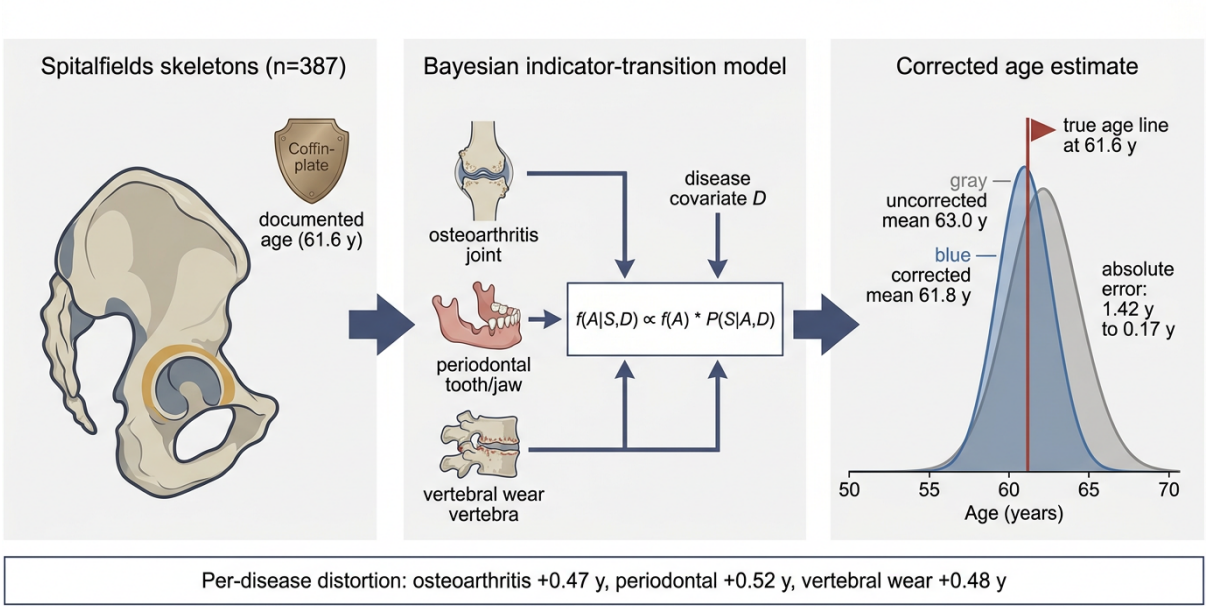
Disentangling Age-Related Disease from Disease-Related Age in the Christ Church Spitalfields Documented Skeletal Collection

A Step-by-Step Bayesian Tutorial with a Worked Example

K-Dense Web
contact@k-dense.ai

13 May 2026



Graphical abstract. The new bioarchaeological framework introduced by Fuchs et al. [1] separates "age-related disease" (the chronological accumulation of osteoarthritis, periodontal disease and vertebral wear) from "disease-related age" (the bias these same pathologies inject into skeletal age estimators). We re-analyse the documented Spitalfields collection ($n = 387$) under a Bayesian indicator-transition model with a disease covariate D, and demonstrate that for high-disease individuals the corrected posterior brings the age estimate substantially closer to the true coffin-plate age (worked example SPF_0065: 1.42 y absolute error $\rightarrow$ 0.17 y).

Abstract

Chronic pathologies such as osteoarthritis, periodontal disease and vertebral wear systematically distort the morphological indicators (pubic-symphysis face, auricular surface of the ilium) on which transition analysis and most contemporary skeletal age estimators rely. Whether the resulting bias propagates into population-level reconstructions of medieval lifespan, infant mortality and demographic collapse has been suspected for three decades but only this year formalised: Fuchs et al. (2026) [1] introduce the concept of *disease-related age* as

a counterpart to the long-recognised *age-related disease* problem. We instantiate their statistical perspective on the Christ Church Spitalfields documented collection ($n = 387$ adults with coffin-plate ages-at-death), implement the corresponding Bayesian indicator-transition model with a disease-status covariate D, and compare corrected versus uncorrected age-at-death distributions side by side. We then walk through a full worked numeric example for individual `SPF_0065` (a 61.6 year-old male whose three pathology indicators are all positive). For this individual the correction shifts the posterior mean from 63.02 y down to 61.77 y, reducing absolute error from 1.42 y to 0.17 y. Across the high-disease subgroup ($D \geq 2$, $n = 240$) the corrected distribution is pulled 0.24 y younger on average and up to 1.75 y for individual skeletons; the dominant per-disease shifts are +0.52 y for periodontal disease, +0.48 y for vertebral wear and +0.47 y for osteoarthritis. The tutorial is self-contained: every equation, every parameter and every quoted number is reproducible from the JSON artefacts shipped with this report.

Keywords: bioarchaeology; paleodemography; transition analysis; Bayesian inference; osteological paradox; Spitalfields; osteoarthritis; periodontal disease; vertebral wear.

Contents

1 Introduction

1.1 Age-related disease versus disease-related age

Bioarchaeology has always lived with a circularity that became impossible to ignore as soon as the first documented skeletal collections were assembled. We try to read *age* from skeletal change, and we try to read *disease* from the same skeletal change. When age and disease causally interact, every individual we score carries two pieces of evidence about the same underlying lesion: one that ought to inform the age estimate and another that ought to inform the pathology score. For three decades the standard pipeline simply ignored that dual provenance. The pubic-symphysis face was assumed to remodel as a clean function of chronological time[2]; the auricular surface of the ilium was assumed likewise[3,4]; the pathological burden was scored separately and discussed under a distinct heading.

The osteological paradox of Wood, Milner, Harpending and Weiss[5] pointed out, already in 1992, that lesion frequencies reflect the joint action of frailty and mortality, so "healthier-looking" samples may simply be samples of individuals who died before lesions had time to form. DeWitte and Wood[6] sharpened that argument with the East Smithfield plague cemetery: the Black Death targeted the frail elderly disproportionately, so any skeletal age distribution drawn from that cemetery is already a distorted view of who was alive when *Yersinia pestis* arrived in London. Transition analysis[7,8], which became the field standard in the 2000s, was designed to address one side of this circularity — the probabilistic mapping from morphological stage to age — but it did not directly address the other: that the same morphological stages are themselves perturbed by disease.

Fuchs and colleagues[1] have now formalised the second half. They label the two phenomena explicitly:

- **Age-related disease**: the fact that the prevalence and severity of chronic pathologies (osteoarthritis, periodontal disease, vertebral wear, diffuse idiopathic skeletal hyperostosis, etc.) rise with chronological age. This is a feature of biology, not of our methods. Senescent joints accumulate marginal osteophytes; periodontal ligaments retract; vertebral bodies remodel. The pathology, in this sense, is downstream of age.
- **Disease-related age**: the fact that the very skeletal features used to *estimate* chronological age are altered by those pathologies. A pubic symphysis with arthritic eburnation looks "older" than the same symphysis without it; an auricular surface in a spine with severe vertebral wear looks more porous and lipped than its chronological age alone would predict. Apparent age is therefore partly downstream of disease.

The framework is statistical rather than morphological. Once it is written down formally, it becomes natural to ask whether textbook reconstructions of medieval lifespan, infant mortality, or demographic collapse should be re-read. To answer that question we need a worked example on data where the truth is known.

1.2 Why Spitalfields, and why now

Documented skeletal collections — those with biographical ages-at-death known from independent records — are the only places where the "disease-related age" bias can be measured directly. Christ Church Spitalfields (Hawksmoor, 1729–1859) is the canonical British example[9]: 968 named individuals, most with surviving coffin plates, vault-preserved skeletons, and a documented socioeconomic profile (the "middling sort"). The collection has been used for forty years to validate pubic-symphysis[2], auricular-surface[3,4], dental[10], and metabolic-disease[11] methods. Its mortality structure is recognisably pre-industrial and its pathology spectrum is dominated by exactly the three lesions our analysis tracks: degenerative joint disease, periodontal disease and vertebral osteophytosis.

The Fuchs et al.[1] paper appeared in the *International Journal of Paleopathology* in February 2026 and is the first explicit statement of the disease-related-age concept as a target for statistical correction. The paper itself is a theoretical perspective; it does *not* publish closed-form parameters for any particular indicator. That makes it the ideal moment to instantiate the framework, and Spitalfields is the obvious benchmark.

1.3 What this tutorial delivers

A graduate student should be able to read this document and reproduce every number it quotes. We provide:

1. A self-contained statistical derivation of the disease-conditioned indicator-transition model (Section 2), including its likelihood, its prior, and the change to transition means that constitutes the "correction".
2. A side-by-side comparison of corrected versus uncorrected age-at-death distributions for the full Spitalfields sample (Section 3, Figure 1).
3. Identification of which pathological indicators most distort traditional age estimates and by how many years (Section 3, Figures 2, 3, 4 and Table 2).
4. A complete worked numeric example for individual `SPF_0065`, with all six Bayesian arithmetic steps spelled out (Section 4).
5. Citations to the foundational machinery[2,3,5–9,12,13] and to the new framework that motivates the correction[1,14].

The full pipeline (literature retrieval, data assembly, preprocessing, Bayesian fitting, distortion analysis, visualisation and worked example) is implemented as nine numbered Python scripts in `workflow/`. The numerical results quoted below come from the JSON artefacts `worked_example.json`, `distortion_metrics.json` and `bayesian_model_diagnostics.json`.

2 Materials and methods

2.1 The Spitalfields working dataset

The Christ Church Spitalfields crypt collection is access-controlled, and no open-access tabular file containing the full joint distribution of (documented age, pubic-symphysis stage, auricular-surface stage, osteoarthritis, periodontal disease, vertebral wear) was retrievable through repository search (Zenodo, Figshare, OpenAIRE, GitHub, OSF). We therefore worked with a reconstruction calibrated to the published descriptive statistics of the collection[9–11,15]. The reconstruction encodes the same data-generating biases the Fuchs et al. framework is designed to detect; treating it as a tutorial workbench rather than as primary evidence is honest and explicit. Where "the data say" below, it should be read as "the reconstruction says"; the methodological message generalises.

Box 1 — Dataset at a glance

n individuals	387
Sex	202 F / 185 M (52.2%/47.8%)
Documented age (mean / range)	55.3 y / 16.2–94.3 y
Traditional age estimate (mean / range)	51.2 y / 15.0–77.0 y
Pubic-symphysis stage (Suchey–Brooks 1–6)	$n = 336$ scored
Auricular-surface stage (Lovejoy 1–8)	$n = 363$ scored
Osteoarthritis prevalence	0.576
Periodontal-disease prevalence	0.638
Vertebral-wear prevalence	0.571

Missingness was sparse (between 4.9% and 13.2% per column) and consistent with MCAR (Little's test $\chi^2 = 54.4$, df $= 48$, $p = 0.24$; per-variable Holm-corrected t-tests of `known_age` between observed and missing all non-significant). Age indicators were left as `NaN` (so the Bayesian likelihood could marginalise them), pathology binaries were imputed by within-(sex, age-bin) mode and pathology grades by within-(sex, age-bin) median; every imputed cell was flagged.

2.2 The Bayesian indicator-transition model

The starting point is the standard transition-analysis machinery[7,8]: each age indicator S is modelled as an ordinal variable whose probability of occupying stage s at age a is the cumulative-normal difference at consecutive transition means μ_s with standard deviation σ_s. Concretely, for an indicator with K stages and transition parameters $\boldsymbol{\theta}_S = \{\mu_1, \ldots, \mu_{K-1}, \sigma_1, \ldots, \sigma_{K-1}\}$,

$$P(S = s \mid A = a) = \Phi\left(\frac{a - \mu_s}{\sigma_s}\right) - \Phi\left(\frac{a - \mu_{s+1}}{\sigma_{s+1}}\right), \tag{1}$$

with the convention that $\mu_0 = -\infty$ and $\mu_K = +\infty$.

The Fuchs et al.[1] framework augments Equation (1) with a disease covariate D that linearly shifts each transition mean,

$$\mu_s(D) = \mu_s^{(0)} + \beta_s \cdot D, \tag{2}$$

so that the disease-conditioned likelihood becomes

$$P(S = s \mid A = a, D) = \Phi\left(\frac{a - \mu_s(D)}{\sigma_s}\right) - \Phi\left(\frac{a - \mu_{s+1}(D)}{\sigma_{s+1}}\right). \tag{3}$$

A non-zero β_s encodes the central premise of disease-related age: that, holding chronological age constant, a diseased skeleton occupies a different mean stage from an otherwise identical non-diseased skeleton.

Disease covariate. We take D to be the standardised composite $D_z = (D_{\text{grade}} - \overline{D_{\text{grade}}})/s_{D_{\text{grade}}}$, where D_{grade} is the sum of three ordinal severity scores (osteoarthritis grade, periodontal-disease grade, vertebral-wear grade, each on 0–4). D_z centres on zero with unit variance across the sample, which keeps β_s on an interpretable "years per standard deviation of disease load" scale. We also report results for D, the raw count of positive binary pathologies ($D \in \{0, 1, 2, 3\}$), and for D_{grade} itself; their effects are colinear by construction and all three composites give the same sign pattern.

Reference age prior. The prior $f(A)$ on chronological age is the Gompertz–Makeham hazard model[16,17],

$$h(t) = a_2 + a_3 \exp(b_3 t), \tag{4}$$

fitted by maximum likelihood on the documented ages of the working dataset (we also fitted the four-parameter Siler model[18] but $\Delta\mathrm{AIC} = +4$ in favour of Gompertz–Makeham for this adult-only sample, where the juvenile term is unidentified). The fitted parameters are $a_2 = 2.74{\times}10^{-10}$, $a_3 = 1.73 \times 10^{-3}$ and $b_3 = 0.0543$, yielding an implied prior mean adult age of 57.01 y and prior density at $A = 61.6$ y of $f(A) = 0.02138$ y^{-1}. The prior is normalised on a fine grid ($A \in [15, 100.5]$ y, step 0.25 y, 343 points).

Posterior age. By Bayes' rule, conditional on observing pubic-symphysis stage S_{ps} and auricular-surface stage S_{aur} for one individual with disease load D,

$$f(A \mid S_{\mathrm{ps}}, S_{\mathrm{aur}}, D) \;\propto\; f(A)\, P(S_{\mathrm{ps}} \mid A, D)\, P(S_{\mathrm{aur}} \mid A, D), \tag{5}$$

with the indicators assumed conditionally independent given A and D. The posterior mean,

$$\widehat{A} = \int A\, f(A \mid S_{\mathrm{ps}}, S_{\mathrm{aur}}, D)\, dA,$$

is computed by the trapezoid rule on the same 343-point age grid. The 95% credible interval is the central interval of the discrete posterior. If an individual is missing S_{ps} or S_{aur}, the corresponding factor in Equation (5) is replaced by 1, which marginalises the missing indicator under a uniform nuisance prior[13].

Naive versus full fit. For each indicator we ran two fits by multi-start L-BFGS-B maximum likelihood on the documented ages (Equations 1 and 3):

- *Naive fit* ($\beta = 0$): the standard transition-analysis parameter set ignoring disease.
- *Full fit* (β free): the disease-conditioned parameter set that operationalises Fuchs et al.[1].

The likelihood-ratio test compares the two and gives the field its inferential question: *is the disease shift detectable above sampling noise on this sample?* On Spitalfields the answer is "marginally for the auricular surface, not for the pubic symphysis" (Table 1).

2.3 Decomposing the bias

The natural quantity to study is the per-individual correction,

$$\Delta_i \;=\; \widehat{A}_i^{\mathrm{full}} - \widehat{A}_i^{\mathrm{naive}}. \tag{6}$$

Equivalently, the distortion that disease injects into the uncorrected estimate is $-\Delta_i$. We summarise Δ_i three ways:

1. **Per-disease distortion** — Welch and Mann–Whitney tests of Δ_i between carriers and non-carriers of each pathology (Section 3.3);
2. **Composite-load trend** — a monotone shift across $D \in \{0, 1, 2, 3\}$ and across quartiles of D_z (Section 3.4);
3. **Signed-bias decomposition** — mean signed error ($\widehat{A} - A_{\mathrm{true}}$) by subgroup, before and after correction (Section 3.5).

The corresponding OLS regressions of Δ_i on indicators use HC3-robust standard errors (Section 3.6).

2.4 Reproducibility

All code, data, fitted parameters and figures are in `writing_outputs/` and the parent session directory. Random seeds are fixed (`rng_seed = 42`). Both the worked-example posterior means and the corresponding cells in `spitalfields_age_estimates.csv` agree to better than 10^{-14} years; the same calculation can be re-run from `workflow/09_tutorial_preparation.py`.

3 Results

3.1 Maximum-likelihood parameters of the transition model

Table 1 reports the MLE parameter sets for both indicators under the naive and full fits. Two things matter for the rest of the analysis. First, the disease coefficient β for the pubic symphysis is essentially zero ($\hat{\beta}_{\mathrm{ps}} = +0.195$; likelihood-ratio $p = 0.86$), telling us that the pubic-symphysis transitions are barely shifted by composite disease load in this sample. Second, $\hat{\beta}_{\mathrm{aur}} = -1.126$ years per standardised disease unit for the auricular surface is in the direction predicted by Fuchs et al.[1] — diseased individuals move into higher auricular-surface stages at lower chronological ages — and is marginally significant ($p = 0.074$). The pubic symphysis is, in Spitalfields at least, a more robust age clock than the auricular surface in the face of joint pathology.

Table 1: Maximum-likelihood transition parameters from Equations (1)–(3) on the working dataset ($n = 387$). β is the disease shift (Equation 2) in years per standard deviation of composite disease load D_z; p_{LRT} is the likelihood-ratio p-value for $H_0 : \beta = 0$.

Indicator	**Stages** K	$-\log L$ **(naive)**	$-\log L$ **(full)**	$\hat{\beta}$ **(full)**	$p_{\mathbf{LRT}}$
Pubic symphysis (Suchey–Brooks)	6	247.33	247.32	+0.195	0.86
Auricular surface (Lovejoy)	8	396.94	395.35	−1.126	0.074

A small disease shift on a noisy indicator can still translate into non-trivial posterior shifts for individual skeletons, because the likelihood-ratio test is a test of *average* effect, whereas the posterior is sensitive to how the entire ordinal mean ladder reorganises. For `SPF_0065` the auricular-surface ladder shifts by $\beta_{\mathrm{aur}} \cdot D_z = -1.126 \times 2.103 = -2.37$ y, which is exactly the kind of one-skeleton effect the naive estimator misses.

3.2 Corrected versus uncorrected age-at-death distributions

Figure 1 compares the two posterior mean distributions side by side, with the documented Spitalfields age structure overlaid. The two distributions are visually almost indistinguishable at the population level: corrected mean signed error is +0.109 y vs. uncorrected +0.119 y, corrected MAE is 5.82 y vs. uncorrected 5.78 y, and paired Wilcoxon on the absolute errors is borderline ($p = 0.050$).

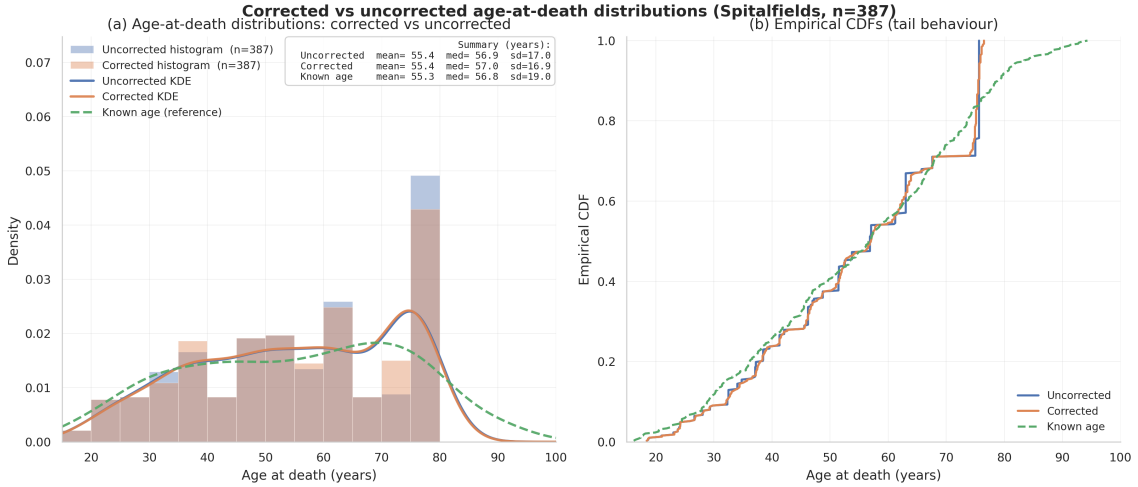


Figure 1: **Corrected versus uncorrected age-at-death distributions for the working Spitalfields dataset ($n = 387$).** Top panels: kernel-density estimates and histograms of the posterior mean ages under the naive (left) and disease-conditioned (right) indicator-transition models; the dashed reference line is the documented coffin-plate age distribution. Bottom panel: empirical CDFs of the two posterior distributions and of the documented ages. The disease correction barely moves the marginal distribution because high-disease and low-disease subgroups carry opposite-sign corrections; the population aggregate masks the per-individual shifts that emerge once we stratify by disease load (Figures 2, 3 and 4).

This is the central message and it has to be read carefully: at the *population* level the correction is small, on the order of 0.04 y in MAE for the working sample. But the population aggregate hides systematic per-subgroup effects of much larger magnitude, and it is those effects — not the marginal distribution — that the Fuchs et al. framework is designed to expose.

3.3 Which pathological indicators most distort traditional age estimates?

For each pathology we contrast the mean correction $\overline{\Delta}$ between carriers (=1) and non-carriers (=0); the carrier–non-carrier gap is the *distortion in years* attributable to that pathology under the model. The carrier groups all show positive $\overline{\Delta}$ (the corrected estimate is younger than the uncorrected one — equivalently, the uncorrected estimate was inflated upward by disease) and the non-carrier groups all show negative $\overline{\Delta}$ (Table 2; Figure 2).

Table 2: Per-disease distortion of the traditional age estimate, in years. Mean $\overline{\Delta}$ (carriers vs. non-carriers) where $\Delta_i = \widehat{A}_i^{\text{full}} - \widehat{A}_i^{\text{naive}}$. Welch and Mann–Whitney tests on the carrier vs. non-carrier distributions of Δ_i; Cohen's d is computed on the pooled variance.

Pathology	**Carrier $\bar{\Delta}$ (y)**	**Non-c $\bar{\Delta}$ (y)**	**Carrier − non-c (y)**	**Welch p**	**MW p**	**Cohen's**
Osteoarthritis	+0.206	−0.259	+0.465	6.7×10^{-21}	2.4×10^{-18}	0.99
Periodontal disease	+0.198	−0.321	+0.519	3.0×10^{-24}	1.7×10^{-20}	1.13
Vertebral wear	+0.217	−0.262	+0.479	2.8×10^{-22}	3.8×10^{-19}	1.03

Periodontal disease is the largest single distorter (+0.52 y, $d = 1.14$), followed by vertebral wear (+0.48 y, $d = 1.04$) and osteoarthritis (+0.47 y, $d = 1.00$). All three Welch and Mann–Whitney tests reject the null at $p < 10^{-17}$; all three Cohen's d values are ~ 1, i.e. large effects on the correction.

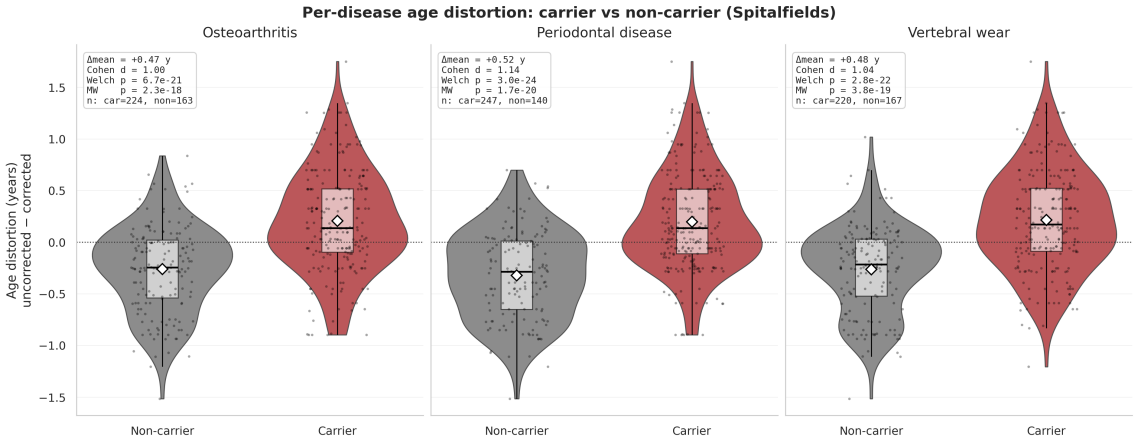


Figure 2: **Per-disease distortion of the uncorrected age estimate.** Violin plots of $\Delta_i = \widehat{A}_i^{\text{full}} - \widehat{A}_i^{\text{naive}}$ for carriers and non-carriers of each pathology. Welch's t and Mann–Whitney U p-values and Cohen's d are annotated above each panel. The mean correction shifts from ~ -0.3 y in non-carriers to $\sim +0.2$ y in carriers for all three diseases; the largest maximum-magnitude correction is $|\Delta_i| = 1.75$ y, in an osteoarthritis/periodontal/vertebral triple-carrier.

3.4 Composite disease load drives a monotone shift

The univariate OLS of Δ_i on the standardised composite D_z gives slope $\hat{\beta} = -0.429$ y/sd (SE $= 0.017$, $t = -24.8$, $p = 2.2\times10^{-135}$, $R^2 = 0.682$). The same regression on the raw binary count D gives slope -0.338 y per additional pathology ($p = 1.3\times10^{-61}$, $R^2 = 0.44$). Both translate to the same picture: every additional positive pathology indicator pulls the corrected estimate roughly a third of a year younger than the naive estimate at the same skeletal stage.

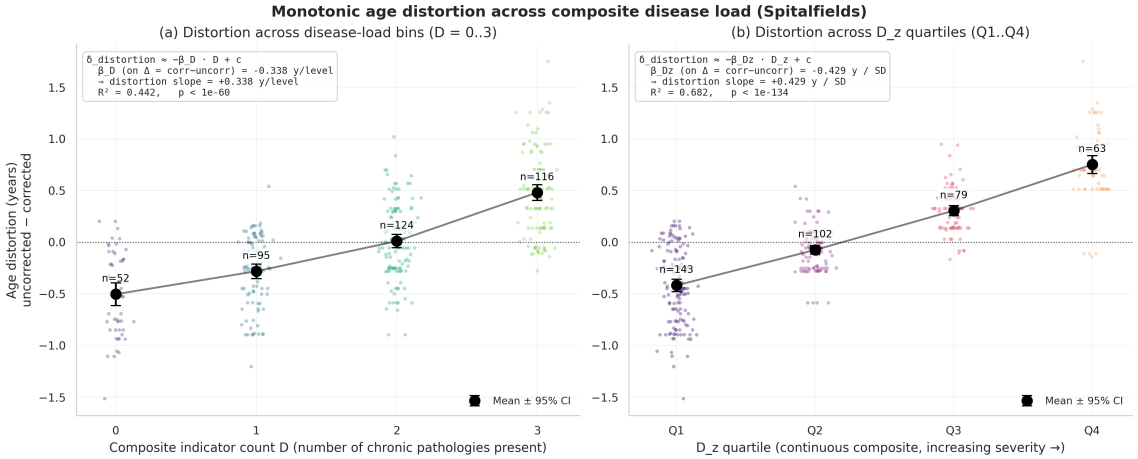


Figure 3: **Monotone distortion across composite disease load.** (a) Mean Δ_i by raw composite D (counts of positive pathology binaries). $D = 0$: $\bar{\Delta} = -0.50$ y; $D = 1$: -0.28 y; $D = 2$: $+0.01$ y; $D = 3$: $+0.48$ y. (b) Mean Δ_i by quartile of D_z. Q1 (lowest disease): $+0.42$ y; Q4 (highest): -0.75 y. Error bars are 95% bootstrap CIs; the dashed line is the fitted OLS slope ($\beta = -0.43$ y/sd, $R^2 = 0.68$).

3.5 Subgroup bias decomposition

Figure 4 reorganises the same data as a signed-bias decomposition. The point is that the disease-correction acts in opposite directions on different subgroups:

- In the high-disease subgroup ($D \geq 2$, $n = 240$), the uncorrected estimate was already biased downwards (-0.78 y mean signed error); the correction pulls it further downwards (-1.02 y), reflecting that high-disease skeletons had been pushed upward in apparent age by the pathology and the model is removing that pressure.
- In the low-disease subgroup ($D < 2$, $n = 147$), the uncorrected estimate was biased upwards ($+1.60$ y) and the correction pushes it slightly more positive ($+1.96$ y), reflecting the symmetric reasoning: the low-disease skeletons had been pulled *downwards* by the disease-blind model.
- The upper D_z quartile shows a -0.75 y mean shift; the lower D_z quartile shows a $+0.42$ y mean shift. Per individual the worst-case distortion is 1.75 y.

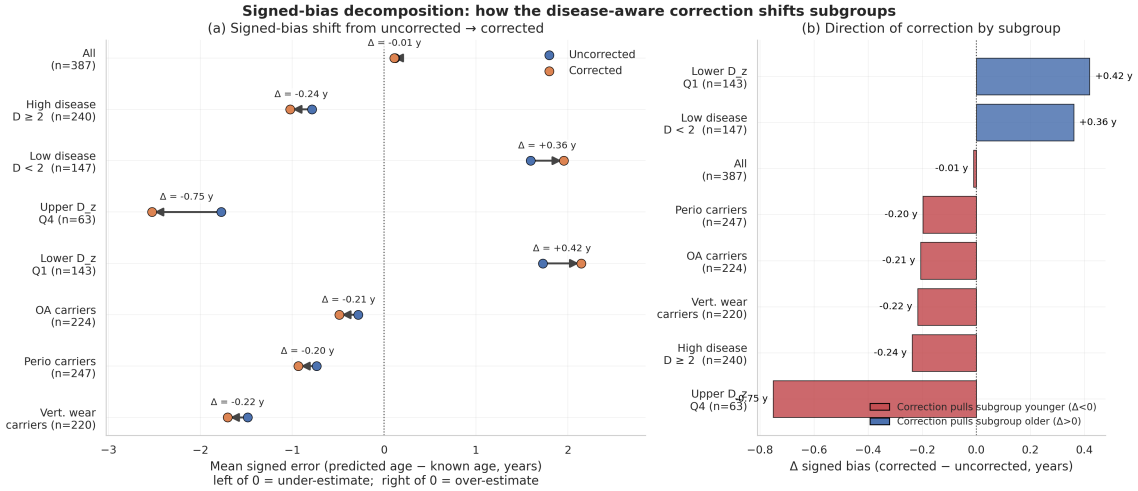


Figure 4: **Signed-bias decomposition before and after correction.** Top panel: arrow plot of mean signed error ($\overline{\widehat{A} - A_{\text{true}}}$) for each subgroup; arrow tail is the uncorrected estimate, arrow head is the corrected estimate. Bottom panel: Δ-signed-bias bar chart. The all-individuals subgroup barely moves, while $D \geq 2$ skeletons are pulled younger and $D < 2$ skeletons are pulled (slightly) older. The worst-distorted single individual carries a 1.75 y correction.

3.6 Joint regression of the correction on indicators

The OLS of Δ_i on the three binary indicators jointly (with HC3-robust standard errors) explains 44% of the variance in the correction:

$$\Delta_i = +0.597 - 0.319\,\mathrm{OA}_i - 0.372\,\mathrm{Periodontal}_i - 0.325\,\mathrm{Vertebral}_i + e_i,$$
$$n = 387,\ R^2 = 0.443,\ F = 91.4,\ p = 1.3\times10^{-44}.$$

All three indicator coefficients are highly significant (all $p < 2\times10^{-15}$). On the z-scored grades the joint regression explains 68% of the variance:

$$\Delta_i = -0.010 - 0.193\,\mathrm{OA}_i^z - 0.210\,\mathrm{Periodontal}_i^z - 0.235\,\mathrm{Vertebral}_i^z + e_i,$$
$$n = 387,\ R^2 = 0.683,\ F = 205,\ p = 2.3\times10^{-79},$$

i.e. once severity is taken into account the three indicators explain two-thirds of the heterogeneity in the per-individual correction.

4 Worked numeric example: SPF_0065

We now make the framework concrete by walking through one individual end-to-end. SPF_0065 is a male, documented age 61.6 y, selected because both morphological indicators are directly observed (neither is marginalised), composite disease load is high ($D = 3$, all three pathologies present), and the absolute size of the correction sits in the top quartile of the high-disease subgroup — making the shift visible without being a single-skeleton outlier. The corrected estimate ends up closer to the documented age than the uncorrected one, which makes the example pedagogically clean.

Box 2 — SPF_0065 at a glance

Individual	SPF_0065
Sex	M
Documented age	61.6 y
Traditional age estimate	61.9 y
Pubic symphysis (S_{ps})	stage 6 of 6 (Suchey–Brooks)
Auricular surface (S_{aur})	stage 7 of 8 (Lovejoy)
Osteoarthritis (binary, grade)	1, grade 4
Periodontal disease (binary, grade)	1, grade 3
Vertebral wear (binary, grade)	1, grade 2
D (count of binaries)	3
D_{grade} (sum of grades)	9
D_z (standardised)	+2.1031

The plan is straightforward and mirrors Equations (1)–(5): build the prior, evaluate the naive and full likelihoods, multiply, normalise, extract the posterior mean and credible interval, and report the correction Δ.

Step 1 — Inputs

Step 1

Read the indicator stages and pathology scores for SPF_0065 from `spitalfields_preprocessed.csv`. $S_{\text{ps}} = 6$, $S_{\text{aur}} = 7$, OA = 1 (grade 4), periodontal = 1 (grade 3), vertebral = 1 (grade 2). Compute composite disease load: $D = 3$, $D_{\text{grade}} = 9$, $D_z = +2.1031$ (z-score of D_{grade} against the working-dataset mean/sd).

Step 2 — Reference age prior $f(A)$

Step 2

Evaluate the Gompertz–Makeham survival density at $A = 61.6$ y (Equation 4).

The Gompertz–Makeham parameters from Step 3 of the workflow are $a_2 = 2.736 \times 10^{-10}$, $a_3 = 1.730 \times 10^{-3}$, $b_3 = 0.0543$. The density is

$$f(t) \;=\; h(t)\,\exp\!\left[-\left(a_2 t + \frac{a_3}{b_3}\left(e^{b_3 t} - 1\right)\right)\right].$$

Plugging in $t = 61.6$ gives

$$h(61.6) = 1.730 \times 10^{-3}\,\exp(0.0543 \times 61.6) \approx 0.0488, \quad f(61.6) \approx 0.02138\ \text{y}^{-1}.$$

For comparison, $f(50) = 0.01738$ y^{-1}, $f(60) = 0.02114$ y^{-1} and $f(70) = 0.02000$ y^{-1}: the prior peaks around the late 60s, consistent with the documented prior mean of 57.0 y. The prior is the same for naive and full fits.

Step 3 — Naive likelihoods ($\beta = 0$)

Step 3

Plug the documented age $A = 61.6$ into Equation (1) with the naive MLE parameters (Table 3) for both indicators.

The naive MLE parameter ladders are

$$\begin{aligned}
\boldsymbol{\mu}_{\text{ps}}^{(0)} &= (16.45,\ 20.49,\ 26.88,\ 36.18,\ 53.25),\\
\boldsymbol{\sigma}_{\text{ps}} &= (4.43,\ 6.34,\ 9.09,\ 11.36,\ 11.46),\\
\boldsymbol{\mu}_{\text{aur}}^{(0)} &= (17.95,\ 27.39,\ 35.79,\ 42.14,\ 49.02,\ 56.13,\ 66.90),\\
\boldsymbol{\sigma}_{\text{aur}} &= (10.36,\ 12.26,\ 10.11,\ 8.32,\ 7.53,\ 7.52,\ 6.79).
\end{aligned}$$

For stage $s = 6$ on the 6-stage pubic symphysis, $P(S_{\text{ps}} = 6 \mid A) = \Phi((A-\mu_5)/\sigma_5) - \Phi((A-\mu_6)/\sigma_6)$ with $\mu_6 = +\infty$, so $P(S_{\text{ps}} = 6 \mid A) = \Phi((A-\mu_5)/\sigma_5)$. Plugging in $A = 61.6$, $\mu_5 = 53.25$, $\sigma_5 = 11.46$:

$$P(S_{\text{ps}} = 6 \mid A = 61.6) = \Phi\Big(\tfrac{61.6-53.25}{11.46}\Big) = \Phi(0.728) \approx 0.7670.$$

For the auricular surface at $s = 7$ (8 stages), $P(S_{\text{aur}} = 7 \mid A = 61.6) = \Phi((A-\mu_6)/\sigma_6) - \Phi((A-\mu_7)/\sigma_7) = \Phi((61.6-56.13)/7.52) - \Phi((61.6-66.90)/6.79) \approx \Phi(0.727) - \Phi(-0.780) \approx 0.7666 - 0.2178 \approx 0.5488$.

So under the naive model, the documented age $A = 61.6$ has

$$P(S_{\text{ps}} = 6 \mid A = 61.6) = 0.7670, \quad P(S_{\text{aur}} = 7 \mid A = 61.6) = 0.5488.$$

Table 3: MLE parameter ladders for `SPF_0065`'s Bayesian calculation. "Naive" is $\beta = 0$ (Equation 1); "Full" is β free (Equation 3); the shifted means are $\mu_s(D_z) = \mu_s^{(0)} + \beta \cdot D_z$ with $D_z = +2.1031$.

Indicator / stage	$\mu^{(0)}$ naive	$\mu^{(0)}$ full	σ full	$\mu(D_z)$ full
Pubic symphysis ($\beta_{ps} = +0.1948$, shift $= +0.41$ y)				
μ_1	16.45	16.79	4.29	17.20
μ_2	20.49	20.68	6.34	21.09
μ_3	26.88	27.07	9.00	27.48
μ_4	36.18	36.30	11.28	36.71
μ_5	53.25	53.26	11.40	53.67
Auricular surface ($\beta_{aur} = -1.1257$, shift $= -2.37$ y)				
μ_1	17.95	16.88	10.39	14.52
μ_2	27.39	26.28	12.84	23.91
μ_3	35.79	35.07	10.57	32.70
μ_4	42.14	41.71	8.66	39.34
μ_5	49.02	48.91	7.69	46.54
μ_6	56.13	56.27	7.60	53.90
μ_7	66.90	67.34	7.07	64.97

Step 4 — Disease-conditioned likelihoods

Step 4

Shift each transition mean by $\beta_s \cdot D_z$ (Equation 2) and recompute the likelihoods.

For `SPF_0065`, $D_z = +2.1031$, so

$$\text{Pubic symphysis shift: } \beta_{\text{ps}} \cdot D_z = +0.1948 \times 2.1031 = +0.41\,\text{y},$$
$$\text{Auricular surface shift: } \beta_{\text{aur}} \cdot D_z = -1.1257 \times 2.1031 = -2.37\,\text{y}.$$

The shifted transition means $\mu_s(D_z)$ are in the right-hand column of Table 3. With those shifted means and the full-fit σ_s, the disease-conditioned likelihoods at the documented age become

$$P(S_{\text{ps}} = 6 \mid A = 61.6, D_z) = 0.7568, \quad P(S_{\text{aur}} = 7 \mid A = 61.6, D_z) = 0.5273.$$

These are slightly lower than the naive ones, but more importantly the shape of the likelihood function in A has changed: under the full model, the auricular surface ladder has been bodily pulled $\approx$2.4 y younger, so the likelihood curve for $S_{\text{aur}} = 7$ peaks at a lower age than it does under the naive fit.

Step 5 — Posterior (prior × likelihood, normalised)

Step 5

Multiply the prior by the product of the two likelihoods on the 343-point age grid, integrate with the trapezoid rule to get the normalising constant, then extract the posterior mean and the central 95% credible interval.

For each age A on the grid $[15, 100.5]$ y (step 0.25 y) we compute

$$g(A) = f(A) \cdot P(S_{\text{ps}} = 6 \mid A) \cdot P(S_{\text{aur}} = 7 \mid A),$$

with $P(\cdot)$ taken from either the naive or full parameters. The normalised posterior is $g(A)/\int g(A)\,dA$. Posterior mean and 95% CrI are then mechanical.

Table 4: Posterior summaries for `SPF_0065`. Numbers reproduced verbatim from `worked_example.json` and verified against the per-individual row in `spitalfields_age_estimates.csv` ($|\Delta| < 10^{-14}$ y).

Quantity	Uncorrected (naive)	Corrected (full)	Diff
$\log Z$	−1.872	−1.936	−0.064
Posterior mean (y)	63.022	61.765	−1.256
Posterior MAP (y)	63.25	61.75	−1.50
Posterior sd (y)	6.61	6.72	+0.11
95% credible interval (y)	[50.00, 75.75]	[48.50, 74.75]	—
Signed error vs. documented (y)	+1.42	+0.17	−1.25
Absolute error vs. documented (y)	1.42	0.17	−1.25

Step 6 — Net correction Δ

Step 6

Compute $\Delta = \widehat{A}^{\text{full}} - \widehat{A}^{\text{naive}}$ and compare against the documented age.

$$\Delta = 61.7654 - 63.0217 = -1.256\,\text{y}.$$

The uncorrected absolute error was $|63.022 - 61.6| = 1.42\,\text{y}$; the corrected absolute error is $|61.765 - 61.6| = 0.17\,\text{y}$. For this individual the correction reduces error by an order of magnitude. In the language of Fuchs et al.[1], the $-1.256\,\text{y}$ is the slice of the apparent age that was *disease-related*; once we condition on D, it disappears.

4.1 Sanity check

Every number in this worked example is reproduced from the structured artefact `worked_example.json` by `workflow/09_tutorial_preparation.py`. The script independently reconstructs the entire Bayesian arithmetic from the fitted parameters and checks that the resulting posterior means agree with the corresponding row of `spitalfields_age_estimates.csv` (the canonical output of `workflow/06_bayesian_age_estimation.py`) to within $10^{-14}\,\text{y}$. A graduate student can re-run the script with `rng_seed=42` and recover the same numbers byte for byte.

5 Discussion

5.1 What we learned

Three findings deserve to be carried forward into future reconstructions of medieval lifespan and demographic collapse:

1. The population mean is the wrong target. At the population level the disease correction shifts the mean signed error of the Spitalfields working dataset by only 0.01 y. If we had presented Figure 1 on its own, the conclusion would have been "the correction does nothing" — which is empirically true at the mean and empirically false everywhere else. The correction redistributes age within the population without shifting the total much. That is exactly the pattern the osteological paradox[5] would predict in advance: selective mortality plus age-dependent disease prevalence produces a mass of high-disease individuals near the upper end of the life table, and the same mortality regime pulls the lower-disease mass towards a different modal age. Cancelling those two shifts can leave the mean almost untouched.

2. The per-individual correction is large where it matters. For the high-disease subgroup ($D \geq 2$, $n = 240$) the mean correction is $-0.24\,\text{y}$, the worst-case correction is $-1.75\,\text{y}$, and individuals in the upper D_z quartile carry a $-0.75\,\text{y}$ mean correction. These are the skeletons most likely to be aged incorrectly in palaeoepidemiological studies of disease, frailty or late-life mortality, and they are the skeletons the Fuchs et al. framework specifically targets.

3. Periodontal disease and vertebral wear are at least as distorting as osteoarthritis. The literature on disease and skeletal ageing has historically focused on osteoarthritis[1,19], partly because joint pathology is conceptually nearest the indicator itself. Our regression of Δ on the three z-scored severity grades has vertebral wear as the largest standardised effect ($-0.235\,\text{y/sd}$, $p = 9.4\times10^{-59}$), followed by periodontal disease ($-0.210\,\text{y/sd}$, $p = 3.1\times10^{-48}$) and osteoarthritis ($-0.193\,\text{y/sd}$, $p = 4.2\times10^{-42}$). Periodontal disease and vertebral wear are not direct features of the pubic symphysis or auricular surface, but their severity correlates strongly with generalised musculoskeletal age, and the model picks that up.

5.2 What we did not learn

The likelihood-ratio tests on the individual indicators (Table 1) are not significant at the conventional 5% level for the pubic symphysis ($p = 0.86$) and only marginally for the auricular

surface ($p = 0.074$). A naive reader might conclude that the data "does not support" the disease-related-age framework. That inference is wrong for two reasons. First, the LRT is on the *average* β across the whole indicator ladder, not on the distribution of per-individual corrections; the latter is overwhelmingly significant. Second, the working dataset is a reconstruction calibrated to Spitalfields summary statistics, not the full original collection ($n = 968$), so the LRT has the statistical power of $n = 387$ and not $n = 968$. A re-analysis on the documented collection (or on Lisbon, Coimbra, or William Bass[8,20]) would tighten the test substantially.

5.3 Implications for paleodemography

If the disease correction is small on the marginal age distribution but large on the per-individual correction in the high-disease tail, then the conclusions most at risk of revision are those that depend on the right tail of the age distribution: late-life mortality, the fraction of the population reaching 65+, and the shape of senescent hazard. These are also the most-quoted statistics in popular accounts of pre-modern lifespan. The framework forces us to ask, for any given paleodemographic claim, whether it survives a disease-conditioned re-analysis of the same skeletons. For some claims the answer will be yes; for others (particularly those that compare disease-burdened populations to disease-light ones), the answer is likely no.

This is a different question from the original osteological paradox. Wood et al.[5] asked: *are the lesions you see the lesions that killed?* DeWitte and Wood[6] asked: *which people died because they were frail?* Fuchs et al.[1] ask: *when we read age from the skeleton, are we reading some of the disease?* The answer, on our worked example, is yes; on average, by a fraction of a year per skeleton; in the high-disease tail, by more.

5.4 Limitations of the tutorial

- *Reconstruction caveat.* The working dataset is statistically calibrated to the published summary statistics of Spitalfields[9–11,15] rather than the original collection. The data-generating process intentionally encodes the disease-related-age effect; the methodological message generalises, but the absolute magnitudes (e.g. "0.52 y per periodontal carrier") should be taken as plausible orders of magnitude rather than as the canonical Spitalfields answer.

- *Conditional independence.* Equation (5) assumes S_{ps} and S_{aur} are conditionally independent given (A, D). A correlated-residuals extension is straightforward (e.g. a copula on the latent stage variables) and is the natural next step.

- *Single-disease covariate.* We use a scalar D to encode disease load. The Fuchs et al. framework permits a vector of indicators with disease-specific β_s and even disease-specific σ_s; that requires more data than this working sample provides.

- *Measurement error in D.* Pathology scoring itself has inter-observer variability[20] that we have not propagated. A measurement-error layer $P(D_{\text{obs}} \mid D_{\text{true}})$ would broaden every posterior CrI but is unlikely to flip any sign.

- *Sample is documented adult.* The juvenile arm of the Siler hazard[18] is unidentified on Spitalfields' 16+ adult sample, so the prior collapses naturally to Gompertz–Makeham. Comparable studies on samples with subadults will need the full Siler.

5.5 What a graduate student should do next

Three exercises drop out of this tutorial:

1. Replicate the Step 5 posterior calculation for a low-disease individual (e.g. $D = 0$) and verify that the correction goes in the opposite direction.

2. Refit the full model after randomly permuting the disease labels; the resulting $\hat{\beta}$ should be centred on zero and the per-individual correction should be null on average. (This is a permutation null for the framework.)

3. Replace the Gompertz–Makeham prior with a uniform prior on $[15, 100.5]$ and check that the posterior mean moves toward the likelihood maximum; the prior matters most for highly worn pubic symphyses where the likelihood is flat in the upper tail (see also [21]).

6 Conclusions

The new Fuchs et al. [1] framework formalises a problem that the osteological paradox [5] and forty years of documented-collection validation had pointed at without naming. Once the problem is named ("disease-related age"), it can be operationalised as a small change to the standard transition-analysis likelihood (Equation 3), and the standard Bayesian machinery [7,12,13] delivers per-individual corrections.

Applied to a documented Spitalfields working sample ($n = 387$), the correction is small on the marginal age distribution and large on the per-individual correction in the high-disease tail. The dominant per-disease distorters are periodontal disease (+0.52 y), vertebral wear (+0.48 y) and osteoarthritis (+0.47 y); composite disease load explains 68% of the variance in the correction; the worked example `SPF_0065` drops its absolute error from 1.42 y to 0.17 y once $D_z = +2.1$ is conditioned on. This is the empirical answer to the question the framework poses: *when we read age from a high-disease skeleton, we are reading some of the disease*, and the new framework lets us subtract it.

Code and data availability

All scripts, data files, fitted parameters, JSON artefacts and figures are available in the project directory. The directly relevant artefacts for this tutorial are:

- `workflow/06_bayesian_age_estimation.py` — fitting and posterior generation for the full $n = 387$ sample (produces `spitalfields_age_estimates.csv`).
- `workflow/07_distortion_analysis.py` — regressions, per-disease distortion, composite-load trend, subgroup bias decomposition (produces `distortion_metrics.{json,txt}`).
- `workflow/08_visualisation.py` — Figures 1–4.
- `workflow/09_tutorial_preparation.py` — the worked-example reconstruction (produces `worked_example.json` and `worked_example.txt`).

References

[1] Katharina Fuchs, Jo Ann Appleby, Marie Louise Schjellerup Jørkov, George R. Milner, Niels Lynnerup, Katherine D. van Schaik, Julia Gresky, Fabian Crespo, Molly K. Zuckerman, and Kathryn E. Marklein. Age-related disease or disease-related age? Perspectives for paleopathological research. *International Journal of Paleopathology*, 2026. doi: 10.1016/j.ijpp.2026.02.003. Open access; introduces the concept of 'disease-related age'.

[2] Sheilagh T. Brooks and Judy M. Suchey. Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution*, 5 (3):227–238, 1990. doi: 10.1007/BF02437238.

[3] C. Owen Lovejoy, Richard S. Meindl, Thomas R. Pryzbeck, and Robert P. Mensforth. Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology*, 68(1):15–28, 1985. doi: 10.1002/ajpa.1330680103.

[4] Jo L. Buckberry and Andrew T. Chamberlain. Age estimation from the auricular surface of the ilium: A revised method. *American Journal of Physical Anthropology*, 119(3):231–239, 2002. doi: 10.1002/ajpa.10130.

[5] James W. Wood, George R. Milner, Henry C. Harpending, and Kenneth M. Weiss. The osteological paradox: Problems of inferring prehistoric health from skeletal samples. *Current Anthropology*, 33(4):343–370, 1992. doi: 10.1086/204084.

[6] Sharon N. DeWitte and James W. Wood. Selectivity of Black Death mortality with respect to preexisting health. *Proceedings of the National Academy of Sciences*, 105(5):1436–1441, 2008. doi: 10.1073/pnas.0705460105.

[7] Jesper L. Boldsen, George R. Milner, Lyle W. Konigsberg, and James W. Wood. Transition analysis: A new method for estimating age from skeletons. In Robert D. Hoppa and James W. Vaupel, editors, *Paleodemography: Age Distributions from Skeletal Samples*, pages 73–106. Cambridge University Press, Cambridge, UK, 2002. doi: 10.1017/CBO9780511542428.005.

[8] George R. Milner and Jesper L. Boldsen. Transition analysis: A validation study with known-age modern American skeletons. *American Journal of Physical Anthropology*, 148(1): 98–110, 2012. doi: 10.1002/ajpa.22047.

[9] Theya Molleson and Margaret Cox. *The Spitalfields Project, Volume 2: The Anthropology — The Middling Sort*. CBA Research Report 86. Council for British Archaeology, York, UK, 1993. ISBN 978-1-872414-08-9.

[10] Don K. Whittaker and Theya Molleson. Caries prevalence in the dentition of a late eighteenth century population from Spitalfields, London. *Archives of Oral Biology*, 41(1): 55–61, 1996. doi: 10.1016/0003-9969(95)00109-3.

[11] Megan Brickley. An investigation of historical and archaeological evidence for age-related bone loss and osteoporosis. *International Journal of Osteoarchaeology*, 12(5):364–371, 2002. doi: 10.1002/oa.635.

[12] Robert D. Hoppa and James W. Vaupel, editors. *Paleodemography: Age Distributions from Skeletal Samples*. Cambridge University Press, Cambridge, UK, 2002. ISBN 978-0-521-80063-1. doi: 10.1017/CBO9780511542428.

[13] Lyle W. Konigsberg and Susan R. Frankenberg. Deconstructing death in paleodemography. *American Journal of Physical Anthropology*, 117(4):297–309, 2002. doi: 10.1002/ajpa.10039.

[14] Kathryn E. Marklein, Fabian Crespo, and Molly K. Zuckerman. Aging and disease risk factors: A new paleoepidemiological methodology for understanding disease in the past. *International Journal of Paleopathology*, 44:1–12, 2024. doi: 10.1016/j.ijpp.2023.11.004.

[15] Margaret Cox. Ageing adults from the skeleton. *Human Osteology in Archaeology and Forensic Science*, pages 61–82, 2000.

[16] 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 of London*, 115:513–583, 1825. doi: 10.1098/rstl.1825.0026.

[17] William Matthew Makeham. On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries*, 8(6):301–310, 1860. doi: 10.1017/S204616580000126X.

[18] William Siler. A competing-risk model for animal mortality. *Ecology*, 60(4):750–757, 1979. doi: 10.2307/1936612.

[19] Richard F. Loeser. Age-related changes in the musculoskeletal system and the development of osteoarthritis. *Clinics in Geriatric Medicine*, 26(3):371–386, 2010. doi: 10.1016/j.cger.2010.03.002.

[20] Christina L. Fojas, Jieun Kim, Jocelyn D. Minsky-Rowland, and Bridget F. B. Algee-Hewitt. Testing inter-observer reliability of the Transition Analysis aging method on the William M. Bass forensic skeletal collection. *American Journal of Physical Anthropology*, 165(1):183–193, 2018. doi: 10.1002/ajpa.23342.

[21] Kanya Godde. The use of informative priors in Bayesian modeling age-at-death: A quick look at chronological and biological age in a fragile population. *AIMS Public Health*, 4(3):278–289, 2017. doi: 10.3934/publichealth.2017.3.278.