A linked mixture model of coronary atherosclerosis

Bret Beheim[*]

Department of Human Behavior, Ecology and Culture,
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Keywords: Agatston Scores, Coronary Calcium, Bayesian Methods, Generative Inference

I. INTRODUCTION

The coronary calcium (CAC) score was first proposed by Agatston, *et al.* [1] to quantify the amount of calcified plaque in the coronary arteries, and remains the standard metric for assessing arterial aging [2, 3]. Three decades of research with large cohorts of patients, in both cross-sectional and longitudinal designs, has shown a patient's CAC score to be highly predictive of subsequent major coronary events [4–8], even among patients with few additional risk factors [9, 10]. Many studies have found CAC varies systematically by nationality [11, 12] and ethnicity [13–16], with some populuations showing very little CAC even into advanced age [17], leading to debates over whether atherosclerosis itself is a consequence of sedentary, industrialized lifestyles [18, 19].

Despite the centrality of the CAC score in the study of atherosclerosis, its significance in clinical practice is ambiguous [20, 21], and it lacks a consistent statistical model from study to study. A given analysis may employ ordered-categorical [8, 22–24], linear [25, 26], log-linear [21, 27–29], quantile [30, 31] and proportional hazards [32–34] regression models. With some exceptions, e.g. [26, 30, 35–38], the CAC score is almost always modelled as a cross-sectional value that varies between patients, rather than as an atherosclerotic growth process within an individual.

To better connect the CAC score to the dynamic physiology of arterial aging, we propose a simple growth model for the onset and progression of calcified arterial plaques. Using high-throughput computational simulations to generate 540 realistic CAC studies (360 cross-sectional and 180 longitudinal studies, 113,760 patients, 268,200 observations), this model can identify structural mis-specifications in many of the common log-linear models of CAC, and indicate specific statistical modifications which dramatically improve the predictive accuracy of CAC modelling across the life course.

———————
[*] bret_beheim@eva.mpg.de

II. METHODS

A generative model of CAC growth

Although CAC is modelled in the current literature described above by a large variety of methods, we can synthesize two key empirical details from which to build a generative model [39]. First, calcified plaques typically appear in the coronary artery in middle age, and before a patient's age of onset, they will show a CAC score of 0 if given a computerized tomography scan. After the age of onset, an individual's CAC steadily accumulates at one or more plaque sites in the coronary artery, increasing at a non-linear pace set in part by variation in genetics, behavior, or risk exposures [9, 26, 35, 40], with CAC measurements of longitudinal study participants increasing between 14.3% and 29.7% annually [41, 42]. Given this, let us assume that the rate of growth in CAC at a given moment t after the age of onset t_0 is proportional to both immediate risk factors and the current amount of coronary calcium, that is,

$$\frac{dy}{dt} = y(t)\,k(t), \tag{1}$$

for CAC score $y(t)$ and positive growth function $k(t)$. The CAC score is thus given by the integral of all growth velocities since the age of onset t_0,

$$y(t) = \int_{t_0}^{t} y(x)\,k(x)\,dx. \tag{2}$$

In principle, we could estimate the shape of the rate function $k(t)$ as we observe the score y change with time. However, since we ordinarily do not have more than one CAC measurement per patient, we can instead appeal to the *average* growth rate between t_0 and t,

$$k = (t - t_0)^{-1} \int_{t_0}^{t} k(x)\,dx,$$

a constant. We can then solve Equation 1 as

$$\ln y(t) = k\,(t - t_0), \tag{3}$$

for any $t \geq t_0$. Since rate k describes an exponential growth process, it is useful to re-express k as a *doubling time*. That is, for any CAC score y at time t, we can describe the time $t + d$ at which the score is doubled,

which is true if $d = (\ln 2)/k$. Longitudinal studies of CAC report an average doubling time between 2.7 and 5.8 years [41, 42]. An example atherosclerosis trajectory and linear approximation following this approach for a single patient is given in Figure 1.

Allowing for a small amount of measurement error, we can estimate the average k for a set of observations by log-linear regression. For each observation j, with CAC score y_j, and patient age t_j,

$$y_j \sim \text{LogNormal}(\mu_j, \sigma),$$

$$\mu_j = k\,(t_j - t_0). \tag{4}$$

That is, each observation varies around some average CAC score μ_j with standard deviation σ due to measurement error and other unobserved factors. Given a dataset recording a single individual's CAC trajectory over time, and the log-linear regression model $\mu_j = a + bt_j$, we could estimate the model parameters in Equation 4 as $t_0 = (-a/b)$, $k = b$ and $d = \ln(2)/b$.

Structural mis-estimation of disease onset and progression

Studying CAC on the logarithmic scale was originally proposed by Agatston, *et al.* [1] and remains the most common statistical approach, commonly with $\ln(\text{CAC}+1)$ or $\ln(\text{CAC}\,|\,\text{CAC}>0)$ as dependent variables. However, the growth model described in Equation 4 indicates the existence of two major estimation biases with such methods (Figure 2). Principally, the existence of distinct CAC age trajectories across patients is not a structural component of log-linear regression, which treats each patient-observation as following a single population average log-linear path from a single age of onset, t_0. As a consequence, standard regression models of CAC systematically under-estimate the age of atherosclerotic onset (Figure 2, left) and rate of progression (Figure 2, right) in direct proportion to between-patient variability. Under realistic growth conditions, a standard deviation in atherosclerotic onset among patients of only 3 years can induce estimation biases as severe as five decades.

An additional bias is introduced by studies that alter the scores of patients with no atherosclerosis (CAC = 0). Since 0 is undefined on the logarithmic scale, many studies first re-assign these patients a CAC score of 1 [6, 12, 16, 21, 28, 29, 43–47], indicating these patients have already begun atherosclerosis. The common justification given for the log-transform is to normalize the distribution of CAC [1, 12, 16, 33]. Yet depending on the cohort, a majority of patients may have a CAC score of 0. Consequently, such logarithmic outcome distributions are extremely zero-inflated, e.g. Figure 1 in [12]. By estimating the average growth rate k from a realistic population in which some are growing CAC at rate $k = 1$ and others are not growing CAC at all, we can expect log-linear regression models to introduce additional biases of

at least -5.8 years in CAC age of onset (Figure 2, left, red) and 0.9 years in CAC doubling time (Figure 2, right, red) versus studies that exclude non-atherosclerotic patients before performing the regression [15, 17, 26, 27, 48, 49].

Alternate statistical specifications

To prevent inter-individual variation from manifesting as bias in statistical analyses, we can modify the log-linear model in Equation 4 with explicit CAC growth trajectories for each patient i by

$$y_j \sim \text{LogNormal}(\mu_j, \sigma),$$

$$\mu_j = k_i\,(t_j - t_{0i}), \tag{5}$$

$$t_{0i} \sim \text{Normal}(t_0, \sigma_{t0}),$$

$$k_i \sim \text{Normal}(k, \sigma_k).$$

That is, the time of initial onset for each individual i, t_{0i}, is assumed to vary around some population-average age t_0 by standard deviation σ_{t0}, and similarly for individual-specific average growth rate k_i.

Rather than exclude patients who have no CAC, or introduce bias by altering their values, another approach is to employ generalized linear regression to model the probability CAC > 0 [4, 5, 8–11, 32, 38, 50–55]. If we define the log-odds that a patient's CAC score at observation j is non-zero as $\theta_j = \text{logit}(\text{Pr}(y_j > 0))$, then for any individual i observed at age t_j,

$$\theta_j = \frac{t_j - \ell_i}{v}, \tag{6}$$

$$\ell_i \sim \text{Normal}(\ell, \sigma_\ell).$$

The probability CAC is observed to be nonzero rises sinusoidally from 0 to 1 over the life course, reaching 50% probability at age ℓ_i for patient i. We can expect ℓ_i to vary by individual according to parameters v and σ_ℓ, but also systematically by risk factors, e.g. patient sex, which can be incorporated into a multivariate logistic regression model. Unlike log-linear models above, this approach can utilize all observations, including those for which $y_j = 0$, at the cost of being unable to distinguish non-zero CAC magnitudes.

Models of CAC > 0 can be combined with models of CAC magnitudes in so-called mixture models [56], first introduced in the CAC literature by Kaplan, *et al.* [17] with the use of a zero-inflated Negative Binomial model. Such models incorporate two distinct probability distributions to describe both zero outcomes and non-zero magnitudes simultaneously. Here we employ the related concept of a hurdle model, which allows an additional link
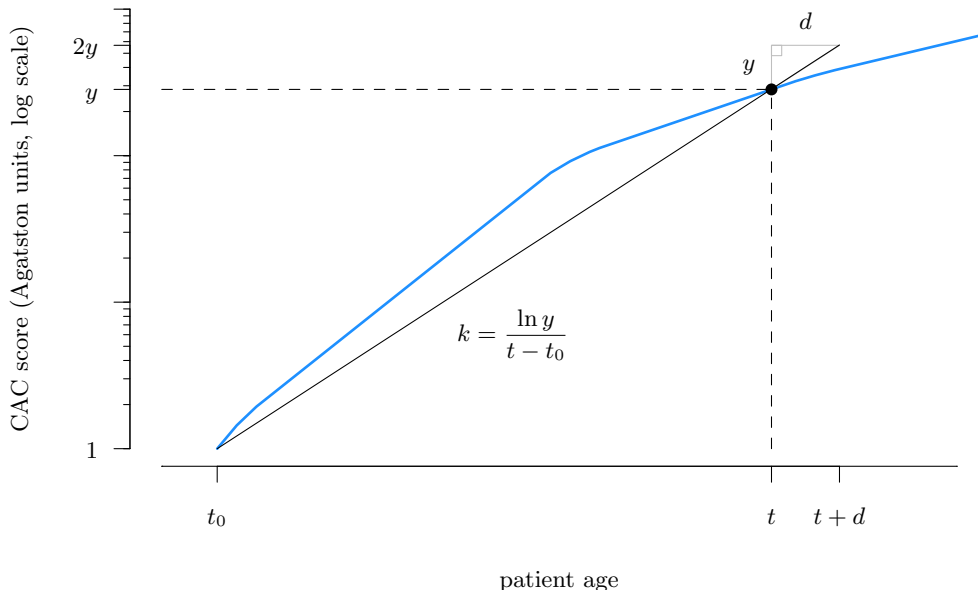
FIG. 1. A single patient's coronary atherosclerosis progression over time, as measured by CAC score (blue). Onset at age t_0 is defined as the initial appearance of calcified plaque (CAC > 0), which is equivalent to a CAC score of 1 in most studies. Before t_0 the patient's CAC is 0. All CAC measurements after t_0 are given by Equation 2. Average growth rate k and doubling time $d = \ln(2)/k$ are defined by a linear approximation (black line) using the location of t_0 and observation y. In this growth trajectory, the patient's instantaneous CAC growth rate $k(t)$ intermittently decreases. Because we are usually confined to a single measurement y at some time t, neither the location of t_0 nor the rate function $k(t)$ are directly observable, but can be approximated by log-linear regression methods (Equations 4, 5).

between these two distributions. Specifically, in modelling the probability of CAC > 0 by age, we are simultaneously describing the age of onset as a random quantity. If we re-interpret the inverse-logit link as a cumulative distribution function, onset age for each individual i must follow a Logistic probability distribution with mean ℓ_i and standard deviation $v\pi/\sqrt{3}$. Therefore, both the hurdle and log-normal components of this mixture share a common parameter in the individual onset age, as

$$t_{0i} \sim \text{Logistic}(\ell_i, v).$$

Thus, the ages of onset in a multi-trajectory log-linear model are random quantities that follow the parameters of the hurdle model. During Bayesian statistical estimation, the hurdle component's ℓ_i and v are updated by CAC presence/absence observations, adding additional resolution on the location of t_{0i} in the log-linear model. In turn, the log-linear estimation of t_{0i} is updated by observed magnitudes of nonzero CAC, and so inform the locations of ℓ_i and v in the hurdle component (Figure 3). Taken together, this fully defines the linked hurdle-lognormal mixture model. A complete list of variables is given in Table III.

Ranking statistical models by generative simulation

In order to characterize gross performance differences between the above methods of studying CAC growth, we simulated 540 datasets for populations experiencing a variety of realistic atherosclerotic growth scenarios (Table I), varying in the number of patients observed (from 1 to 1,000 patients), the number of observations per patient and the impact of standard risk factors (age and sex) on initial onset and progression. We can summarize patient data for N = 113,760 patients seen over all 540 studies in Table II.

For each study, we consider each of the seven models described above: log-linear regressions of $\ln(\text{CAC} + 1)$ or $\ln(\text{CAC} \mid \text{CAC} > 0)$ on patient age and logistic regressions of $\Pr(\text{CAC} > 0)$ with or without individual trajectories for each patient i, and mixture models including a logistic hurdle and log-linear magnitude component with or without a hierarchical link between t_{0i} and ℓ_{0i}. The seven models were fit to each of the 540 datasets, statistical estimates were calculated and extracted, and, given the true values for each individual trajectory were known, the average bias of each parameter estimate was com-
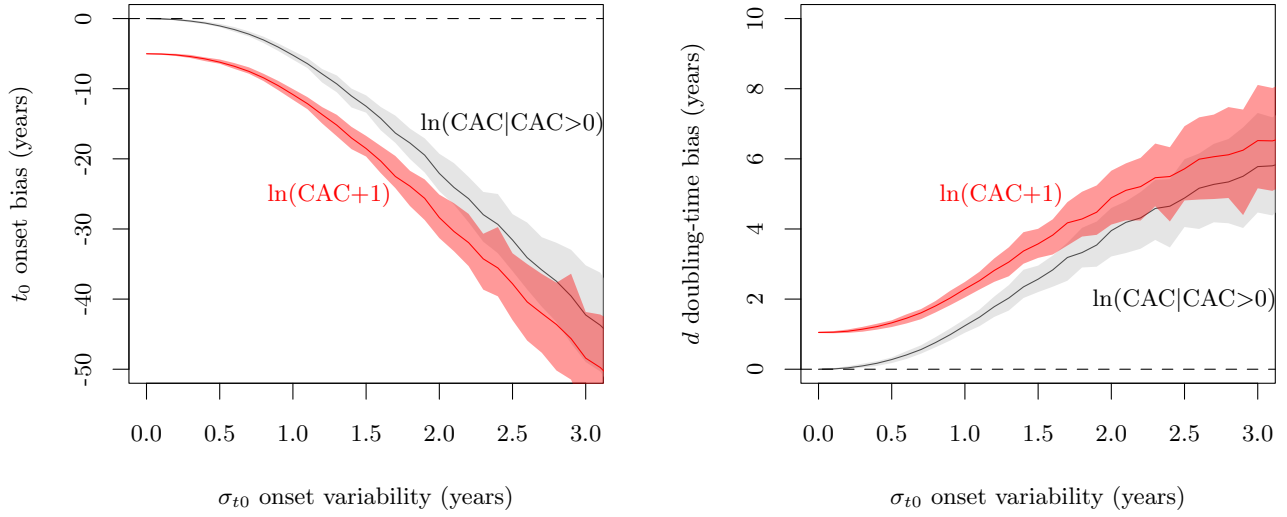
FIG. 2. The relationship between inter-individual variance in t_0 and bias in linear regression estimates of the age of onset (left) and the doubling time (right) of CAC for a simulated cohort of 5,000 patients ($t_0 = 40$, $k = 1$). Experimentally controlling the simulation standard deviation in individual age of onset, 100 cross-sectional datasets are constructed, each selecting a single random observation from each patient growth trajectory (average scan age 60 ± 10 years). For each dataset, parameter estimates are calculated in a log-linear regression of $\ln(\text{CAC}+1)$ (red) or $\ln(\text{CAC} \,|\, \text{CAC}>0)$ (black) on patient age, and the true onset age t_0 in the simulation is subtracted from onset estimates $-a/b$. Average cross-sectional estimates are shown with 89% confidence regions. The $\ln(\text{CAC}+1)$ estimates are, on average, -5.8 years more biased than $\ln(\text{CAC} \,|\, \text{CAC}>0)$ on t_0 and 0.9 years more biased in d over the observed range of σ_{t0}. Reproducible R code for this simulation and figure is available in the Supplementary Materials.

puted. To assess bias in estimating the importance of other risk factors besides patient age, we include patient sex (male or female) as an additional covariate in each model. All simulations were performed in R version 4.0.4 and models were fit with Stan by the `cmdstanr` library version 0.3.0.9, with all analysis code available in the Supplementary Materials.

III. RESULTS

Model performance is summarized in Table IV. Traditional log-linear models of $\ln(\text{CAC}+1)$ and $\ln(\text{CAC} \,|\, \text{CAC}>0)$ show large, systematic estimation errors across cross-sectional studies, of -11.5 years (SE: 0.2, P<0.001) and -9.7 years (SE: 0.3, P<0.001), respectively. Similarly, these models systematically under-estimate the average rate of progression, with a doubling-time bias of 2.4 years (SE: 0.1, P<0.001) for $\ln(\text{CAC}+1)$ and 1.8 years (SE: 0.1, P<0.001) for $\ln(\text{CAC} \,|\, \text{CAC}>0)$ over the true doubling time.

In both cross-sectional and longitudinal designs, estimation of onset ages (population average t_0 and patient-specific t_{0i}) is least biased in models that account for CAC presence / absence. In cross-sectional designs, a

logistic regression model estimating the population average age of onset t_0 is biased by 0.4 years (SE: 0.4), and not statistically distinct from zero bias (P=0.367). Logistic regression estimates of individual patient ages t_{0i} are similarly biased by only -0.1 years (SE: 0.1, P=0.430). The linked hurdle log-normal model is biased in cross-sectional designs on average -2.7 years (SE: 0.2, P<0.001) but out-performs all other log-linear models in estimation of both age of onset and in estimation of the doubling time.

Less systematic bias is observed across models in the estimation of the importance of sex as a predictor of onset and progression. In both cross-sectional and longitudinal designs, all seven models were able to assess the true covariate effect without clear bias, although standard log-linear models show the largest standard errors in bias in both t_0 and d (Table IV).

IV. DISCUSSION

Generally speaking, logistic regression and mixture models out-perform log-linear models in estimating population and individual-specific ages of CAC onset, and models that explicitly account for individual patient tra-
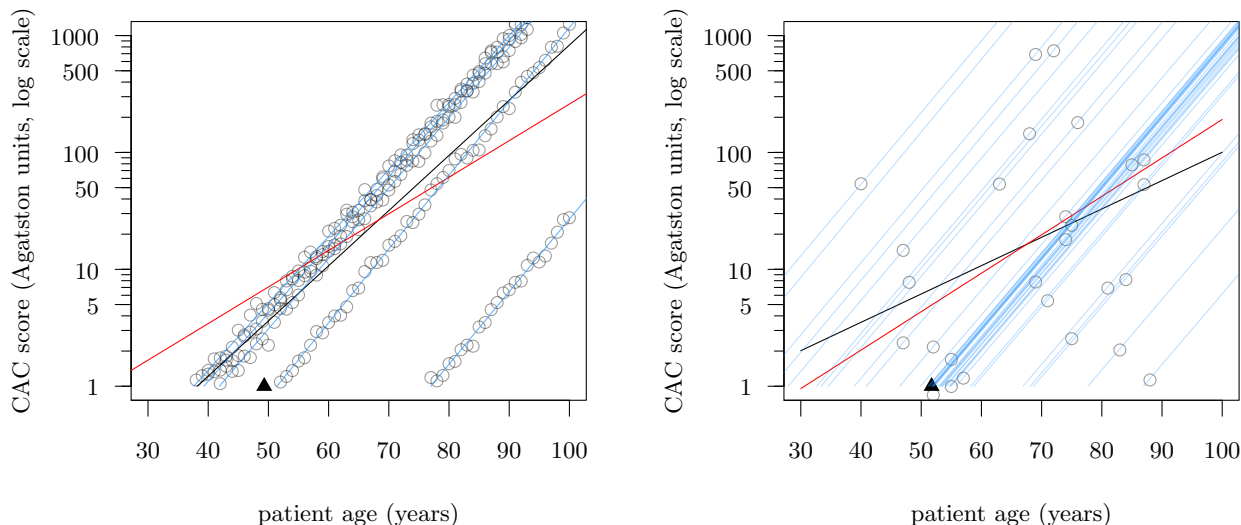
FIG. 3. Observed CAC trajectories for 5 longitudinal study patients (left) and 40 cross-sectional patients (right), with average counterfactual predictions from fitted $\ln(\text{CAC} \mid \text{CAC} > 0)$ (black), $\ln(\text{CAC} + 1)$ (red) and linked hurdle-lognormal (blue) models. In each figure, the true population average t_0 is indicated by a black triangle. All statistical models are fit in R and Stan, with prior uncertainty equal to simulation uncertainty. Reproducible R code for this simulation and figure is available in the Supplementary Materials.

jectories out-perform those that do not. These advantages are most pronounced in cross-sectional study designs, which constitute the vast majority of existing CAC datasets.

Because logistic regression models are able to accurately utilize the full set of patient observations, including those from the period before CAC growth begins, they are able to characterize the onset of atherosclerosis more accurately than log-linear models, which must either exclude patients observed with zero CAC or incorrectly alter these scores to accomodate the logarithmic scale, producing large estimation errors.

Although large biases in estimation are observed in some models in patient-specific CAC onset and progression, substantially smaller, unsystematic errors are found in estimation of the impact of sex. As a proxy for other risk factors such as smoking, obesity and physical activity, this covariate's predictive value is thus less likely to be altered significantly by choice of model in realistic datasets.

While the innovations in model design described above show marked improvements in estimation accuracy, they come with increased costs, in terms of model complexity and computation time, and successful posterior updating of the full linked-hurdle log-normal may not be feasible for very large samples of patients. However, Table IV indicates that estimation performance can be improved for specific parameters without the use of the full linked hurdle-lognormal model: for parameter d, the

$\ln(\text{CAC} \mid \text{CAC} > 0)_i$ model is, on average, less biased than either standard log-linear model, and for t_0, the logistic regression models with or without individual trajectories are both very accurate.

More generally, although the simulation studies presented here cover a broad range of realistic epidemiological scenarios, they are strictly limited to the context of the generative model described by Equations 1 and 2. The true process of CAC onset and progression is necessarily more complex, and is only part of the larger process of coronary atherosclerosis, which can be seen as beginning with arterial lesions, and progress to stenosis and vascular occlusion even in the absence of CAC [3, 57–59]. Incorporating more physiological details, such as the process of soft plaque accumulation prior to calcification, plaque density and stability [60] or the site-specific formation of individual calcified plaques [20], can potentially lead to further modelling innovations, and consequently greater improvements in estimation accuracy.

V. CONCLUSION

Currently, the study of atherosclerosis involves a wide variety methods of modelling the CAC score, with multiple statistical designs employed even within the same study. Both because of this awkwardness, and because the CAC score only suimmarizes the underlying distribution of calcified plaque, some have suggested alternative

	Longitudinal	Cross-sectional
Total Studies	180	360
Inter-Individual Variation		
No Variation	90	0
Variation on t_0, d	90	180
Sex Effect		
No sex differences	60	120
Sex difference in t_0	120	240
Sex difference in t_0, d	60	120
Prior Accuracy		
Regularizing Priors	90	180
Simulation Priors	90	180
Study Size		
$N = 1$ Patient	60	0
$N = 5$ Patients	60	0
$N = 20$ Patients	60	60
$N = 50$ Patients	0	60
$N = 100$ Patients	0	60
$N = 200$ Patients	0	60
$N = 500$ Patients	0	60
$N = 1000$ Patients	0	60

TABLE I. Tabulation of conditions across 540 simulated CAC studies. Here "longitudinal" studies involve 100 annual observations over the entire course of each patient's life (all individuals live to age 100), while "cross-sectional" studies simulate realistic CAC datasets by randomly sampling a single observation from each patient's adult measurements. As sex is a commonly-observed risk factor across studies, some populations here include a root-mean-square difference between men and women of 5.7 years on t_0 and 1.4 years on d. If present, between-individual variation is set by model parameters $\sigma_{t0} = 4.0$ years and $\sigma_d = 1.0$ years. Bayesian statistical models use either "simulation" priors, set to have the same uncertainty as the generative simulation itself, or "regularizing" priors, which reflect larger, empirically-plausible ranges for each parameter [56]. Generative R code to create these simulated populations can be found in the Supplementary Materials.

	Women	Men	Total
Number of Patients	56,901	56,859	113,760
Longitudinal	786	774	1,560
Cross-sectional	56,115	56,085	112,200
Number of Observations	134,715	133,485	268,200
Longitudinal	78,600	77,400	156,000
Cross-sectional	56,115	56,085	112,200
Onset t_0 (years)	49.8 (10.3)	50.1 (10.4)	49.9 (10.3)
Progression k	1.5 (0.3)	1.5 (0.3)	1.5 (0.3)
Doubling time d (years)	5.0 (1.1)	5.0 (1.1)	5.0 (1.1)
CAC category (%)			
0	42.8	43.0	42.9
1 to 9	20.4	20.6	20.5
10 to 99	19.0	19.1	19.1
100 to 399	8.1	8.0	8.1
400 to 999	3.6	3.5	3.5
≥ 1000	5.7	5.4	5.6

TABLE II. Characteristics of patients across 360 cross-sectional and 180 longitudinal CAC studies. Means (standard deviations) are given for t_0 and d.

Variable	Description
$k(t)$	instantaneous growth rate in CAC at age t
$y(t)$	true patient CAC score at age t
y_j	observed CAC score at observation j
t_j	patient age at observation j
t_0	population average age of onset
t_{0i}	age of onset for patient i
k	population average growth rate in CAC
d	population average doubling time in CAC
k_i	average growth rate in CAC for patient i
d_i	average doubling time in CAC for patient i
a	log-linear regression intercept
b	log-linear regression slope
v	scaling term for log-odds of CAC > 0
θ_j	log-odds of CAC > 0 for patient i in measurement j
ℓ	expected population average age of onset
ℓ_i	expected age of onset for patient i
σ	standard deviation on each CAC measurement y_j
σ_{t0}	standard deviation on t_{0i}
σ_k	standard deviation on k_i
σ_l	standard deviation on ℓ_i

TABLE III. Glossary of variables used.

metrics to studying atherosclerosis progression [20, 21]. However, the above computational meta-analysis shows the current statistical approaches to CAC modelling have substantial structural biases, tending to represent atherosclerotic growth as beginning earlier and progressing slower than it really does. Modifications to these models that recognize patient-specific growth trajectories are substantially more accurate in realistic simulation studies with both longitudinal and cross-sectional designs, and so we can expect such models to similarly out-perform traditional methods in the analysis of real patient data.

Supplementary materials

All statistical models and simulation data and software are available at https://github.com/babeheim/linked-calcium-growth.

model	design	t_0 (years)	t_0 sex (years)	d (years)	d sex (years)
$\text{logit}(\text{Pr}(\text{CAC} > 0))_i$	cross-sectional	-0.1 (0.1)	-0.0 (0.2)	-	-
$\text{logit}(\text{Pr}(\text{CAC} > 0))$	cross-sectional	0.4 (0.4)	-0.8 (0.7)	-	-
linked hurdle-lognormal$_i$	cross-sectional	**-2.7 (0.2)**	0.0 (0.2)	**0.5 (0.1)**	0.0 (0.0)
$\ln(\text{CAC} \,|\, \text{CAC} > 0)_i$	cross-sectional	**-5.9 (0.3)**	0.1 (0.2)	**1.1 (0.1)**	0.0 (0.0)
hurdle-lognormal$_i$	cross-sectional	**-6.1 (0.3)**	-0.0 (0.2)	**1.2 (0.1)**	0.0 (0.0)
$\ln(\text{CAC} \,|\, \text{CAC} > 0)$	cross-sectional	**-9.7 (0.3)**	0.3 (0.6)	**1.8 (0.1)**	-0.1 (0.2)
$\ln(\text{CAC} + 1)$	cross-sectional	**-11.5 (0.2)**	-0.3 (0.4)	**2.4 (0.1)**	0.0 (0.3)
linked hurdle-lognormal$_i$	longitudinal	-0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	-0.0 (0.0)
$\text{logit}(\text{Pr}(\text{CAC} > 0))$	longitudinal	0.0 (0.1)	-0.0 (0.0)	-	-
$\text{logit}(\text{Pr}(\text{CAC} > 0))_i$	longitudinal	0.0 (0.0)	-0.1 (0.3)	-	-
$\ln(\text{CAC} \,|\, \text{CAC} > 0)_i$	longitudinal	**-0.8 (0.2)**	-0.1 (0.3)	**0.1 (0.0)**	0.0 (0.0)
hurdle-lognormal$_i$	longitudinal	**-0.8 (0.2)**	-0.1 (0.3)	**0.1 (0.0)**	0.0 (0.0)
$\ln(\text{CAC} \,|\, \text{CAC} > 0)$	longitudinal	**-3.5 (0.3)**	-0.3 (0.6)	**0.5 (0.0)**	-0.1 (0.2)
$\ln(\text{CAC} + 1)$	longitudinal	**-25.7 (0.4)**	-0.2 (0.6)	**5.1 (0.2)**	-0.2 (0.4)

TABLE IV. Estimation bias (mean and standard error) over all 540 studies, in years. Statistical models that include individual-specific trajectories (subscript i) calculate bias estimates using the true t_{0i} and d_i values of each patient; models without individual-specific trajectories estimate population-level parameters, which are compared against the true population average t_0 and d within each simulated study. Biases on "sex" represent the estimated average difference between men and women minus the true generative difference (or lack thereof). Since $\text{E}(t_{0i}) = \ell_i$, logistic regression models use ℓ_i or ℓ as equivalents to t_{0i} and t_0. Estimates in boldface are statistically distinct from zero bias by two-tailed t-tests with $\alpha = 0.001$.

[1] A. S. Agatston, W. R. Janowitz, F. J. Hildner, N. R. Zusmer, M. Viamonte, and R. Detrano, Journal of the American College of Cardiology **15**, 827 (1990), publisher: American College of Cardiology Foundation.

[2] J. D. Berry, H. Ning, L. V. Horn, and D. M. Lloyd-Jones, New England Journal of Medicine , 9 (2012).

[3] T. Nakahara, M. R. Dweck, N. Narula, D. Pisapia, J. Narula, and H. W. Strauss, JACC: Cardiovascular Imaging **10**, 582 (2017).

[4] R. Vliegenthart, European Heart Journal **23**, 1596 (2002).

[5] G. T. Kondos, J. A. Hoff, A. Sevrukov, M. L. Daviglus, D. B. Garside, S. S. Devries, E. V. Chomka, and K. Liu, Circulation **107**, 2571 (2003).

[6] Y. Arad, K. J. Goodman, M. Roth, D. Newstein, and A. D. Guerci, Journal of the American College of Cardiology **46**, 158 (2005).

[7] M. J. LaMonte, S. J. FitzGerald, T. S. Church, C. E. Barlow, N. B. Radford, B. D. Levine, J. J. Pippin, L. W. Gibbons, S. N. Blair, and M. Z. Nichaman, American Journal of Epidemiology **162**, 421 (2005).

[8] M. J. Budoff, L. J. Shaw, S. T. Liu, S. R. Weinstein, P. H. Tseng, F. R. Flores, T. Q. Callister, P. Raggi, D. S. Berman, and T. P. Mosler, Journal of the American College of Cardiology **49**, 1860 (2007).

[9] A. J. Taylor, J. Bindeman, I. Feuerstein, F. Cao, M. Brazaitis, and P. G. O'Malley, Journal of the American College of Cardiology **46**, 807 (2005).

[10] S. G. Lakoski, Archives of Internal Medicine **167**, 2437 (2007).

[11] R. D. Santos, K. Nasir, J. A. Rumberger, M. J. Budoff, J. B. Braunstein, R. Meneghelo, M. Barreiros, A. Pereirinha, J. A. Carvalho, R. S. Blumenthal, and P. Raggi, Atherosclerosis **187**, 378 (2006).

[12] A. Schmermund, N. Lehmann, L. F. Bielak, P. Yu, P. F. Sheedy, A. E. Cassidy-Bushrow, S. T. Turner, S. Moebus, S. Möhlenkamp, A. Stang, K. Mann, K.-H. Jöckel, R. Erbel, and P. A. Peyser, Atherosclerosis **195**, e207 (2007).

[13] J. A. Hoff, E. V. Chomka, A. J. Krainik, M. Daviglus, S. Rich, and G. T. Kondos, The American Journal of Cardiology **87**, 1335 (2001).

[14] D. E. Bild, American Journal of Epidemiology **156**, 871 (2002).

[15] D. E. Bild, R. Detrano, D. Peterson, A. Guerci, K. Liu, E. Shahar, P. Ouyang, S. Jackson, and M. F. Saad, Circulation **111**, 1313 (2005).

[16] M. J. Budoff, K. Nasir, S. Mao, P. H. Tseng, A. Chau, S. T. Liu, F. Flores, and R. S. Blumenthal, Atherosclerosis **187**, 343 (2006).

[17] H. Kaplan, R. C. Thompson, B. C. Trumble, L. S. Wann, A. H. Allam, B. Beheim, B. Frohlich, M. L. Sutherland, J. D. Sutherland, J. Stieglitz, D. E. Rodriguez, D. E. Michalik, C. J. Rowan, G. P. Lombardi, R. Bedi, A. R. Garcia, J. K. Min, J. Narula, C. E. Finch, M. Gurven, and G. S. Thomas, The Lancet **389**, 1730 (2017).

[18] R. C. Thompson, A. H. Allam, G. P. Lombardi, L. S. Wann, M. L. Sutherland, J. D. Sutherland, M. A.-T. Soliman, B. Frohlich, D. T. Mininberg, J. M. Monge, C. M. Vallodolid, S. L. Cox, G. Abd el Maksoud, I. Badr, M. I. Miyamoto, A. el-Halim Nur el din, J. Narula, C. E. Finch, and G. S. Thomas, The Lancet **381**, 1211 (2013).

[19] M. D. Gurven and D. E. Lieberman, Evolution and Human Behavior **41**, 330 (2020).

[20] D. S. Berman, Y. Arnson, and A. Rozanski, JACC: Cardiovascular Imaging **9**, 1417 (2016), publisher: American College of Cardiology Foundation.

[21] Blaha Michael J., Budoff Matthew J., Tota-Maharaj Rajesh, Dardari Zeina A., Wong Nathan D., Kronmal Richard A., Eng John, Post Wendy S., Blumenthal Roger S., and Nasir Khurram, JACC: Cardiovascular Imaging **9**, 1407 (2016), publisher: American College of Cardiology Foundation.

[22] G. Koulaouzidis, R. Nicoll, D. Charisopoulou, T. McArthur, P. J. Jenkins, and M. Y. Henein, International Journal of Cardiology , 5 (2013).

[23] M. G. Silverman, M. J. Blaha, H. M. Krumholz, M. J. Budoff, R. Blankstein, C. T. Sibley, A. Agatston, R. S. Blumenthal, and K. Nasir, European Heart Journal **35**, 2232 (2014).

[24] A. Fujiyoshi, K. Miura, T. Ohkubo, T. Kadowaki, S. Kadowaki, M. Zaid, T. Hisamatsu, A. Sekikawa, M. J. Budoff, K. Liu, and others, American Journal of Epidemiology **180**, 590 (2014).

[25] G. Koulaouzidis, P. Jenkins, and T. McArthur, International Journal of Cardiology **168**, 1647 (2013).

[26] M. J. Budoff, R. Young, V. A. Lopez, R. A. Kronmal, K. Nasir, R. S. Blumenthal, R. C. Detrano, D. E. Bild, A. D. Guerci, K. Liu, S. Shea, M. Szklo, W. Post, J. Lima, A. Bertoni, and N. D. Wong, Journal of the American College of Cardiology **61**, 1231 (2013).

[27] M. H. Criqui, J. O. Denenberg, J. H. Ix, R. L. McClelland, C. L. Wassel, D. E. Rifkin, J. J. Carr, M. J. Budoff, and M. A. Allison, JAMA **311**, 271 (2014).

[28] R. L. McClelland, N. W. Jorgensen, M. Budoff, M. J. Blaha, W. S. Post, R. A. Kronmal, D. E. Bild, S. Shea, K. Liu, K. E. Watson, A. R. Folsom, A. Khera, C. Ayers, A.-A. Mahabadi, N. Lehmann, K.-H. Jöckel, S. Moebus, J. J. Carr, R. Erbel, and G. L. Burke, Journal of the American College of Cardiology **66**, 1643 (2015).

[29] J. M. Hughes-Austin, A. Dominguez III, M. A. Allison, C. L. Wassel, D. E. Rifkin, C. G. Morgan, M. R. Daniels, U. Ikram, J. B. Knox, C. M. Wright, and others, JACC: Cardiovascular Imaging **9**, 152 (2016).

[30] H. M. Ahmed, M. J. Blaha, K. Nasir, S. R. Jones, J. J. Rivera, A. Agatston, R. Blankstein, N. D. Wong, S. Lakoski, M. J. Budoff, G. L. Burke, C. T. Sibley, P. Ouyang, and R. S. Blumenthal, American Journal of Epidemiology **178**, 12 (2013).

[31] Y. Ohmoto-Sekine, R. Yanagibori, K. Amakawa, M. Ishihara, H. Tsuji, K. Ogawa, R. Ishimura, S. Ishiwata, M. Ohno, T. Yamaguchi, and Y. Arase, Journal of Cardiology **67**, 449 (2016).

[32] K. Nasir, L. J. Shaw, S. T. Liu, S. R. Weinstein, T. R. Mosler, P. R. Flores, F. R. Flores, P. Raggi, D. S. Berman, R. S. Blumenthal, and M. J. Budoff, Journal of the American College of Cardiology **50**, 953 (2007).

[33] A. R. Folsom, Archives of Internal Medicine **168**, 1333 (2008).

[34] Z.-h. Hou, B. Lu, Y. Gao, S.-l. Jiang, Y. Wang, W. Li, and M. J. Budoff, JACC: Cardiovascular Imaging **5**, 990 (2012).

[35] R. L. McClelland, K. Nasir, M. Budoff, R. S. Blumenthal, and R. A. Kronmal, The American Journal of Cardiology **103**, 59 (2009).

[36] A. P. DeFilippis, M. J. Blaha, C. E. Ndumele, M. J. Budoff, D. M. Lloyd-Jones, R. L. McClelland, S. G. Lakoski, M. Cushman, N. D. Wong, R. S. Blumenthal, J. Lima, and K. Nasir, Journal of the American College of Cardiology **58**, 2076 (2011).

[37] W. Arguelles, M. M. Llabre, F. J. Penedo, M. L. Daviglus, R. L. Sacco, K. Liu, M. Szklo, J. F. Polak, J. Eng, G. L. Burke, and N. Schneiderman, International Journal of Cardiology **174**, 51 (2014).

[38] S. P. Whelton, M. G. Silverman, J. W. McEvoy, M. J. Budoff, R. Blankstein, J. Eng, R. S. Blumenthal, M. Szklo, K. Nasir, and M. J. Blaha, JACC: Cardiovascular Imaging **8**, 1393 (2015).

[39] A. Gelman, J. Hill, and A. Vehtari, *Regression and other stories* (Cambridge University Press, 2021).

[40] M. J. Budoff, K. Nasir, R. L. McClelland, R. Detrano, N. Wong, R. S. Blumenthal, G. Kondos, and R. A. Kronmal, Journal of the American College of Cardiology **53**, 345 (2009).

[41] E. S. Houslay, Heart **92**, 1207 (2006).

[42] Y. Ikegami, I. Inoue, K. Inoue, Y. Shinoda, S. Iida, S. Goto, T. Nakano, A. Shimada, and M. Noda, npj Aging and Mechanisms of Disease **4**, 7 (2018).

[43] H. S. Oei, European Heart Journal **25**, 48 (2004).

[44] R. Detrano, D. E. Bild, K. Liu, D. A. Bluemke, and K. Watson, New England Journal of Medicine , 10 (2008).

[45] S. Möhlenkamp, N. Lehmann, S. Moebus, A. Schmermund, N. Dragano, A. Stang, J. Siegrist, K. Mann, K.-H. Jöckel, and R. Erbel, Journal of the American College of Cardiology **57**, 1455 (2011).

[46] N. D. Wong, V. A. Lopez, M. Allison, R. C. Detrano, R. S. Blumenthal, A. R. Folsom, P. Ouyang, and M. H. Criqui, Atherosclerosis **214**, 436 (2011).

[47] M. J. Leening, S. E. Elias-Smale, M. Kavousi, J. F. Felix, J. W. Deckers, R. Vliegenthart, M. Oudkerk, A. Hofman, E. W. Steyerberg, B. H. C. Stricker, and J. C. Witteman, JACC: Cardiovascular Imaging **5**, 874 (2012).

[48] R. L. McClelland, H. Chung, R. Detrano, W. Post, and R. A. Kronmal, Circulation **113**, 30 (2006).

[49] L. F. Bielak, P. Yu, K. A. Ryan, J. A. Rumberger, P. F. Sheedy, S. T. Turner, W. Post, A. R. Shuldiner, B. D. Mitchell, and P. A. Peyser, Atherosclerosis **196**, 888 (2008).

[50] R. Vliegenthart, M. Oudkerk, A. Hofman, H.-H. S. Oei, W. van Dijck, F. J. van Rooij, and J. C. Witteman, Circulation **112**, 572 (2005).

[51] A. Schmermund, S. Möhlenkamp, S. Berenbein, H. Pump, S. Moebus, U. Roggenbuck, A. Stang, R. Seibel, D. Grönemeyer, K.-H. Jöckel, and R. Erbel, Atherosclerosis **185**, 177 (2006).

[52] P. L. Lutsey, A. V. Diez Roux, D. R. Jacobs, G. L. Burke, J. Harman, S. Shea, and A. R. Folsom, American Journal of Public Health **98**, 1963 (2008).

[53] T. M. Okwuosa, P. Greenland, S. G. Lakoski, H. Ning, J. Kang, R. S. Blumenthal, M. Szklo, J. R. Crouse, J. A. Lima, K. Liu, and D. M. Lloyd-Jones, The American Journal of Cardiology **107**, 879 (2011).

[54] J.-H. Yu, S. H. Yim, S. H. Yu, J. Y. Lee, J. D. Kim, M. H. Seo, W. S. Jeon, S.-E. Park, C.-Y. Park, W.-Y. Lee, K.-W. Oh, S.-W. Park, and E.-J. Rhee, Endocrinology and Metabolism **28**, 33 (2013).

[55] V. Aboyans, R. L. McClelland, M. A. Allison, M. M. McDermott, R. S. Blumenthal, K. Macura, and M. H. Criqui, Atherosclerosis **214**, 169 (2011).

[56] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, 2nd ed. (Chapman and Hall/CRC, 2020).

[57] J. L. Kelly, D. Thickman, S. D. Abramson, P. R. Chen, S. F. Smazal, M. J. Fleishman, and S. C. Lingam, American Journal of Roentgenology **191**, 50 (2008).

[58] I. Gottlieb, J. M. Miller, A. Arbab-Zadeh, M. Dewey, M. E. Clouse, L. Sara, H. Niinuma, D. E. Bush, N. Paul, A. L. Vavere, J. Texter, J. Brinker, J. A. Lima, and C. E. Rochitte, Journal of the American College of Cardiology **55**, 627 (2010).

[59] F. S. Gabriel, L. F. G. Gonçalves, E. V. d. Melo, A. C. S. Sousa, I. M. F. Pinto, S. M. M. Santana, C. J. O. d. Matos, M. J. S. Souto, F. M. d. S. Conceição, and J. L. M. Oliveira, Arquivos Brasileiros de Cardiologia **110**, 420 (2018).

[60] A. Shioi and Y. Ikari, Journal of Atherosclerosis and Thrombosis **25**, 294 (2018).