

---

# Inferring a measure of physiological age from multiple ageing related phenotypes

---

**David Knowles**  
Cambridge University Engineering Department

**Leopold Parts**  
University of Toronto

**Daniel Glass**  
King’s College London

**John M. Winn**  
Microsoft Research Cambridge

## Abstract

What is ageing? One definition is simultaneous degradation of multiple organ systems. Can an individual be said to be “old” or “young” for their (chronological) age in a scientifically meaningful way? We investigate these questions statistically using ageing related phenotypes measured on the 12,000 female twins in the Twins UK study [4]. We propose a simple linear model of ageing, which allows a latent adjustment  $\Delta$  to be made to an individual’s chronological age to give their “physiological age”, shared across the observed phenotypes. We note problems with the analysis resulting from the linearity assumption and show how to alleviate these issues using a non-linear extension. We find more gene expression probes are significantly associated with our measurement of physiological age than to chronological age.

The aim of this work is to infer a measure of the “physiological age” of an individual, in this case using the ten ageing related phenotypes shown in Table 1.

Phenotype	Correlation
Lens nuclear dip (cataract)	0.63
FEV (lung function)	-0.61
Grip strength	-0.38
Telomere length	-0.33
Bone mineral density	-0.38
Total moles	-0.22
DHEAS	-0.19
Eye wrinkles	0.16
Neck wrinkles	0.21
Hair loss	0.18

Table 1: Phenotypes used for inferring our measure of physiological age, and Pearson correlation coefficient to chronological age. Most are self explanatory; DHEAS is an important hormone.

We are interested in whether this measure correlates more strongly to low level phenotypes, such as gene expression measurements, than chronological age, or the individual phenotypes used. This continues our previous work of finding low dimensional summaries of high dimensional phenotypes, but with the added characteristic of the summary being interpretable as ageing. In fact our model represents “physiological age” as  $a + \Delta$ , where  $a$  is chronological age, and  $\Delta$  is the unobserved variable representing how much older this individual appears biologically than the average.

This work is being performed in collaboration with the Department of Twin Research and Genetic Epidemiology (DTR) at King’s College London. The DTR manages the largest UK adult twin

registry of around 12,000 female monozygotic and dizygotic twins, established in 1992 [4]. The data has characteristics common to many biomedical applications:

1. *High missingness.* Many variables have up to 80% missing, and the level of overlap between phenotypes varies considerably. This level of missingness motivates Bayesian methods which are able to naturally deal with missingness, rather than attempting crude imputation procedures.
2. *Heterogeneous data.* The data contains continuous, categorical (including binary), ordinal and count data.
3. *High dimensional.* The Twins UK database contains over 6000 phenotype and exposure variables, measured at multiple time points. Modern healthcare records are of the same nature. For a subset of 800 individuals we have 10,000 gene expression measurements in three different tissues, and the genotype of 600k Single Nucleotide Polymorphisms (SNPs).

In previous work a general modelling framework was presented for this data [2], allowing these issues to be straightforwardly and rigorously addressed using Bayesian probabilistic models fit using Variational Message Passing [5] under the Infer.NET framework [3].

It is helpful when collaborating with clinicians to be able to explain the core idea of a method without resorting to mathematics. A picture is worth a thousand words, and so is presumably worth a fair number of LaTeX symbols. Figure 1 shows an attempt to achieve this: highlighted by a green circle is a particular 55 year old individual. From the graph of lens nuclear dip (a measure of how bad an individual’s cataracts are) against chronological age, we see that her cataracts are quite bad for her age, with a value of around 4.5 compared to the least squares best fit of around 3.9. In fact, we could anecdotally say that she has the cataracts of a 79 year old, implying an age adjustment  $\Delta = +24$  years.

Our model embodies this idea, but provides additional benefits:

- Optimally aggregating information across multiple phenotypes
- Fitting the noise variance jointly with  $\Delta$ , preventing overfitting: in the example in Figure 1 the individuals unusually high cataract measurement is likely to be at least partly noise, which would reduce our estimate of  $\Delta$
- Can be extended to allow non-linear variation with age

## 1 Models

We here describe the probabilistic models we will use to infer our measure of physiological age. Assume we have observations  $y_{nd}, n = 1 \dots N, d = 1 \dots D$  where  $d$  indexes the phenotypes and  $n$  indexes the individuals, whose chronological ages  $a_n$  are known. Suitable diffuse priors are specified on all random variables, but these are not shown for clarity of exposition.

### 1.1 Linear model

The simplest model we propose is a simple extension of linear regression, as follows:

$$g_{nd} = b_d(a_n + \Delta_n) + c_d \tag{1}$$

where

- $g_{nd}$  is the auxiliary variable associated with phenotype observation  $y_{nd}$
- $c_d$  is the intercept for phenotype  $d$
- $b_d$  is the regression coefficient for phenotype  $d$
- $\Delta_n$  additional ageing of individual  $n$

One can think of  $\Delta_n$  as input noise, but crucially, its value is common to all phenotypes, embodying the assumption that there is global systemic ageing which effects all phenotypes. The Infer.NET [3] code for the core of the model (for continuous phenotypes) is shown in Figure 2.

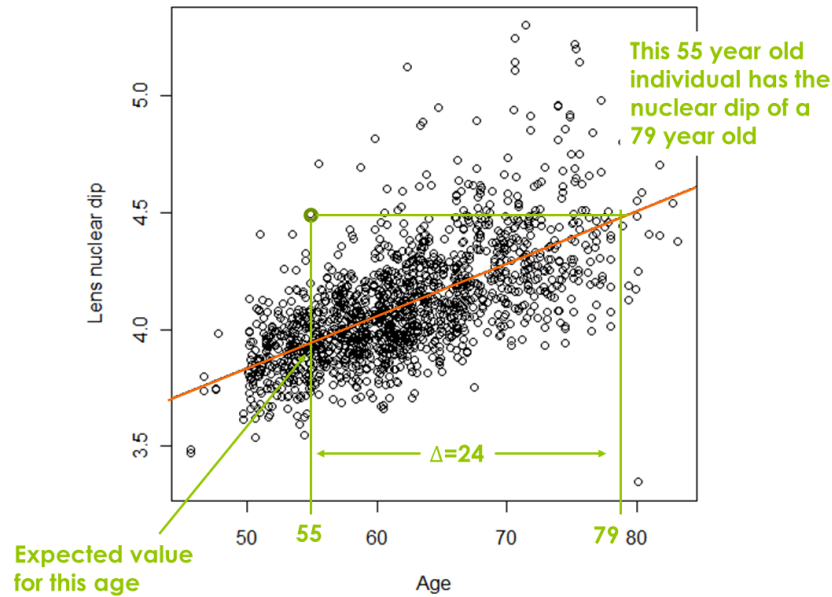


Figure 1: An intuitive explanation of our model. This 55 year old individual has the cataracts of a 79 year old, implying  $\Delta = 24$  years.

```

var n = new Range(6000);
var p = new Range(10);
var a = Variable.Array<double>(n);
var delta = Variable.Array<double>(n);
delta[n] = Variable.GaussianFromMeanAndPrecision(0, 1).ForEach(n);
var b = Variable.Array<double>(p);
b[p] = Variable.GaussianFromMeanAndPrecision(0, 0.1).ForEach(p);
var c = Variable.Array<double>(p);
c[p] = Variable.GaussianFromMeanAndPrecision(0, 0.1).ForEach(p);
var y = Variable.Array<double>(n, p);
var noisePrecisions = Variable.Array<double>(p);
noisePrecisions[p] = Variable.GammaFromShapeAndRate(1, 0.1).ForEach(p);
y[n, p] = Variable.GaussianFromMeanAndPrecision(b[p] * (a[n] + delta[n]) + c[p],
                                                noisePrecisions[p]);

```

Figure 2: Infer.NET [3] code implementing the linear model (Equation 1) for continuous variables, with the key likelihood term highlighted.

How the observations themselves are generated under this model depends on the data type, as follows.

**Continuous variables.** Continuous variables are modeled simply by adding Gaussian noise to  $g_{nd}$ :

$$y_{nd}^c \sim N(y_{nd}^c; 0, \sigma_d^2) \quad (2)$$

$$\sigma_d^2 \sim IG(\sigma_d^2; 1, 1) \quad (3)$$

where IG is the inverse-Gamma distribution.

**Binary variables.** For binary variables we use a logistic link function  $\sigma(x) = (1 + e^{-x})^{-1}$  in an analogous manner to logistic regression.

$$y_{nd}^b \sim \text{Bernoulli}(\sigma(g_{nd})) \quad (4)$$

**Ordinal variables.** Ordered categorical (ordinal) variables are common in biomedical data, for example, severity of a condition. Assume we have a Gaussian predictor variable  $g$  and an observed ordinal variable  $y \in [1, \dots, J]$ . Let the likelihood function be

$$P(Y = j|g) = \sigma(\tau_j - g) - \sigma(\tau_{j-1} - g) = \sigma(g - \tau_{j-1}) - \sigma(g - \tau_j) \quad (5)$$

where the logistic function  $\sigma(x) = 1/(1 + e^{-x})$  and  $\{\tau_j : j = 0..J\}$  are interval boundaries with  $\tau_0 = -\infty, \tau_{j-1} < \tau_j, \tau_J = +\infty$ .

## 1.2 Non-linear model V1

A straightforward mechanism to obtain nonlinear regression models is to use a mixture of experts where the input itself determines the cluster assignment. For example consider the following model on one variable:

$$\pi \sim \text{Dirichlet}(\alpha) \quad (6)$$

$$z_n \sim \text{Discrete}(\pi) \quad (7)$$

$$x_n \sim N(m_{z_n}, v_{z_n}) \quad (8)$$

$$y_n \sim N(b_{z_n}x_n + c_{z_n}, \phi_{z_n}) \quad (9)$$

where  $\pi$  are mixture weights,  $z_n$  are cluster assignments,  $x$  is the input variable,  $y$  is the output,  $m_k, v_k$  are means and variances for a naive Bayes model of which cluster to use based on  $x$ , and  $b_k, c_k$  are coefficients and intercepts for the regression model. This is a finite version of the Dirichlet Process mixture of General Linear Models [1].

Combining this method of obtaining nonlinearity with the ageing model in Equation 1 results in the following model:

$$\pi_d \sim \text{Dirichlet}(\alpha) \quad (10)$$

$$z_{nd} \sim \text{Discrete}(\pi_d) \quad (11)$$

$$a_n \sim N(m_{dz_{nd}} - \Delta_n, v_{dz_{nd}}) \quad (12)$$

$$y_{dn} \sim N(b_{dz_{nd}}[a_n + \Delta_n] + c_{dz_{nd}}, \phi_{dz_{nd}}^{-1}) \quad (13)$$

where

- $\pi_d$  are the mixture weights for phenotype  $d$
- $z_{nd}$  is the cluster assignment variable for individual  $n$  in phenotype  $d$
- $m_{dk}, v_{dk}$  are the mean and variance for naive Bayes in cluster  $k$  of phenotype  $d$
- $b_{dk}, c_{dk}$  are the coefficient and intercept respectively for the  $k$ -th regression model for phenotype  $d$

Here we have a separate nonlinear function (modelled using Equation 9) for each phenotype. Note that we implement this model as  $a_n \sim \dots$  rather than  $a_n + \Delta_n \sim \dots$  since in a probabilistic model a random variable such as  $\Delta_n$  cannot have multiple generating mechanisms ( $a_n$  is non-random so it can simply be replicated). Binary and ordinal variables are treated analogously to the linear model.

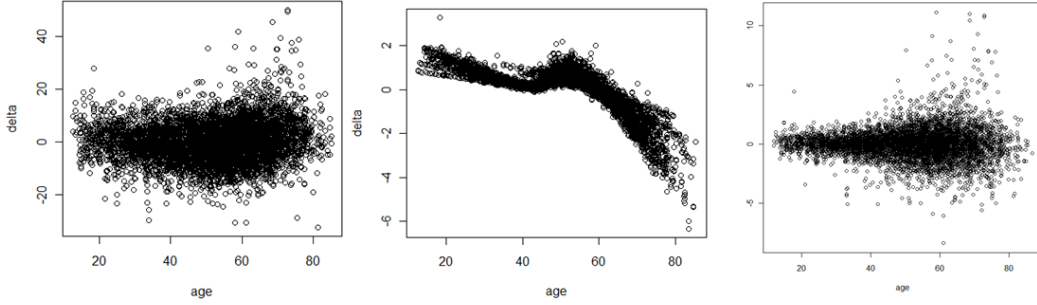


Figure 3: Plots of  $\Delta$  against age (note the different  $y$ -scales) for the three different models. Left: linear model. Centre: Non-linear V1. Right: Non-linear V2.

### 1.3 Non-linear model V2

In the Results section below we will see that for our purpose of inferring  $\Delta$  the nonlinear model proposed above actually has some highly undesirable behaviour. We show this can be alleviated by replacing the naive Bayes link from the input variable  $x$  to the cluster assignment  $z$  with a multinomial regression model, as follows for the one dimensional setting:

$$h_{nk} = s_k x_n + m_k \quad (14)$$

$$z_n \sim \text{Discrete}(\sigma(h_n)) \quad (15)$$

$$y_n \sim N(b_{z_n} x_n + c_{z_n}, \phi_{z_n}^{-1}) \quad (16)$$

Extending this to the ageing model is in fact more straightforward than for the previous model.

$$h_{ndk} = s_{dk}(a_n + \Delta_n) + m_{dk} \quad (17)$$

$$z_{nd} \sim \text{Discrete}(\sigma(h_{nd:})) \quad (18)$$

$$y_{nd} \sim N(b_{dz_{nd}}[a_n + \Delta_n] + c_{dz_{nd}}, \phi_d^{-1}) \quad (19)$$

The binary and ordinal observation models are analogous to those used in the linear case.

## 2 Results

We run various experiments to validate the models and inference, which is performed using Variational Message Passing [5] under the Infer.NET framework [3].

### 2.1 Comparing models

We expect that the  $\Delta$  values we learn will be roughly independent of age, so plotting learnt  $\Delta$  values against age as shown in Figure 2.1 is a useful diagnostic. For the linear model we see the variance of  $\Delta$  is large, and  $\Delta$  shows significant dependence on age. The model uses the additional flexibility provided by  $\Delta$  to help fit non-linear trends of the phenotypes in age, which is clearly undesirable. This motivates using a non-linear model, for example non-linear model V1 (see Section 1.2). However, for this model (with  $K = 2$ ) the dependence of  $\Delta$  on age is actually much worse. This is due to the naive Bayes construction of the dependence of the cluster assignment  $z$  on the input variable  $x$ , which results in a factor  $N(a_n; m_{dz_{nd}} - \Delta_n, v_{dz_{nd}}) = N(\Delta_n; m_{dz_{nd}} - a_n, v_{dz_{nd}})$  in the model. This aggressively regularises  $\Delta_n$  towards  $m_{dz_{nd}} - a_n$ . Typically with  $K = 2$  one cluster is used for younger individuals, and one for older, resulting in the two linear regions in the middle plot around age 30 and 60. To fix this behaviour we introduce the non-linear model described in Section 1.3. Figure 2.1 (right) shows the improvement: the variance of the  $\Delta$  values is much reduced compared to the linear model, and there is no noticeable dependence on age. Cluster assignment for this model with  $K = 3$  is shown in Figure 2.1. The remainder of the paper will assume this model.

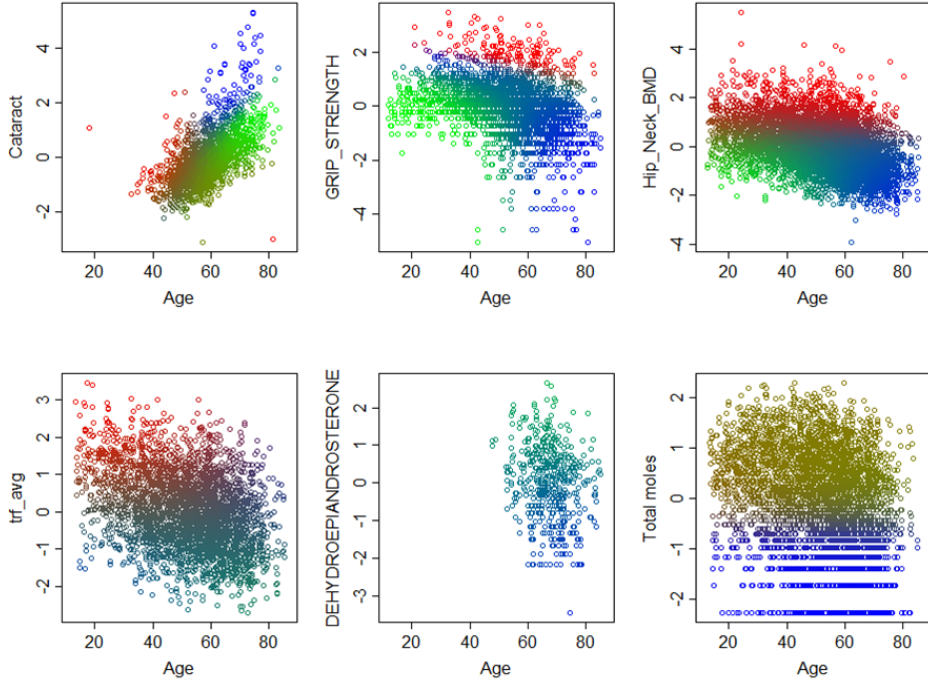


Figure 4: Cluster assignments for six continuous phenotypes for non-linear model V2 with  $K = 3$  mixture components.

Phenotype	Chronological age	Physiological age
Lung function (FEV)	Inf	Inf
Lens nuclear dip	<b>185</b>	80
Grip strength	85	<b>168</b>
Hip/Neck BMD	209	<b>227</b>
Telomere length	<b>80</b>	55
DHEAS	4	<b>7</b>
Total moles	50	<b>51</b>

Table 2: Hold out test. Values are  $\log_{10}(p)$  where  $p$  is the  $p$ -value for the Spearman rank correlation hypothesis test.

## 2.2 Biological validation

In order to validate  $\Delta$  we perform the following imputation experiment. For each phenotype in turn we learn  $\Delta_n$  with this phenotype held out, i.e. using all phenotypes but the current one. If the inferred “physiological age” ( $\Delta_n + a_n$ ) is a biologically meaning estimate of global systemic ageing, then we would expect it to be a better predictor of the held out phenotype than chronological age. We assess this in two ways. Firstly we use the standard Fisher transformation hypothesis test for the Spearman rank correlation coefficient calculated using R, as shown in Table 2.

We see that for grip strength, bone mineral density, DHEAS and total moles, the  $p$ -value is more significant for physiological age than for chronological age, suggesting that these phenotypes indeed exhibit common ageing. The negative result for lens nuclear dip and telomere lengths suggests that these phenotypes do not share any ageing behaviour with the other phenotypes included here. Unfortunately the  $p$ -values for lung function underflowed to 0. A second approach is to use Bayes factors between the non-linear model in Equation 16 conditioned on chronological or physiological age. The results are shown in Figure 2.2, where we are also able to include binary variables unlike for the Spearman test. The results are broadly consistent with the hypothesis testing approach: grip

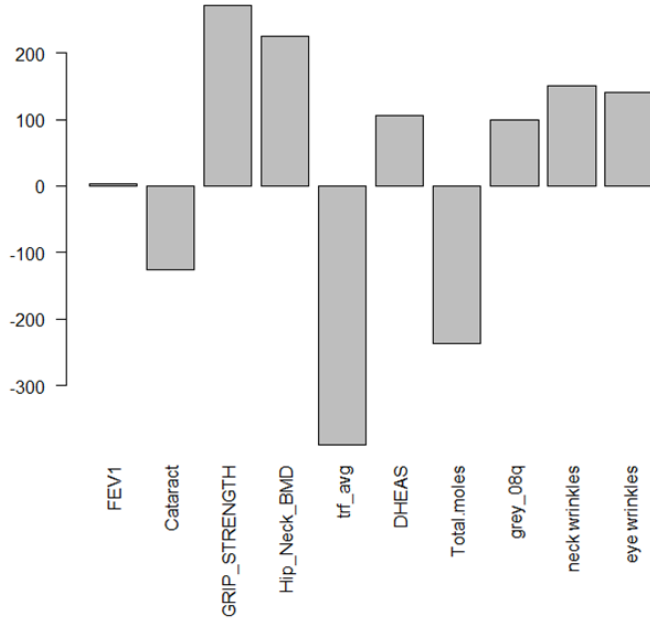


Figure 5: Bayes factors between non-linear models with the input as (learnt) physiological age compared to chronological age.

strength, bone mineral density and DHEAS have positive Bayes factors, implying that learnt physiological age is more predictive for these variables than chronological age. The binary phenotypes grey hair, neck wrinkles and eye wrinkles also all have positive Bayes factors.

### 2.3 Association to gene expression variation

If our measure of physiological age is more biological meaningful than chronological age, then we might expect to find more significant associations to low level molecular phenotypes such as gene expression microarray measurements. For around 900 of the individuals in Twins UK cohort we have gene expression measurements in three tissues: fat, blood (specifically LCLs) and skin. To test for association we use a linear mixed model of the form

$$[\text{expression}] = \underbrace{W \times [\text{age}] + B \times [\text{confounders}]}_{\text{fixed effects}} + \underbrace{M \times [\text{family ID}]}_{\text{random effects}} \quad (20)$$

where *age* will be physiological or chronological, *confounders* includes technical confounders such as batch number, and *family ID* is simply a factor shared by each twin pair. The number of significant associations (separated into up and down regulation) in each tissue for both physiological and chronological age is shown in Table 3. The False Discovery Rate was controlled to 0.05 using permutation testing. For skin and fat, where significant associations are found, physiological age always gives more associations than chronological age.

## 3 Discussion

We have proposed a novel model for inferring an individual’s physiological age from multiple age related phenotypes, including a straightforward extension which can be used when the linearity assumption is problematic. We have some encouraging results: our measure of physiological age is a better predictor of various phenotypes than chronological age, even when learnt without those phenotypes, and is significantly associated with more gene expression probes than chronological

Age	Regulation	Fat	Blood	Skin
Physiological	Down	297	0	1134
Chronological	Down	211	1	1093
Physiological	Up	306	0	967
Chronological	Up	207	0	908

Table 3: Counts of gene expression probes found to have significant association (at an FDR of 0.05) to physiological or chronological age, separated by up or down regulation.

age. More discouraging is that some phenotypes which are clearly correlated with age are actually predicted more poorly by physiological age.

This suggests several ideas for future research. If the problem is that the assumption of global systemic ageing is ill conceived, then maybe learning different  $\Delta$  values for different organ systems could be considered. This could be done in an automated way or using biomedical prior knowledge. Which model fits the data best here is of fundamental relevance to understanding the ageing process: e.g. is there such a phenomenon as accelerated global systematic ageing?

Another potential problem is that outliers in one phenotype have an unreasonably strong effect (leverage) on the final value of  $\Delta$ . Using a heavy tailed likelihood such as a Student-t distribution would be a straightforward extension to address this problem which would not require significant changes to the inference procedure.

From a medical perspective, understanding ageing is clearly pertinent given the ageing Western population. Learning what factors, environmental or systemic, encourage “successful”, healthy ageing is essential.

Finally, it could be of great medical interest to perform a GWAS on  $\Delta$  to discover causative SNPs which result in different rates of ageing.

## References

- [1] L.A. Hannah, D.M. Blei, and W.B. Powell. Dirichlet process mixtures of generalized linear models. *Arxiv preprint arXiv:0909.5194*, 2009.
- [2] David Knowles, Leopold Parts, Daniel Glass, and John M. Winn. Modeling skin and ageing phenotypes using latent variable models in infer.net. In *Predictive Models in Personalized Medicine Workshop, NIPS 2010, 6-11 December 2010, Vancouver, BC, Canada.*, 2010.
- [3] T. P. Minka, J. M. Winn, J. P. Guiver, and D. A. Knowles. Infer.NET 2.4, 2010. Microsoft Research Cambridge. <http://research.microsoft.com/infernet>.
- [4] Tim D. Spector and Alex J. MacGregor. The st. thomas’ uk adult twin registry. *Twin Research*, 5:440–443(4), 1 October 2002.
- [5] John Winn, Christopher M. Bishop, and Tommi Jaakkola. Variational message passing. *Journal of Machine Learning Research*, 6:661–694, 2005.