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

Spencer Farrell[1], Arnold Mitnitski[2], Kenneth Rockwood[2], Andrew Rutenberg[1],

1 Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4R2
2 Division of Geriatric Medicine, Dalhousie University, Halifax, Nova Scotia, Canada B3H 2E1 spencer.farrell@dal.ca, andrew.rutenberg@dal.ca

Abstract

We develop a computational model of human aging that generates individual health trajectories using a set of observed health attributes. Our model consists of a network of interacting health attributes that stochastically damage with age to form health deficits, leading to eventual mortality. We train and test the model for two different cross-sectional observational aging studies that include simple clinical indicators of health. In both studies, we find that the individuals generated from the model resemble the observed data in both health characteristics and mortality. Predicted average health trajectories and survival probabilities also agree with the observed data.

Author summary

Human aging is characterized by the individual accumulation of damage. In later life this damage often presents as laboratory abnormalities, diseases, or an impaired ability to carry out daily tasks and activities. This accumulated damage is also connected with an increased risk of mortality. We develop a model to generate realistic and detailed trajectories of health for aging individuals. Given individual health status at a given age, these predicted trajectories describe the likely conditions or disabilities the individual will have in the future. This provides a computational model of human aging that includes observable health features.

Introduction

Human aging is a complex process of stochastic accumulation of damage [1] that occurs at many organismal scales ranging from the cellular [2] 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 [3, 4], 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 [5–8]. 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, combining

many binary deficits to evaluate overall health provides measures that are strongly associated with both adverse health outcomes and mortality [5–8].

Many recent attempts to generate or forecast individual trajectories of disease progression use longitudinal data. Current models can generate and forecast individual trajectories of disease progression [9–11], or jointly forecast both disease progression and survival [12, 13]. Modeling aging more generally is less common.

There are many limitations due to data availability when modeling aging. Typically large observational studies (10^4+ individuals) with linked mortality are cross-sectional (measuring variables only at study entry, with only irregular follow ups), have short censored survival outcomes, and have a lot of missing data. This complicates generating realistic health trajectories during aging. While a model forecasting general health trajectories during aging using cross-sectional data has been developed recently [14], it did not consider individual survival. Recent work used longitudinal data to build a model that describes the aging of mice with a single state variable, although doesn't incorporate survival or predictions of health trajectories [15]. Here, we present a model that can deal with the data limitations of cross-sectional data with censored survival information to generate individual aging trajectories that includes both health and survival.

Our model is adapted from previous work modeling human aging with stochastic dynamics on a complex network [16–18], which was shown to capture population level aging phenomena, such as Gompertz' law of mortality [19]. This model is 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 [20]. 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 involving no age-dependent programming of damage.

In this work, we consider a coarse-grained adaptation of this previous generic network model, but with far fewer nodes. 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.

We parameterize our model with cross-sectional observational data from the Canadian Study of Health and Aging (CSHA) [21] and the National Health and Nutrition Examination Survey (NHANES) [22]. 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 by maximizing the penalized 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 is representing real aspects of aging, and is not overfitting to the training data.

We find that our model is generally predictive of health outcomes and survival with a number of different measures. Rather than focusing on achieving state-of-the-art performance on any one particular task, the goal of our model is to generate realistic

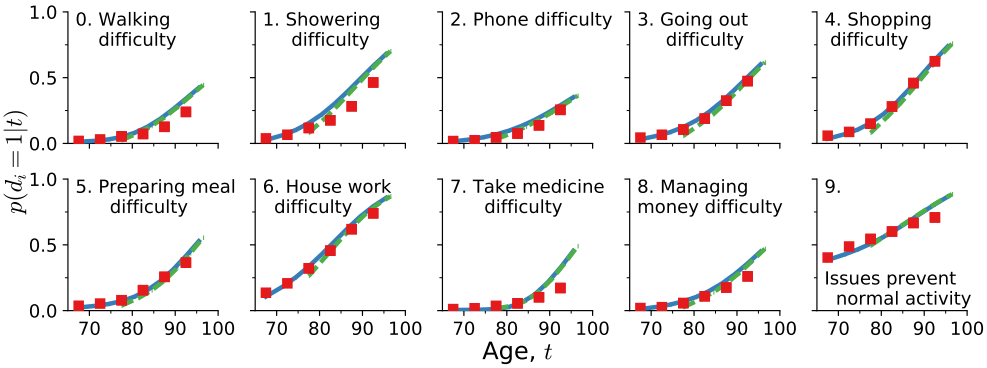

Fig 1. Average predicted trajectories of ten deficit prevalences (as indicated by the subplot titles, numbered 0 to 9) vs. age for individuals from the test data aged $65 - 70$ surviving past 70 (solid blue lines) and aged $75 - 80$ surviving past 80 (dashed green lines). Observed CSHA prevalence is shown in red squares; standard errors are smaller than the point size.

trajectories of individual aging phenotypes from their initial baseline health status.

Results

Health trajectories

Starting from an individual at a baseline age t_0 with deficits $\{d_i(t_0)\}_{i=1}^N$, our model (see Model section below) generates deficit trajectories $\{d_i(t)\}$ describing health for synthetic individuals for each age $t > t_0$ until mortality. 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 ten selected deficits for individuals from the test data aged 65 - 70 that survive past age 70, and individuals aged 75 - 80 that survive past age 80, we forecast health trajectories. 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 Fig. 1 we show health trajectories with deficit prevalence $\hat{p}(d_i = 1|t)$ 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 Supplemental Fig. S4 and with the NHANES data in Supplemental Fig. S5.

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 shown in Fig. 2 and in Supplemental Fig. S6, and 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

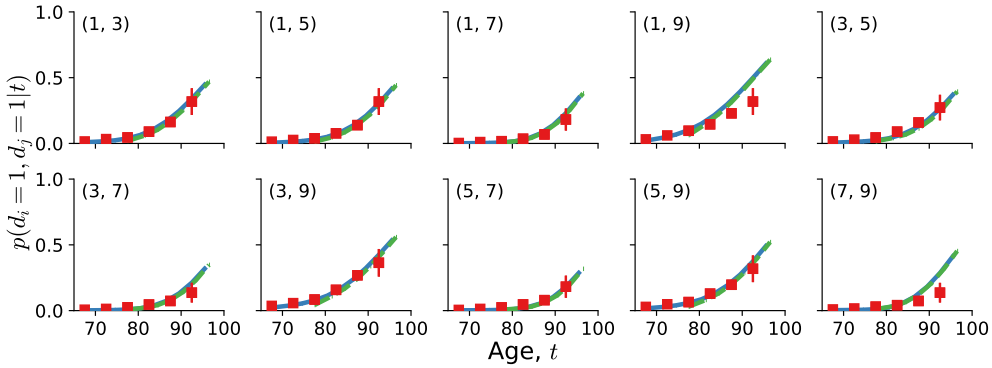

Fig 2. Average predicted trajectories of pairwise deficit prevalence vs. age for individuals from the test data aged $65 - 70$ surviving past 70 (blue solid lines) and aged $75 - 80$ surviving past 80 (dashed green lines). Subplot titles indicate the two deficits included in the pairwise prevalence, with numbers corresponding to titles in Fig. 1. Only odd pairs are shown here; other pairs are shown in Supplemental Fig. S6. CSHA data is shown by red squares; errorbars represent standard errors of the pairwise prevalence.

interactions. Notably, pairwise correlations in the model often perform better than corresponding prevalences – see for example $(1, 7)$ in Fig. 2 – confirming that pairwise correlations (together with higher order correlations, not shown here) are non-trivial predictions of the model.

We can represent overall individual health with the well-established "Frailty Index" (FI), $F = \sum_{i=1}^{N} d_i/N$ [5,8], an index that uses the proportion of deficits as a predictor of health and mortality. In Supplemental Fig. S7 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 Supplemental Fig. S8 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, including partially observed sets of deficits with missing values.

Fig. 3 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.

Individual deficit predictions

Since our training and test data sets have similar prevalences, the good test performances in Figs. 1 and 2 do not rule out overfitting. To assess overfitting, we need to look at predictions at an individual level in the test set. We first check that there is only a small overlap between training and test data sets, see Supplemental Fig. S9.

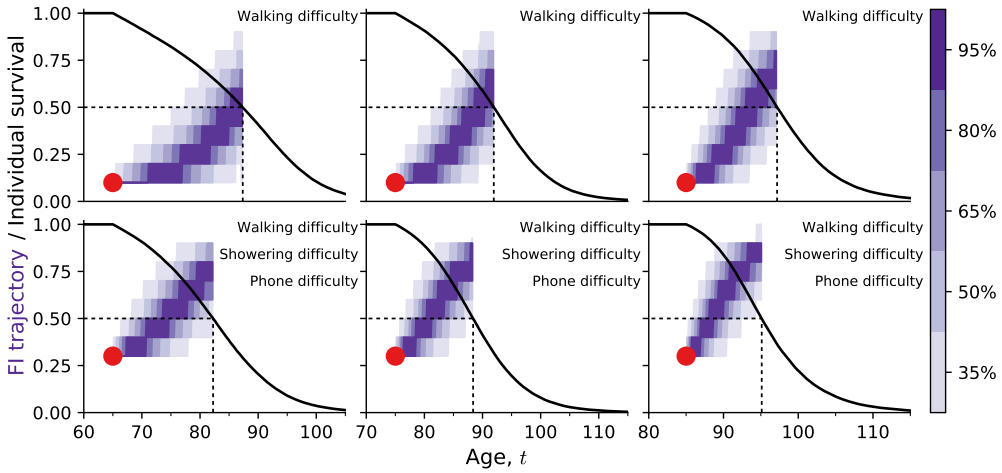

Fig 3. The purple shading indicates the distribution of Frailty Index trajectories vs age for simulated individuals starting from the red circle at a specific age, with the indicated starting set of deficits. Individual survival curves are shown as black lines, and trajectories are conditioned on dying at the median death age (indicated with dashed lines). The scale-bar at right indicates the magnitudes of the trajectory distributions.

We test the model's ability to capture the age-dependent joint distribution of deficits $p(\{x_i\}|t)$ 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 $p(\{x_j\}_{\text{missing}}|\{x_i\}_{\text{observed}}, t)$. Given a known age t^m and known deficits $\{d_i\}^m$ 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 deficits $\{d_{i\neq j}\}$ to estimate the probability of having the left out deficit, $\hat{p}(d_j|\{d_{i\neq j}\}^m, t^m)$. We compare this probability with the actual value of the left out deficit 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.

Fig. 4 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 Fig. 4 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.

Survival

We can also model individual mortality. We take baseline data from the test set for an individual m at age t^m and damage $\{d_i\}^m$, 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

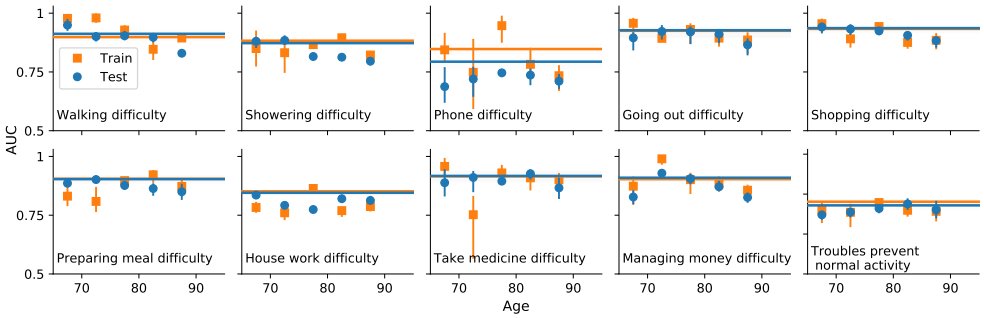

Fig 4. AUC for predicting the value of individual left-out deficits for test (blue circles and lines) and training (orange squares and lines) data. The points show the AUC when stratifying by age. Errorbars indicate 95% confidence intervals. The solid line shows the result for all ages. The baseline AUC is 0.5 for random (uninformative) predictions.

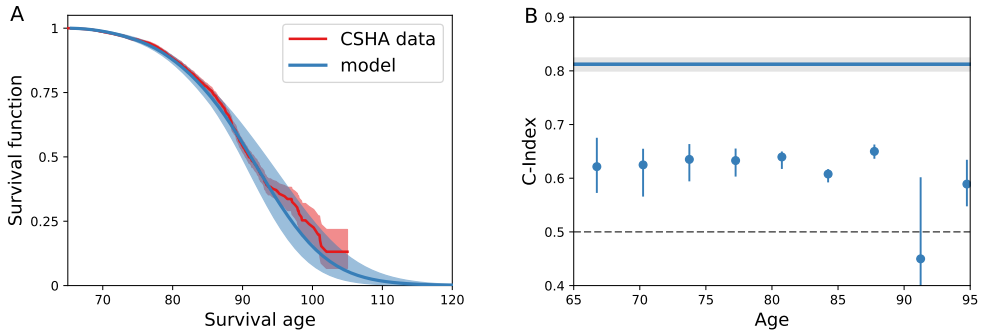

Fig 5. A) Model population average survival function (blue) and CSHA Kaplan-Meier population survival function (red). Error in the model and data are shown as shaded regions for a 95% confidence interval. B) Survival C-Index stratified by age (blue circles) and unstratified (solid blue line). Errorbars on points and grey region around the unstratified line show 95% confidence intervals.

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 [23] of the population survival function from the observed CSHA test data in Fig. 5A, 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 around 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 [24] to measure how well the model discriminates individuals in terms of risk of mortality. Since our model includes potentially complex time-dependent effects, where survival curves can potentially cross, we use a more general age-dependent C-index [25]. We obtain this by comparing the rank ordering between survival probability and known survival age while including censoring, so $C^{\text{td}} = \Pr(\hat{S}_{m_1}(a^{m_1}|t^{m_1}, \{d_i\}^{m_1}) < \hat{S}_{m_2}(a^{m_1}|t^{m_2}, \{d_i\}^{m_2})|a^{m_1} < a^{m_2}, c^{m_1} = 0)$ [25].

Fig. 5B 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, showing that the model captures the increased risk of mortality from specific deficits. We note that stratified values are noisy due to the small number of individuals per age bin, especially at higher ages.

In Supplemental Fig. S10A we show similar results for the C-Index for both training and test sets, which also indicates a lack of overfitting. Similarly, Fig. S10B shows an R^2 measure constructed from Brier Scores [26], a measure of how well predicted and observed survival curves match, that behaves similarly for training and test data. Furthermore, Fig. S10C 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 Fig. S4, as shown in Supplemental Fig. S11 and for survival predictions for the NHANES dataset, as shown in Supplemental Fig. S12.

Summary and Discussion

Our weighted network model (WNM) is trained with cross-sectional data, generates synthetic individuals that resemble the observational data, and can forecast the future health and 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 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 of parameters, with 188 parameters for $N = 10$ binary attributes. A concern with such discrete (binary) health states is that there could be significant overlap between the test and training sets since there are only 2^N possible states. Nevertheless, as we show in Supplemental Fig. S9, a significant fraction of the health states from the test set are not in the training set. Furthermore, from this non-overlapping test set we have shown that our model does not substantially overfit and can make predictions on unseen test individuals.

Our model also 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.

However, we find that when we predict health trajectories until very old ages (80 - 90 years old), our model tends to estimate slightly higher prevalences than are observed in cross-sectional data, 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 of this overestimate is that only $N = 10$ binary deficits are not capturing enough information to predict mortality, so that our maximum likelihood method imposes a middle ground between accurate mortality prediction and accurate health decline by including an artificially rapid health decline in order to induce observed death ages. This effect may be stronger in the NHANES data since health trajectories agree rather poorly, see Supplemental Fig. S5, while survival in Supplemental Fig. S12 is still predictive.

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. This would suggest that these cross-sectional studies could be biased by excluding subjects 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 [27] – this is called "terminal decline" [28]. Using such longitudinal data would allow us to better predict and to better test generated health trajectories for specific individuals, including health near death. This should also improve mortality predictions.

The choice of which deficits to use in our model is arbitrary, the only assumptions we have made are that they are binary and permanent. However, the quality of survival or prevalence predictions (Figs. 4 and 5) does depend on the deficits chosen. 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 [29–31], and environmental variables like socioeconomic status [32], 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 can naturally interact with the damage-accumulation health attributes. Similarly, individual non-damage variables that could be deliberately modified – such as physical activity levels [33, 34] – could be added as nodes with explicit time-dependent values that depend on the individual.

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 [17, 18], 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.

There has been significant work inferring biological networks, both at different scales [35] and using a variety of approaches [36–39]. In the context of human frailty, previous work has created a network representation of health attributes with measures of association or correlation [18, 40, 41]. 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 the Supplemental section on Parameter Robustness, we explore the "robustness" of the network parameters and predictions by sampling an ensemble of parameters around the maximum likelihood estimate [42]. Supplemental Fig. S1 shows that significant deviations from the maximum likelihood parameters still leads to relatively 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 Supplemental Fig. S2 that we find significantly different network parameterizations each time. This indicates that while the model behavior is robust, the network structures themselves are not robustly predicted by the available data.

This lack of robustness for model parameters is not surprising. Due to the complex interactions between many parameters in our model, we expect that many network parameters are "sloppy" [43]. This would lead to robust collective behavior of the system with many different combinations of parameters – i.e. many different networks that are consistent with the observed data. Nevertheless, we show in Supplemental Fig. S3 that how damage propagates from node to node of the network does have some degree of robustness. We show average 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 indicates that the behavior of our weighted network model (WNM) has robust behavior despite some sloppiness of individual parameters. We believe that is why it behaves so well in terms of predicting deficit prevalence and mortality.

Our model differs significantly from other models of disease progression or aging [9–14]. The most important difference is that our model generates trajectories for specific synthetic individuals and directly models stochastic changes to their health state as they age until death – rather than capturing the dynamics with unobserved latent states. We offer a computational model of aging that generates synthetic tracked health trajectories or forecasts 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.

A potential advantage of our computational approach is that model individuals could be manipulated 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.

Model

Model structure

We consider of network of N nodes representing binary health attributes. Each node can be in state $d_i = 0$ for undamaged (healthy) and $d_i = 1$ damaged (deficit). In previous work using a generic network model (GNM) [17, 18], 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, $\Gamma_0^+ \exp(\gamma^+ f_i)$. Repair transitions were also included, but 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 [17, 18].

In this work, we generalize the GNM network to a weighted and directed network, described by a continuous-valued adjacency matrix of weights, w_{ij}. We call this a weighted network model (WNM). We use far fewer 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 $\mu_i(t)$ is implemented as a generic power series, and so our new measure of local damage for the ith node is,

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

$\phi(x) = \max(x, 0)$ is a function that clips negative rates to zero, allowing the model to effectively "turn on" rates at older ages rather than zero. This allows for strong non-linear behaviour even with a low-order power series. $\{w_{ij}, \mu_{ik}\}$ are fitting parameters of the model while n_f is a hyperparameter for $\mu_i(t)$.

The exponential damage rates of the GNM have been replaced by more general power-series with node-dependent coefficients to allow specific nodes to capture distinct behaviour for specific deficits in the observed data. The new damage rate for node i is given by,

$$\Gamma_i^+(t,\{d_j\}) \quad = \quad \phi\Big(\sum_{k=0}^{n_+}\gamma_{ik}^+ f_i^k(t)\Big). \tag{2}$$

$\{\gamma_{ik}^+\}$ are fitting parameters of the model and n_+ is a hyperparameter for the power series.

Mortality occurs as a separate process with a single rate of death,

$$\Gamma_D(t,\{d_j\}) \quad = \quad \phi\Big(\sum_{k=0}^{n_{D_1}}\alpha_k x^k\Big), \tag{3}$$

$$x(t,\{d_j\}) \quad = \quad \phi\Big(\sum_{j=1}^{N}\beta_j d_j + \sum_{l=0}^{n_{D_2}}\eta_l t^l\Big). \tag{4}$$

This 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 [16, 17]). The intermediate damage x depends on each deficit linearly with an age-dependent deficit-independent component represented as a power series. This mortality rate uses fitting parameters $\{\alpha_k,\, \beta_j,\, \eta_l\}$ and n_{D_1}, n_{D_2} as hyperparameters determining the number of terms in the power series.

In total the model has $N_{\rm tot} = N(N + n_f + n_+ + 1) + n_{D_1} + n_{D_2} + 2$ fitting parameters. We restrict parameter values to ensure that f_i, Γ_i^+, and Γ_D are all monotonically increasing functions of age. Details of the requisite parameter bounds are in the Supplemental. 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 Supplemental Fig. S9). 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) [44]. Details of the stochastic simulation are in the Supplemental. 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, $\{d_i(t)\}_{t=0,i=1}^{t_D,i=N}$. Due to the exact nature of the SSA, all transition times are precisely resolved in our model data.

Likelihood

We calculate our likelihood using cross-sectional data. First, we consider observed cross-sectional data of the form $\left\{t^m, \{d_i\}^m, a^m, c^m\right\}_{m=1}^{M}$. 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 $a^m \equiv \min\left(t_D^m, t_c^m\right)$, 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 $c^m = 0$.

By simulating synthetic individuals from the model, we sample and estimate the probability $p(\{d_i\}^m, t_D^m | t^m; \vec{\theta})$ for each individual m in the data. We denote all

parameters by the vector $\vec{\theta}$. For simplicity we split this probability into two separate parts, representing mortality and health respectively:

$$\log p(\{d_i\}^m, t_D^m | t^m; \vec{\theta}) = \log p(t_D^m | \{d_i\}^m, t^m; \vec{\theta}) + \log p(\{d_i\}^m | t^m; \vec{\theta}). \tag{5}$$

For uncensored individuals, we can calculate their likelihood by using their known death age using Eqn. 5. For censored individuals, we also need to integrate the mortality term over all possible death ages above the censoring age,

$$S(a^m | t^m, \{d_i\}^m; \vec{\theta}) = \int_{a^m}^{\infty} p(a' | t^m, \{d_i\}^m; \vec{\theta}) da', \tag{6}$$

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

$$\begin{aligned} L(\vec{\theta}) &= L_{\text{mortality}}^{\text{uncensored}} + L_{\text{health}} + L_{\text{mortality}}^{\text{censored}} \\ &= \sum_{m|c^m=0} \log p(a^m | \{d_i\}^m, t^m; \vec{\theta}) \\ &+ \sum_m \log p(\{d_i\}^m | t^m; \vec{\theta}) \\ &+ \sum_{m|c^m=1} \log S(a^m | t^m, \{d_i\}^m; \vec{\theta}), \end{aligned} \tag{7}$$

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 variables implicitly by sampling all possible combinations of the missing variables. Additional details of the likelihood estimation from simulations are in the Supplemental.

Observed data

We use data from the Canadian Study of Health and Aging (CSHA) [21] to develop and test our model. The CSHA sample has 8547 individuals that range from ages $65-99$ with death ages that are available within a 6 year censoring window. The 10 default deficits used are shown in Fig. 1, alternate deficits are shown in the Supplemental.

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 optimization by having a much larger number of individuals.

We validate our conclusions on the National Health And Nutrition Examination Survey (NHANES) [22]. The NHANES sample has 9504 individuals that range from ages $20-85$ with death ages that are available within a 10 year censoring window. In the same way this data is split into 2352 training individuals and 7151 test individuals.

Parameter optimization

We maximize the log-likelihood in Eqn. 7 using particle swarm optimization [45] in order to train the model and estimate the parameters $\hat{\theta}$. Details of the parameter optimization procedure are in the Supplemental. We use parameter bounds shown in the Supplemental to impose monotonic dependence of damage rates on existing damage. We regularize the fitting as detailed in the Supplemental. 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 $N = 10$.

Supporting information

Appendix S1. Supplemental details

Fig. S1 Prediction robustness.

Fig. S2 Network sloppiness.

Fig. S3 Damage rate robustness.

Fig. S4 Alternate deficit predicted trajectories.

Fig. S5 NHANES predicted trajectories.

Fig. S6 Full pair trajectories.

Fig. S7 Population Frailty Index distributions.

Fig. S8 Simulated population trajectories from birth.

Fig. S9 CSHA state overlap

Fig. S10 Training set mortality predictions.

Fig. S11 Alternate deficits mortality predictions

Fig. S12 NHANES survival predictions

References

1. Kirkwood TBL. Understanding the odd science of aging. Cell. 2005;120:437 – 447.
2. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153:1194–1217.
3. Herndon LA, Schmeissner PJ, Dudaronek JM, Brown PA, Listner KM, Sakano Y, et al. Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. Nature. 2002;419:808 – 814.
4. Kirkwood TBL, Finch CE. The old worm turns more slowly. Nature. 2002;419:794–795.
5. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. The Scientific World. 2001;1:323–36.

6. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: Evidence for a phenotype. The Journals of Gerontology: Series A. 2001;56(3):M146–M157.

7. Kulminski AM, Ukraintseva SV, Akushevich IV, Arbeev KG, Yashin AI. Cumulative index of health deficiencies as a characteristic of long life. Journal of the American Geriatrics Society. 2007;55:935–940.

8. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatrics. 2008;8:24.

9. Schulam P, Saria S. A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure. In: Advances in Neural Information Processing Systems. Johns Hopkins University, Baltimore, United States; 2015. p. 748–756.

10. Alaa AM, van der Schaar M. Forecasting individualized disease trajectories using interpretable deep learning. arXiv. 2018;arXiv:1810.10489v1.

11. Fisher CK, Smith AM, Walsh JR. Machine learning for comprehensive forecasting of Alzheimer's Disease progression. Scientific Reports. 2019;9(1):1–14.

12. Hickey GL, Philipson P, Jorgenson A, Kolamunnage-Dona. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. BMC Medical Research Methodology. 2016;16(117).

13. Lim B, van der Schaar M. Disease-Atlas: Navigating disease trajectories using deep learning. Proceeding of Machine Learning Research. 2018;85:137–160.

14. Pierson E, Koh PW, Hashimoto T, Koller D, Liang P. Inferring multidimensional rates of aging from cross-sectional data. Proc Mach Learn Res. 2019;89:97–107.

15. Avchaciov K, Antoch MP, Andrianova EL, Tarkhov AE, Menshikov LI, Burmistrova O, et al. Identification of a blood test-based biomarker of aging through deep learning of aging trajectories in large phenotypic datasets of mice. bioRxiv. 2020;doi:http://dx.doi.org/10.1101/2020.01.23.917286.

16. Taneja S, Mitnitski AB, Rockwood K, Rutenberg AD. Dynamical network model for age-related health deficits and mortality. Physical Review E. 2016;93:022309.

17. Farrell SG, Mitnitski AB, Rockwood K, Rutenberg AD. Network model of human aging: frailty limits and information measures. Physical Review E. 2016;94:052409.

18. Farrell SG, Mitnitski AB, Rockwood K, Rutenberg AD. Probing the network structure of health deficits in human aging. Physical Review E. 2018;98:032302.

19. Kirkwood TBL. 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. 2015;370:20140379.

20. Rutenberg AD, Mitnitski AB, Farrell SG, Rockwood K. Unifying aging and frailty through complex dynamical networks. Experimental Gerontology. 2018;107:126–129.

21. Canadian Study of Health and Aging Working Group. Canadian study of health and aging: study methods and prevalence of dementia. Canadian Medical Association Journal. 1994;150(6):899.

22. Centers for Disease Control and Prevention National Center for Health Statistics. National Health and Nutrition Examination Survey Data; Updated 2014. Available from: `http://www.cdc.gov/nchs/nhanes.htm`.

23. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 1958;53:457–481.

24. Harrell Jr FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. The Journal of the American Medical Association. 1982;247(18):2543–2546.

25. Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. Statistics in Medicine. 2005;24:3927 – 3944.

26. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Statistics in Medicine. 1999;18:2529–2545.

27. Stow D, Matthews FE, Hanratty B. Frailty trajectories to identify end of life: a longitudinal population study. BMC Medicine. 2018;16(171).

28. Cohen-Mansfield J, Skornick-Bouchbinder M, Brill S. Trajectories of end of life: A systematic review. Journals of Gerontology - Series B Psychological Sciences and Social Sciences. 2018;73(4):564–572.

29. Puts MTE, Lips P, Deeg DJH. Sex differences in the risk of frailty for mortality independent of disability and chronic diseases. Journal of the American Geriatrics Society. 2005;53:40–47.

30. Mitnitski A, Song X, Skoog I, Broe GA, Cox JL, Grunfeld E, et al. Relative fitness and frailty of elderly men and women in developed countries and their relationship with mortality. Journal of the American Geriatrics Society. 2005;53:2184–2189.

31. Gordon EH, Peel NM, Theou O, Howlett SE, Hubbard RE. Sex differences in frailty: A systematic review and meta-analysis. Experimental Gerontology. 2017;89:30–40.

32. Andrew MK, Mitnitski AB, Rockwood K. Social vulnerability, frailty and mortality in elderly people. PLOS One. 2008;3(5):e2232.

33. Fried LP. Interventions for human frailty: Physical activity as a model. Cold Spring Harbor Perspectives in Medicine. 2016;6:a025916.

34. Rogers NT, Marshall A, Roberts CH, Demakakos P, Steptoe A, Scholes S. Physical activity and trajectories of frailty among older adults: Evidence from the English Longitudinal Study of Ageing. PLOS One. 2017;12(2):e0170878.

35. Natale JL, Hofmann D, Damian G H, Nemenman I. Reverse-engineering biological networks from large data sets. Quantitative Biology Theory, Computational Methods, and Models: Chapter 10. 2017;.

36. Butte AJ, Kohane IS. Mutual information relevance Networks: functional genomic clustering using pairwise entropy measurements. Pacific Symposium on Biocomputing. 2000;5:415–426.

37. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, et al. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006;7:S7.

38. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Statistical Applications in Genetics and Molecular Biology. 2005;4:1.

39. Shen Z, Wang WX, Fan Y, Di Z, Lai YC. Reconstructing propagation networks with natural diversity and identifying hidden sources. Nature Communications. 2014;5:4323.

40. Mitnitski AB, Graham JE, Mogilner AJ, Rockwood K. Frailty, fitness and late-life mortality in relation to chronological and biological age. BMC Geriatrics. 2002;2:1.

41. García-Peña C, Ramírez-Aldana R, Parra-Rodriguez L, Gómez-Verján JC, Pérez-Zepeda MU, Gutiérrez-Robledo LM. Network analysis of frailty and aging: Empirical data from the Mexican Health and Aging Study. Experimental gerontology. 2019;128:110747.

42. Zhou B, Hofmann D, Pinkoviezky I, Sober SJ, Nemenman I. Change, long tails, and inference in a non-Gaussian, Bayesian theory of vocal learning in songbirds. Proceedings of the National Academy of Sciences. 2018;115(38):E8358.

43. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. PLoS Comput Biol. 2007;3:10.

44. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry. 1977;81:25.

45. Kennedy J, Eberheart R. Particle swarm optimization. Proceedings of IEEE International Conference on Neural Networks. 1995;IV(1942):1942–1948.