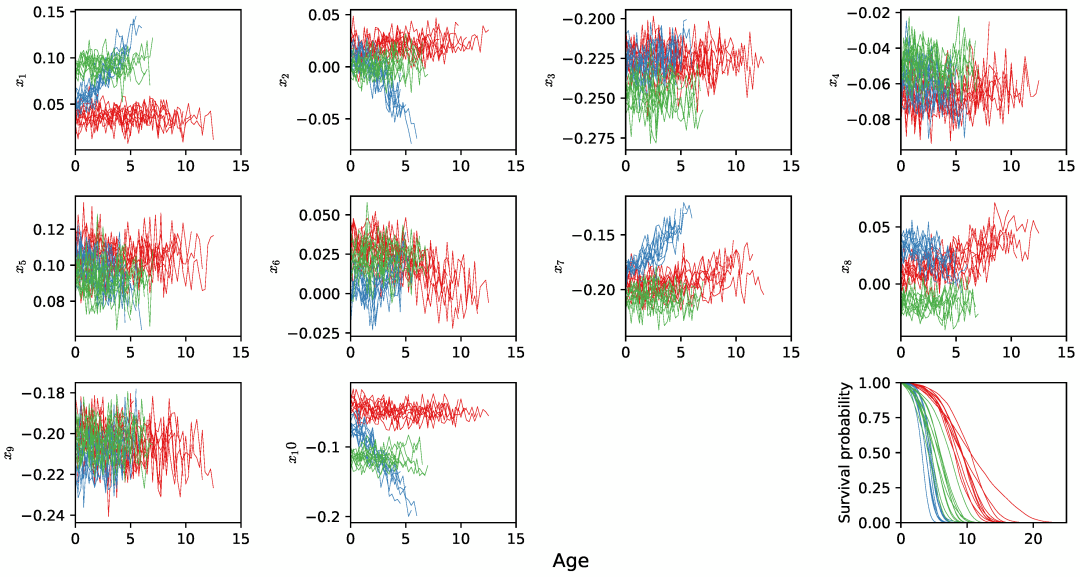


Figure C.1: **Simulated clustering data.**

Figure C.1 shows 10 simulated trajectories and survival curves from this model. 1000 total individuals are simulated from this model, and 800 are used for training, 100 for validation, and 50 for testing. The clustering model remains the same as discussed in Chapter 7, with the same hyperparameters. Figure C.2 shows that the model can accurately infer these clusters.

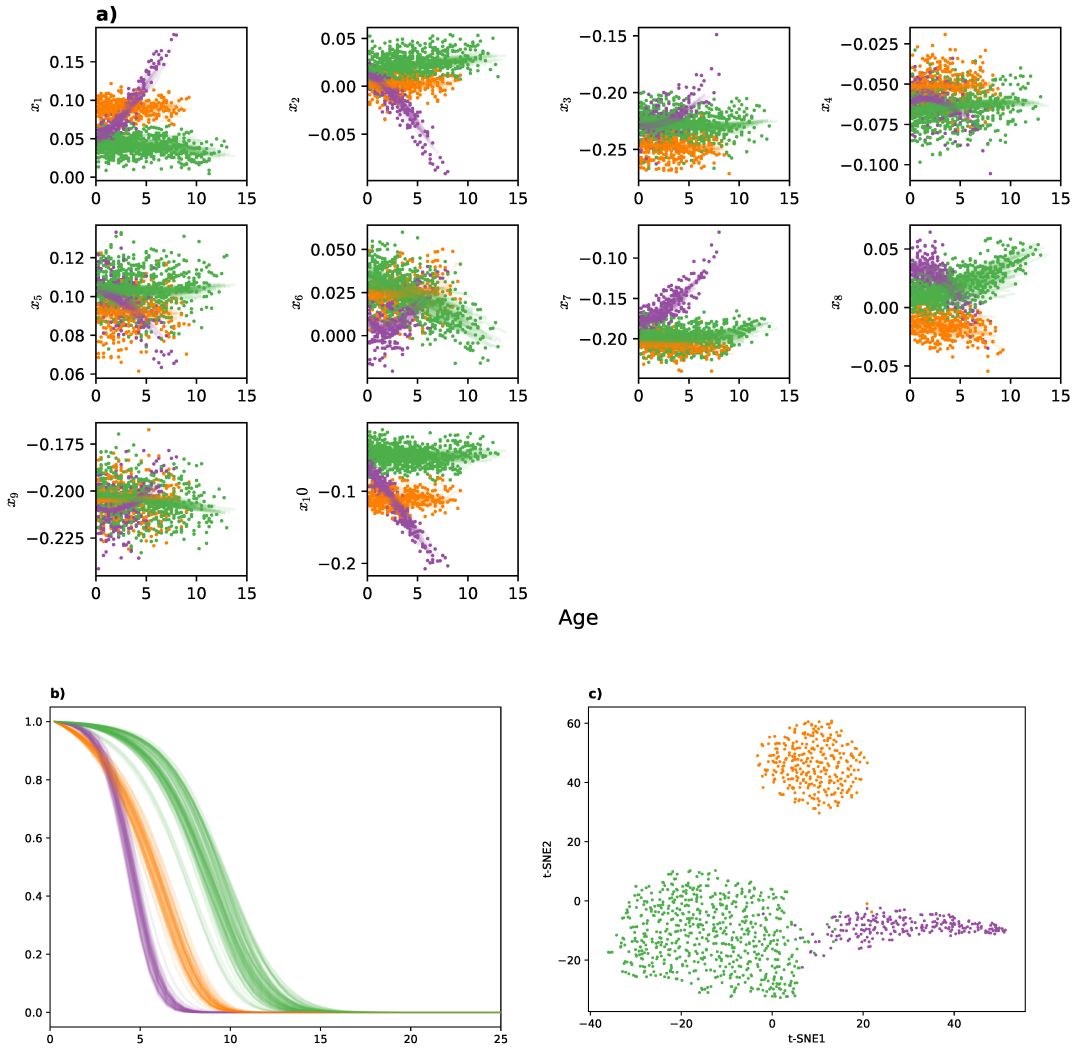


Figure C.2: **Clustering on simulated data.** **a)** Points show the simulated data from the test set, colors show the inferred clusters. Transparent lines show the mean predictions from the model. The distinct trajectory clusters are seen. **b)** Survival curves estimated by the model. **c)** t-SNE dimensional reduction of the observed data, with colors showing inferred clusters. The model cleanly separates these clusters.

Appendix D

Chapter 8 supplemental information

D.1 Supplemental figures

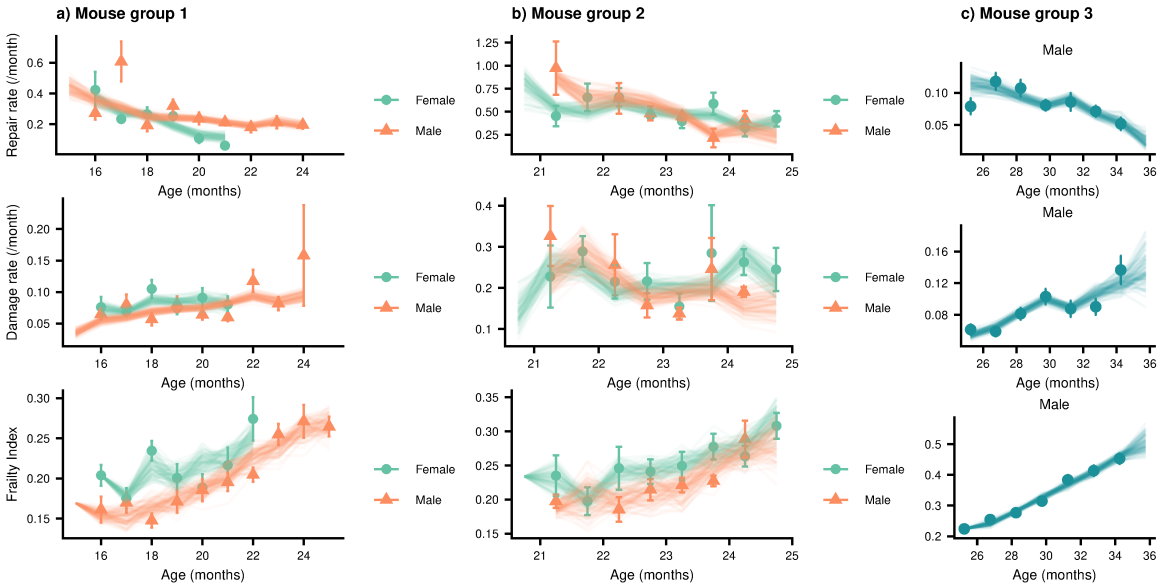


Figure D.1: **Mice control rates sex comparison.** Repair rates, damage rates, and Frailty Index vs age for control mice, with a more direct comparison of the differences with sex when compared to Figure 8.3. In mouse group 1, male mice have higher repair rates and lower damage rates, resulting in a lower Frailty Index. In mouse group 2, damage and repair rates nearly overlap for male/female mice, but the Frailty Index is lower in male mice.

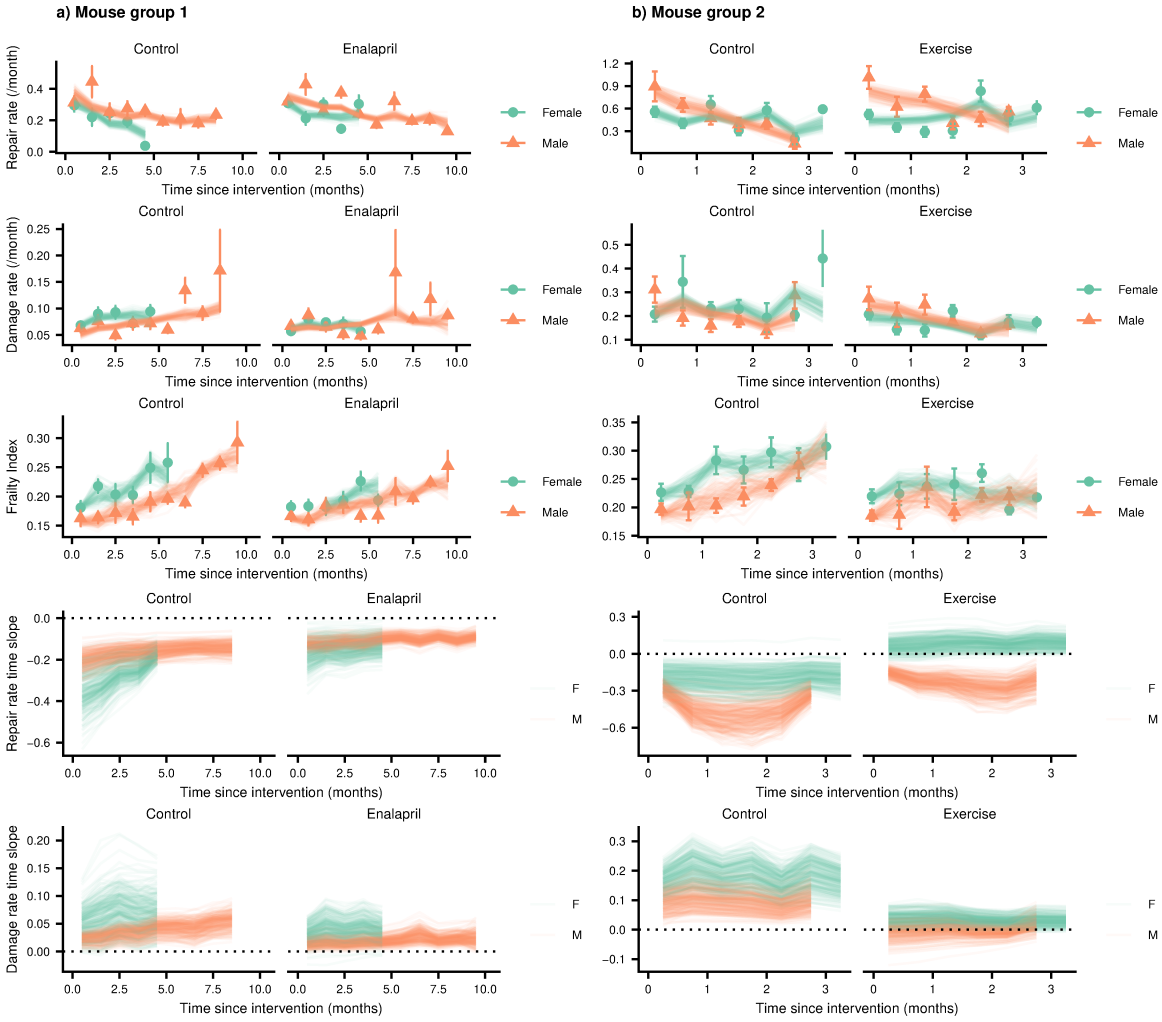


Figure D.2: **Mice interventions rates sex comparison.** Repair rates, damage rates, Frailty Index, repair rate time derivative, and damage rate time derivative vs age for mice with interventions, with a more direct comparison of the differences with sex when compared to Figure 8.4. In mouse group 1, male mice have higher repair rates and lower damage rates, resulting in a lower Frailty Index in both the control and enalapril groups. However, as we saw in Figure 8.4, female mice respond more strongly to the enalapril treatment, resulting in a smaller difference between the male and female mice in the enalapril group. In mouse group 2, we see a similar effect with exercise.

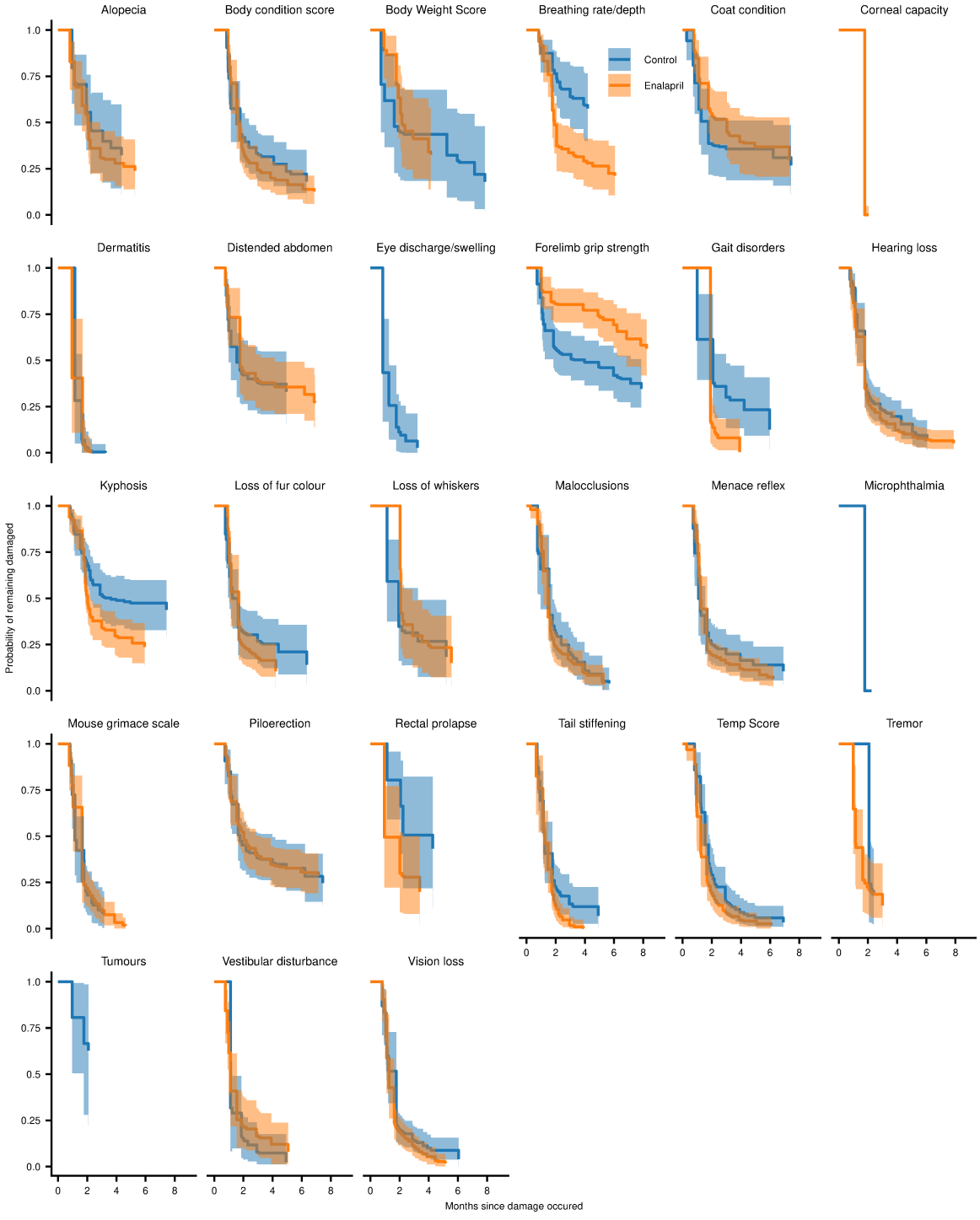


Figure D.3: **Mouse group 1 deficit resilience time-scales.** Probability of remaining damaged vs months since damage occurred (resilience). The colored regions show 95% posterior credible intervals for the survival curves. Survival curves are shown for each deficit.

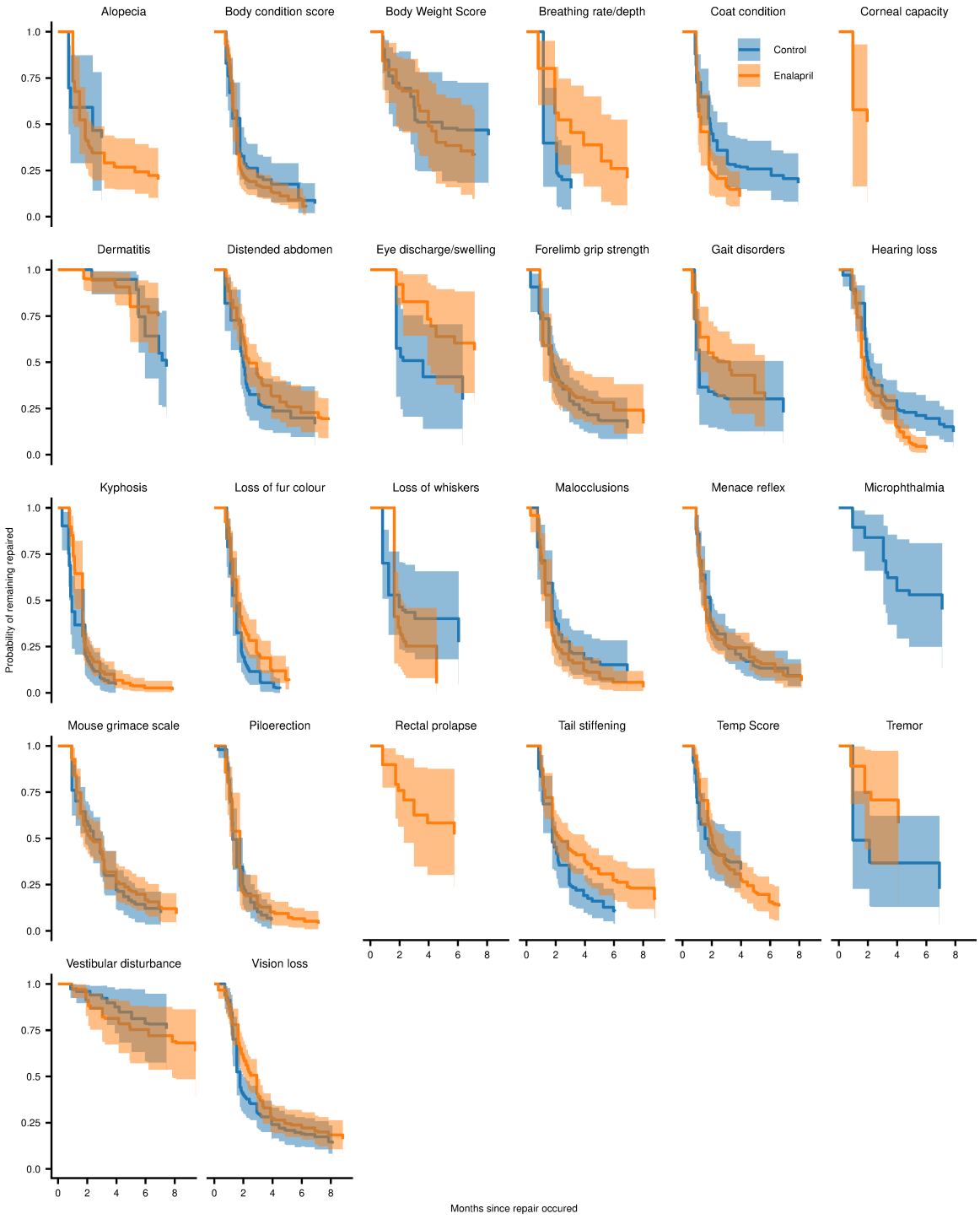


Figure D.4: **Mouse group 1 deficit robustness time-scales.** Probability of remaining undamaged vs months since repair occurred (robustness). The colored regions show 95% posterior credible intervals for the survival curves. Survival curves are shown for each deficit.

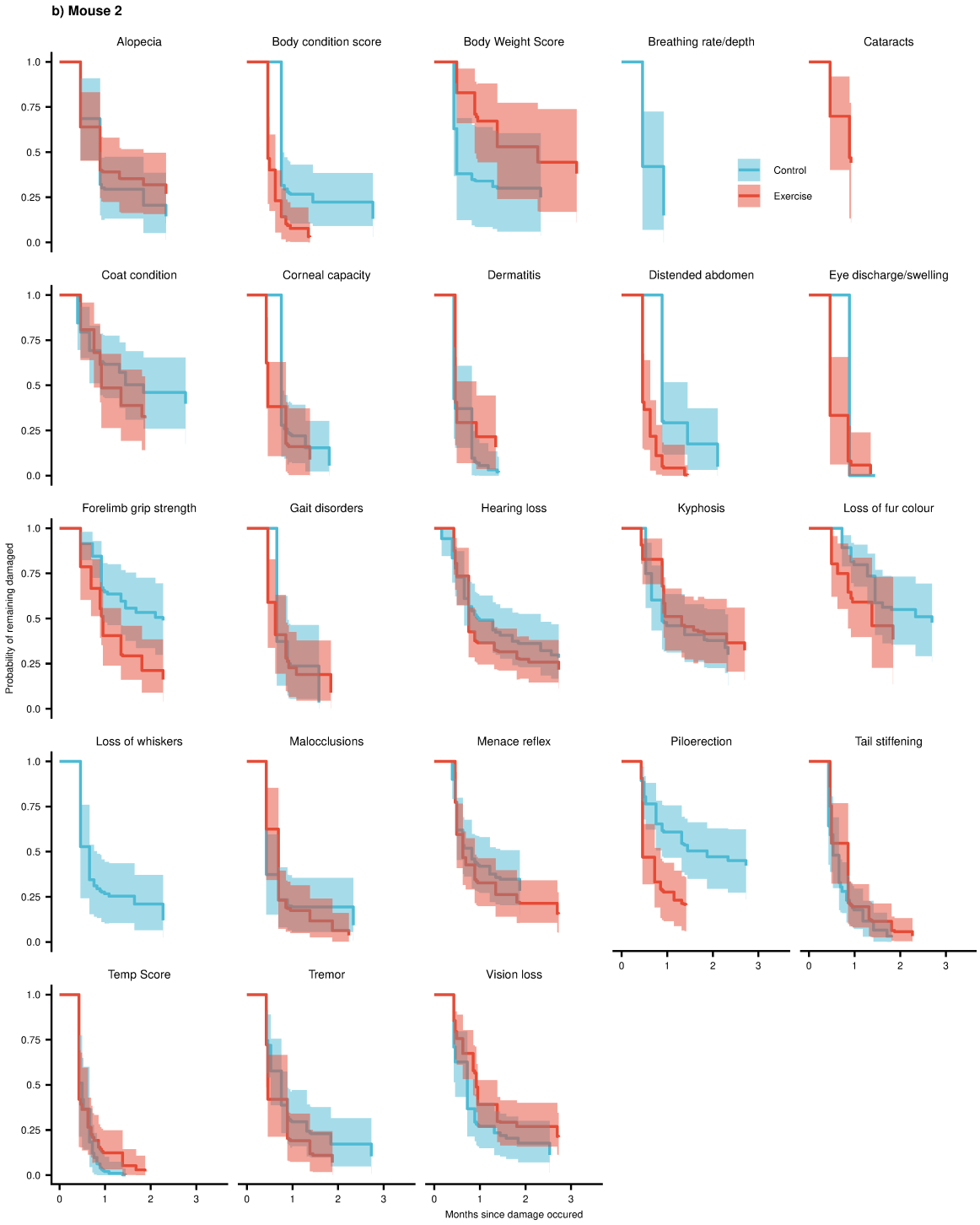


Figure D.5: **Mouse group 2 deficit resilience time-scales.** Probability of remaining damaged vs months since damage occurred (resilience). The colored regions show 95% posterior credible intervals for the survival curves. Survival curves are shown for each deficit.

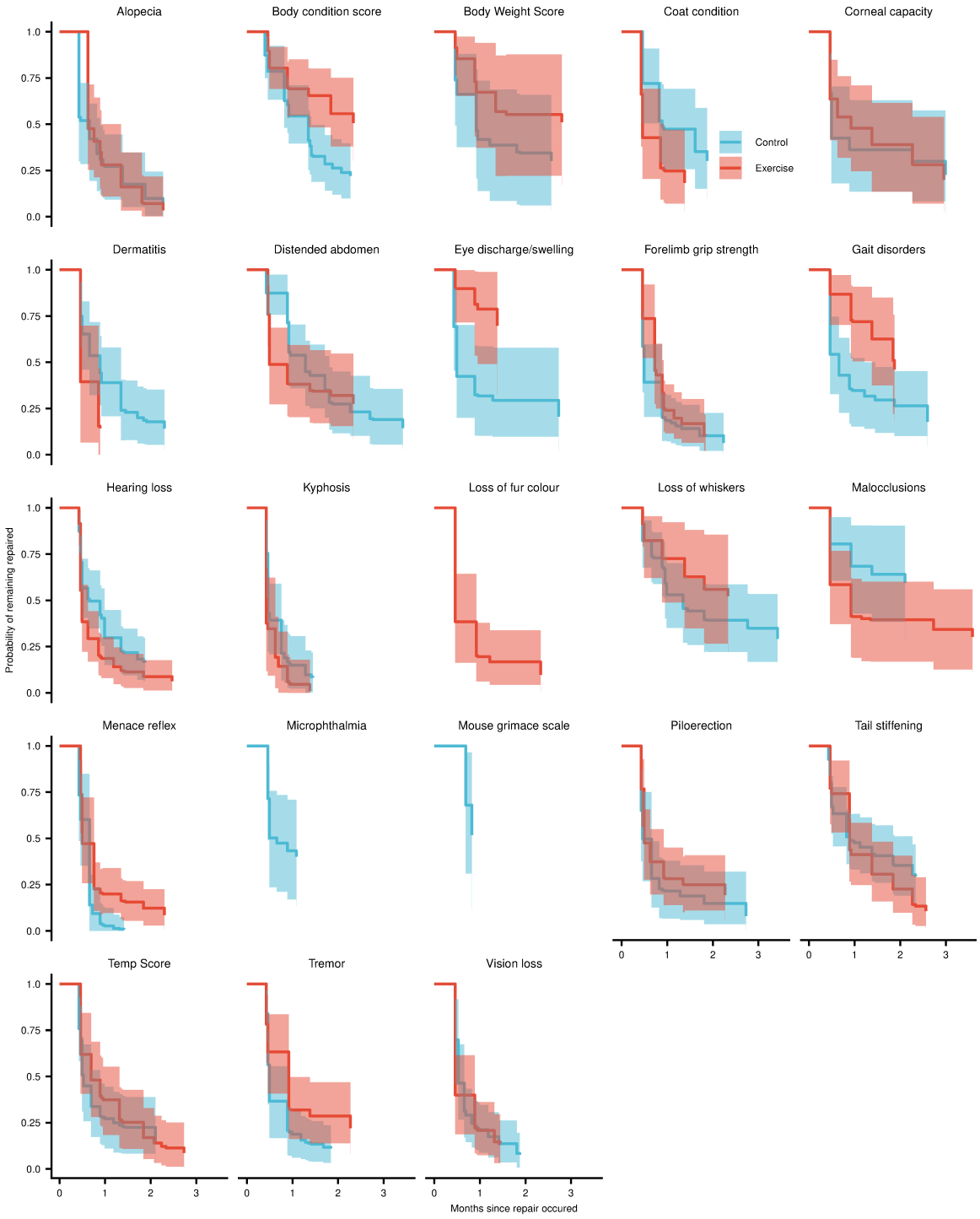


Figure D.6: **Mouse group 2 deficit robustness time-scales.** Probability of remaining undamaged vs months since repair occurred (robustness). The colored regions show 95% posterior credible intervals for the survival curves. Survival curves are shown for each deficit.

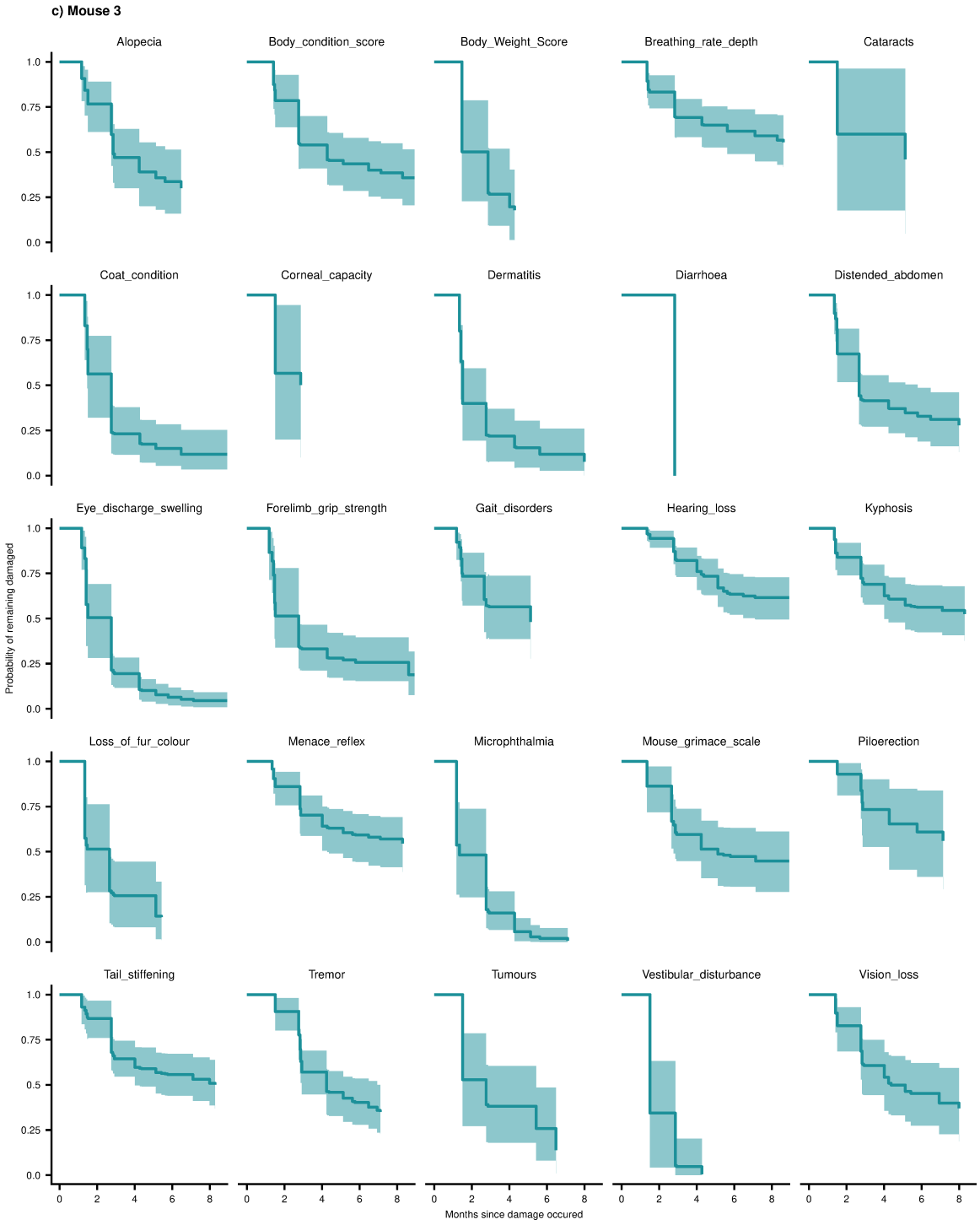


Figure D.7: **Mouse group 3 deficit resilience time-scales.** Probability of remaining damaged vs months since damage occurred (resilience). The colored regions show 95% posterior credible intervals for the survival curves. Survival curves are shown for each deficit.

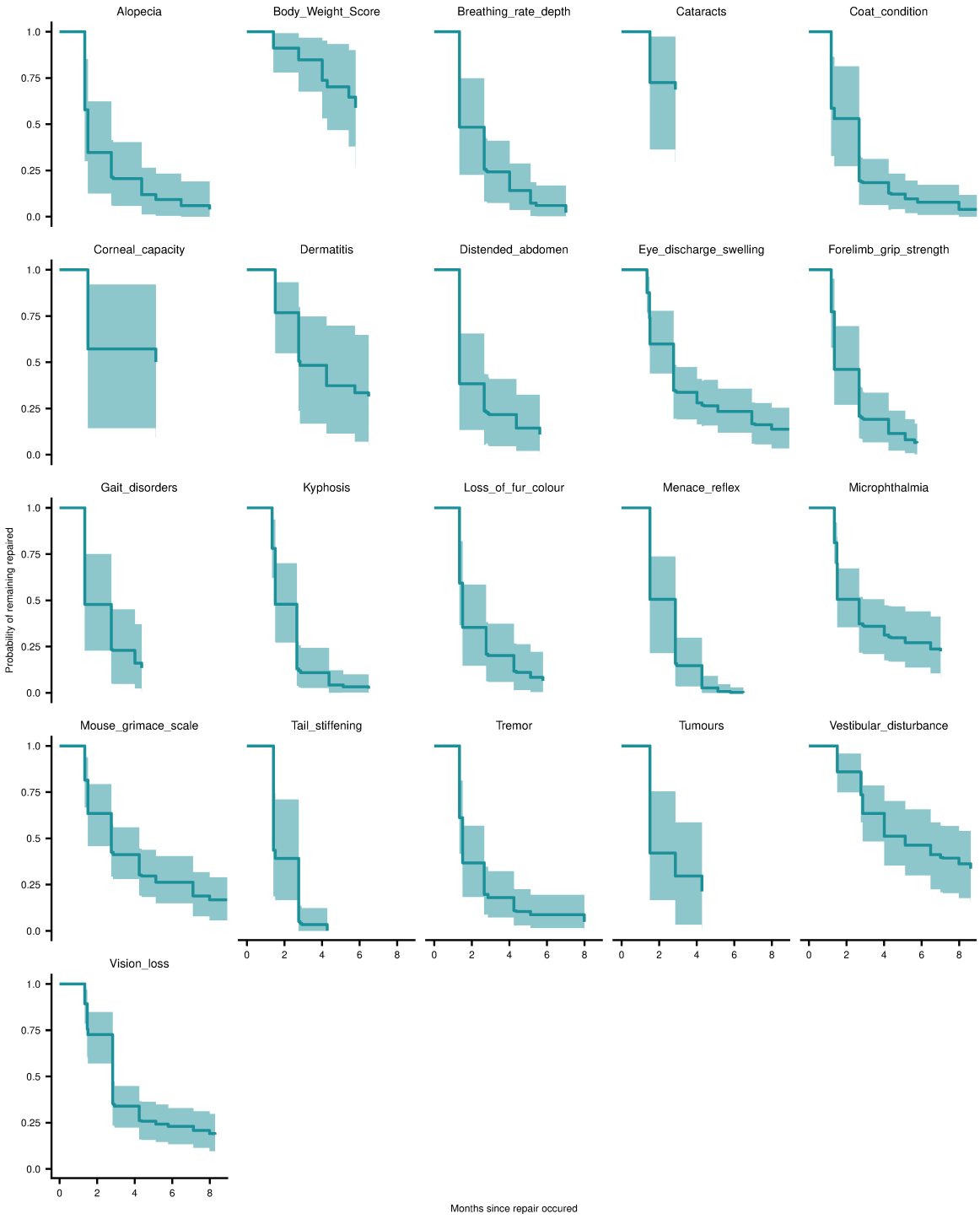


Figure D.8: **Mouse group 3 deficit robustness time-scales.** Probability of remaining undamaged vs months since repair occurred (robustness). The colored regions show 95% posterior credible intervals for the survival curves. Survival curves are shown for each deficit.

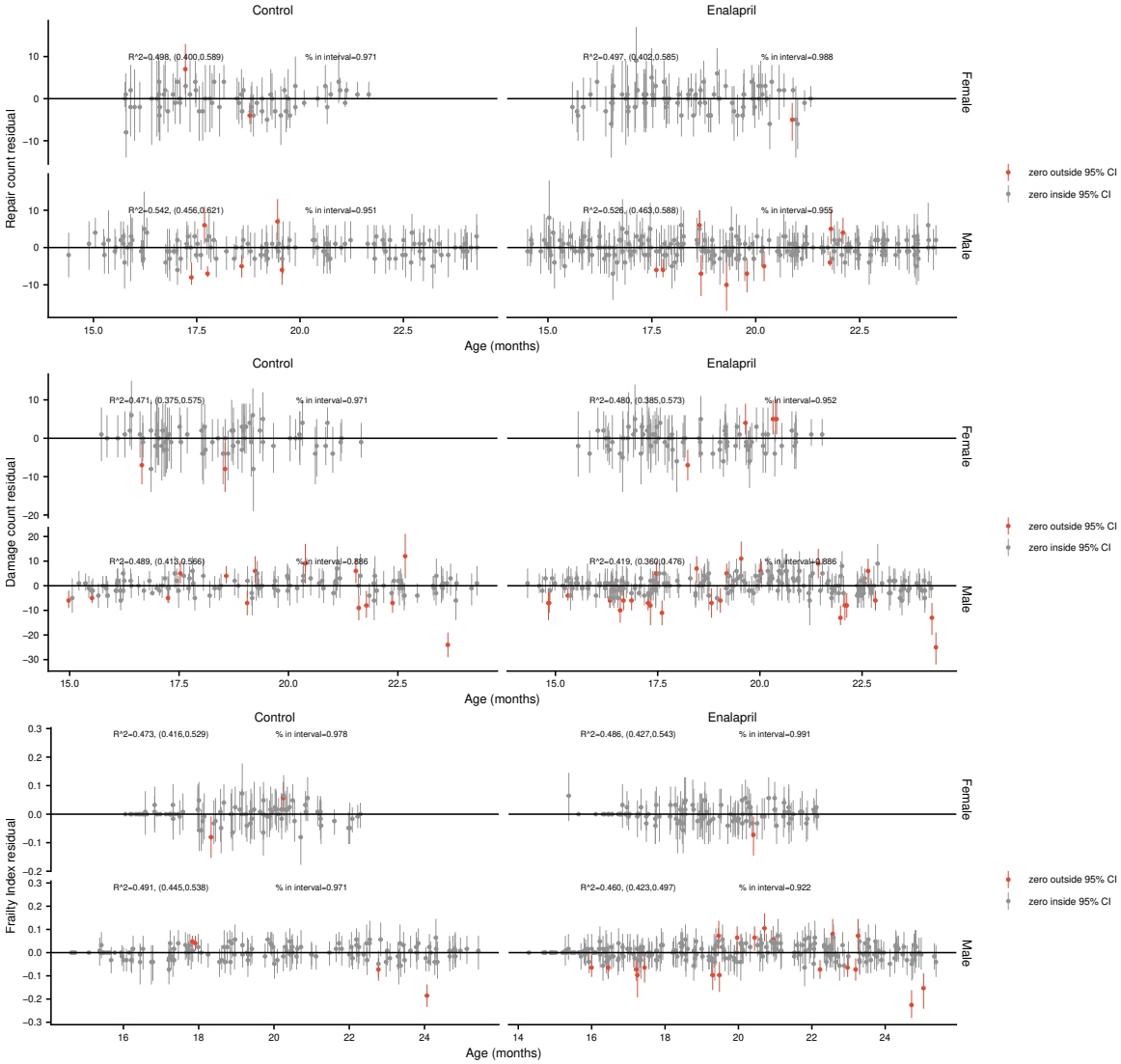


Figure D.9: **Mouse group 1 residuals.** This plot shows the quality of the model fit. Each point shows the median of the difference between the count of deficits sampled from the model (either repaired count, damaged count, or total count) and the count from the observed data, for each individual mouse at each time-point. Random scatter about zero suggests unbiased model fits. Error-bars show 95% credible intervals. Credible intervals which do not overlap zero are coloured red, and the proportion of these intervals that overlap zero is shown. The proportions are close to 0.95, as expected for a model accurately representing the uncertainty in the data. R^2 scores are also shown, with values between 0.4 and 0.6.

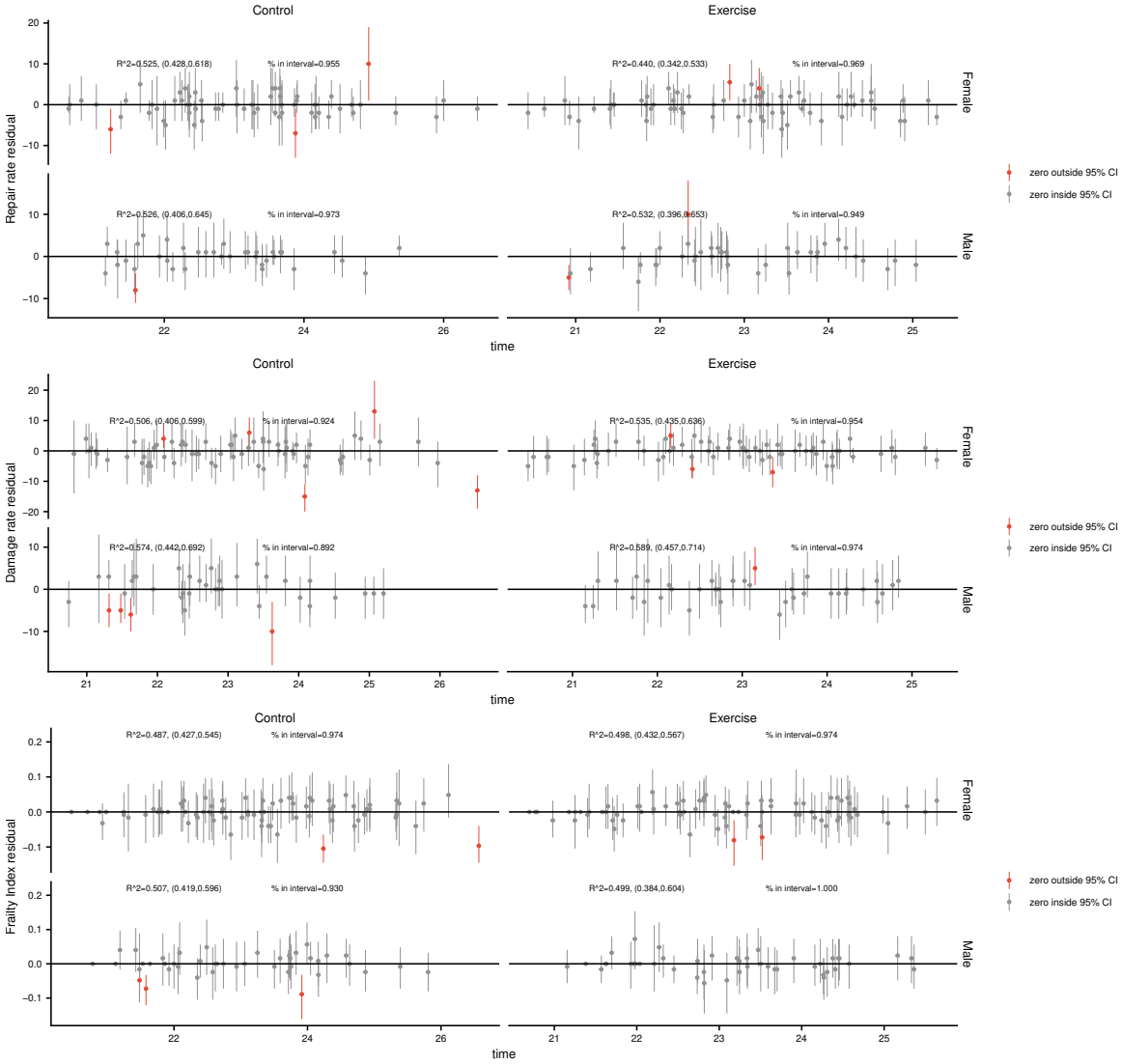


Figure D.10: **Mouse group 2 residuals.** This plot shows the quality of the model fit. Each point shows the median of the difference between the count of deficits sampled from the model (either repaired count, damaged count, or total count) and the count from the observed data, for each individual mouse at each time-point. Random scatter about zero suggests unbiased model fits. Error-bars show 95% credible intervals. Credible intervals which do not overlap zero are coloured red, and the proportion of these intervals that overlap zero is shown. The proportions are close to 0.95, as expected for a model accurately representing the uncertainty in the data. R^2 scores are also shown, with values between 0.4 and 0.6.

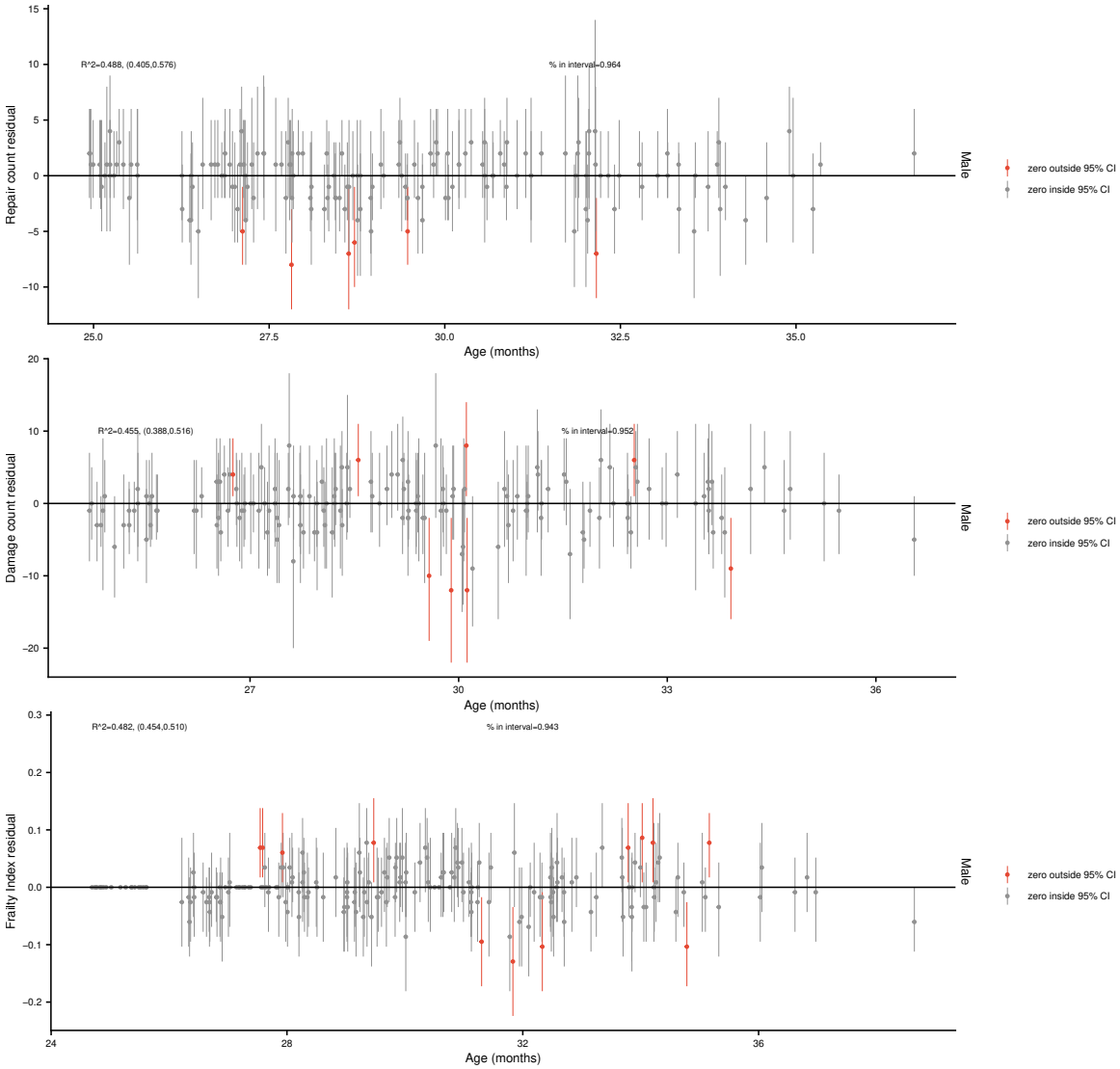


Figure D.11: **Mouse group 3 residuals.** This plot shows the quality of the model fit. Each point shows the median of the difference between the count of deficits sampled from the model (either repaired count, damaged count, or total count) and the count from the observed data, for each individual at each time-point. A random selection of 100 individuals from the ELSA data are shown. Random scatter about zero suggests unbiased model fits. Error-bars show 95% credible intervals. Credible intervals which do not overlap zero are coloured red, and the proportion of these intervals that overlap zero is shown. The proportions are close to 0.95, as expected for a model accurately representing the uncertainty in the data. R^2 scores are also shown, with values between 0.4 and 0.6.

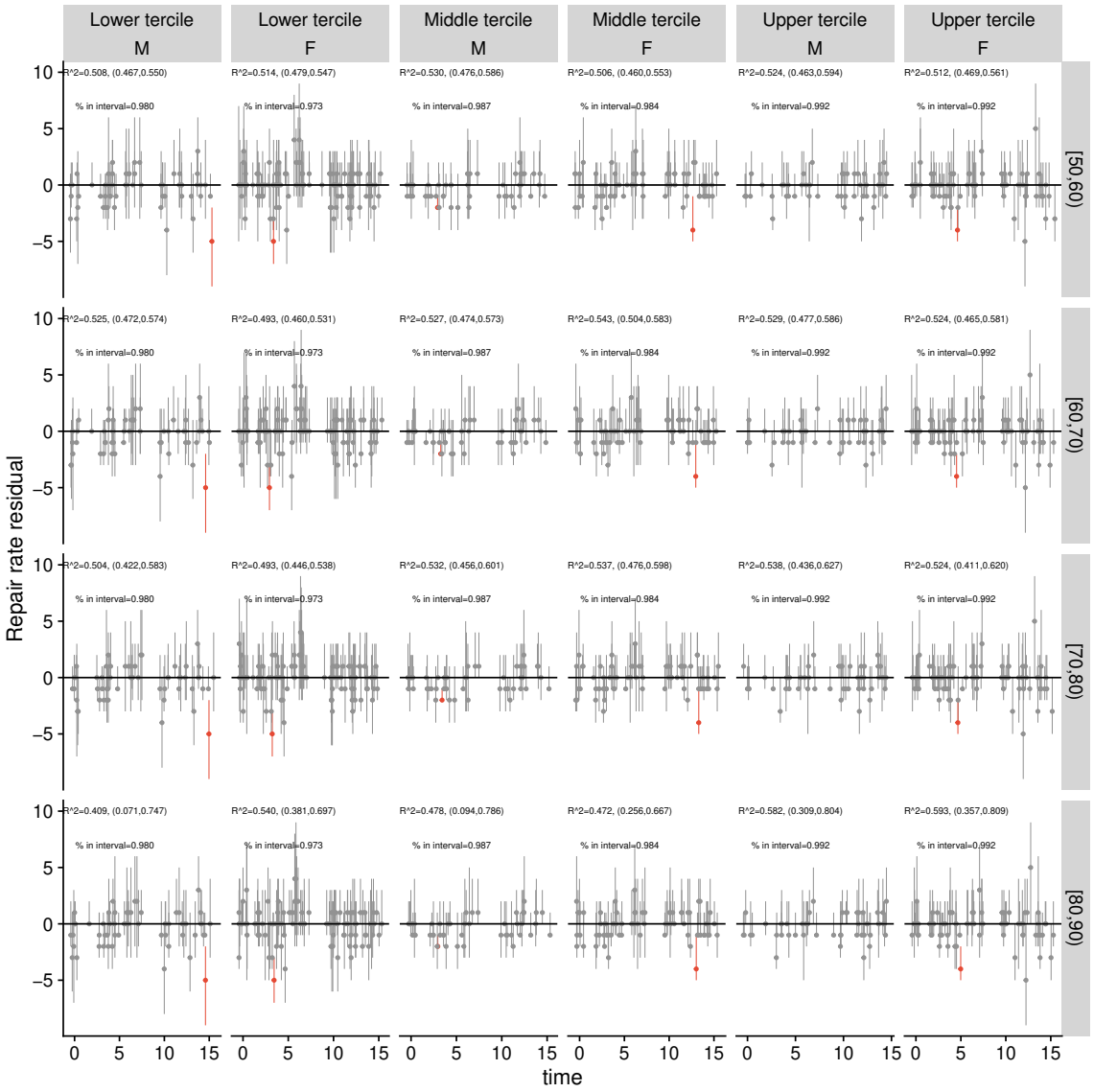


Figure D.12: **Human repair residuals.** This plot shows the quality of the model fit. Each point shows the median of the difference between the count repaired deficits sampled from the model and the count from the observed data, for each individual at each time-point. A random selection of 100 individuals from the ELSA data are shown. Random scatter about zero suggests unbiased model fits. Error-bars show 95% credible intervals. Credible intervals which do not overlap zero are coloured red, and the proportion of these intervals that overlap zero is shown. The proportions are close to 0.95, as expected for a model accurately representing the uncertainty in the data. R^2 scores are also shown, with values between 0.4 and 0.6.

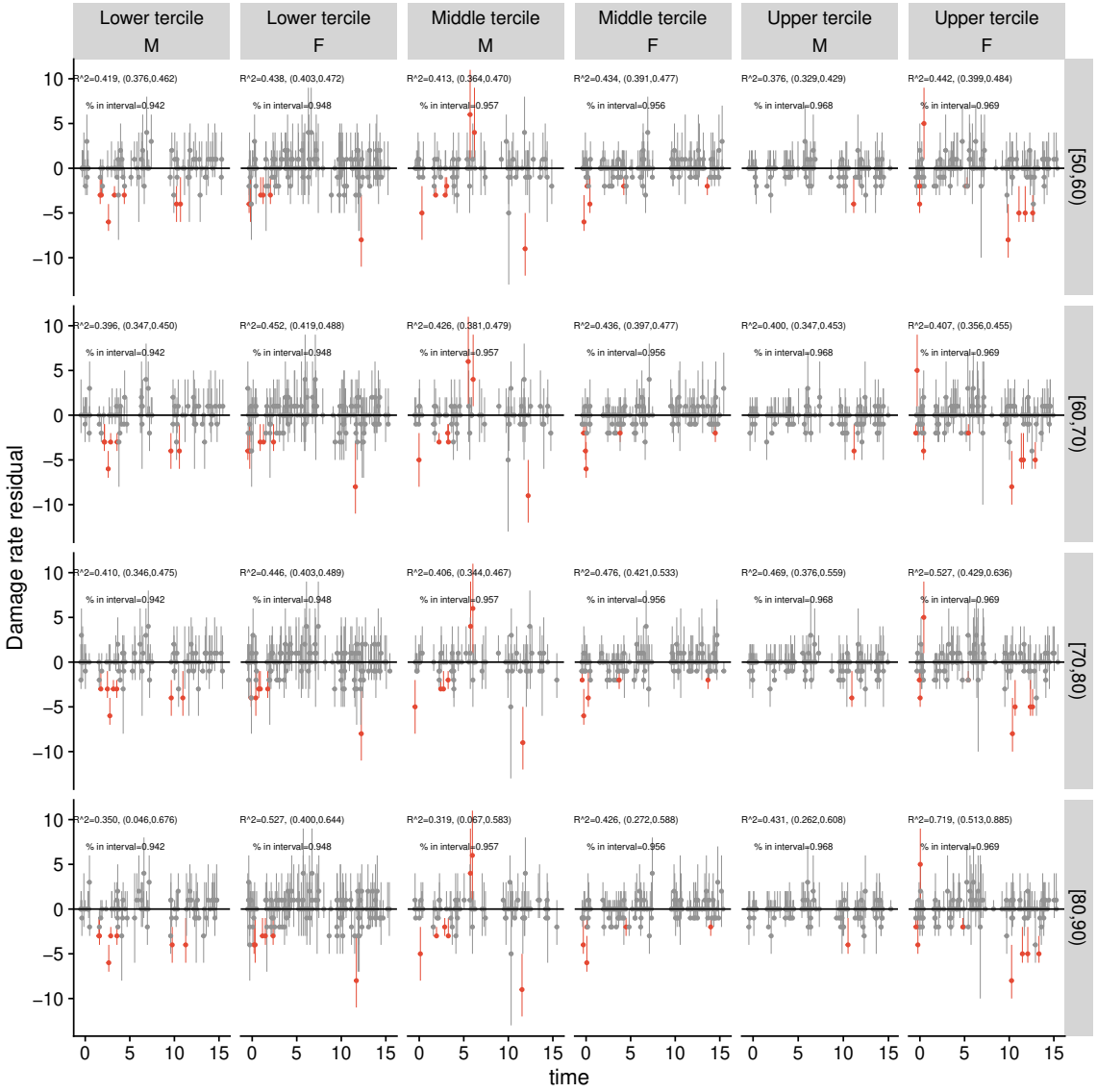


Figure D.13: **Human damage residuals.** This plot shows the quality of the model fit. Each point shows the median of the difference between the count damaged deficits sampled from the model and the count from the observed data, for each individual mouse at each time-point. Random scatter about zero suggests unbiased model fits. Error-bars show 95% credible intervals. Credible intervals which do not overlap zero are coloured red, and the proportion of these intervals that overlap zero is shown.The proportions are close to 0.95, as expected for a model accurately representing the uncertainty in the data. R^2 scores are also shown, with values between 0.4 and 0.6.

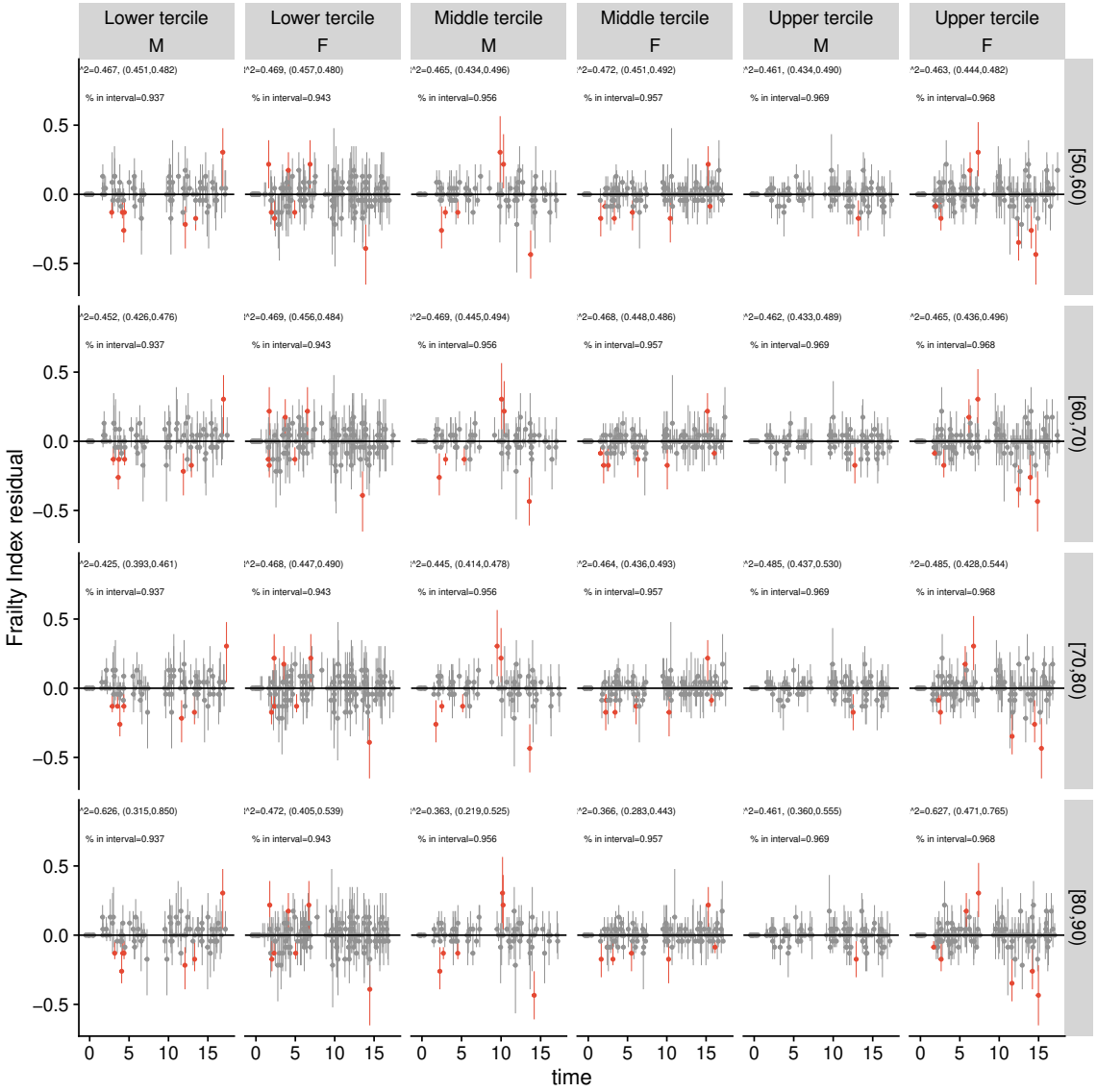


Figure D.14: **Human FI residuals.** This plot shows the quality of the model fit. Each point shows the median of the difference between the count of deficits sampled from the model and the count from the observed data, for each individual mouse at each time-point. Random scatter about zero suggests unbiased model fits. Error-bars show 95% credible intervals. Credible intervals which do not overlap zero are coloured red, and the proportion of these intervals that overlap zero is shown. The proportions are close to 0.95, as expected for a model accurately representing the uncertainty in the data. R^2 scores are also shown, with values between 0.4 and 0.6.