Aggregating Epigenetic Clocks to Study Human Capital Formation*

Giorgia Menta[1,2], Pietro Biroli[3], Divya Mehta[4], Conchita D'Ambrosio[2], Deborah Cobb-Clark[5]

[1]Luxembourg Institute of Socio-Economic Research (LISER), [2]University of Luxembourg, [3]University of Bologna, [4]Queensland University of Technology, [5]University of Sydney

Abstract

Economists have increasingly recognized the role of genetics in economic outcomes. Epigenetics, which studies changes in gene expression without altering DNA sequences, provides insights into how socioeconomic environments affect human biology and influence economic behaviors and outcomes. This paper explores the application of "epigenetic clocks", which aggregate epigenetic data to predict biological aging and health risks, in economics research. We propose a novel integrated measure of epigenetic aging, the Multi EpiGenetic Age (MEGA) clock: several epigenetic clocks are combined to reduce measurement error and improve efficiency. Using data from the Avon Longitudinal Study of Parents and Children (ALSPAC), we apply the MEGA clock in two empirical settings: first, we examine the association between longitudinal exposure to child abuse and epigenetic age acceleration in adolescence; second, we test the association between epigenetic age acceleration and early-adulthood cognitive and socioemotional outcomes. Our findings reveal that (i) exposure to child maltreatment before adolescence is associated with half a year of accelerated epigenetic aging and that (ii) epigenetic aging predicts moderately worse cognitive and socioemotional outcomes in early adulthood. These results highlight the usefulness of epigenetic aging as a metric for understanding the long-term effects of early-life adversity and inform economic policies targeting public health and productivity.

JEL Codes: I12, I14, J24

Keywords: Epigenetic clocks, DNA methylation, Child abuse, Human capital, ALSPAC

*We would like to thank Tiger Mathieson for excellent research assistance. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Welcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and all authors will serve as guarantors for the contents of this paper. A comprehensive list of grants funding the ALSPAC project is available on their website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). The authors are grateful for research support from the Australian Research Council (ARC) (grant numbers DP200100979, CE200100025) and from the Fonds National de la Recherche Luxembourg (Grant C19/SC/13650569/ALAC).

1. Introduction

Economists have long recognized the importance of genetics in shaping economic outcomes, supported by extensive evidence from twin and adoption studies (e.g. Leibowitz 1974, Goldberger 1976, Taubman 1976, Beauchamp et al., 2017; Björklund et al., 2006; Cesarini et al., 2010), as well as recent advances in behavioral genetics (e.g. Okbay et al., 2022; Abdellaoui et al. 2023, Benjamin et al. 2007, 2024) and their integration in economics (e.g. Papageorge and Thom, 2020; Houmark et al., 2024). While genetic variants are fixed at conception, genetic expression is not. Epigenetics, the study of changes in gene expression without alterations to the DNA sequence, is one of the ways in which the environment (*nurture*) can affect human biology (*nature*). Epigenetic modifications can happen as a result of exposures to adverse socioeconomic environments (Cole 2009), which leave a lasting biological footprint (Szyf 2009). While traditionally the domain of biologists, epigenetics is also conceptually relevant to economists and other social scientists, as it provides insights into how environmental exposures, such as socioeconomic status and stress, can have lasting biological impacts that influence economic behaviors and outcomes (Neu, 2023). Understanding these epigenetic influences can provide insights into the long-term impacts of economic policies and social environments on public health and productivity.

One of the most promising epigenetic tools are 'epigenetic clocks', which aggregate DNA methylation data to predict biological aging, mortality, and disease risk. Chronological age alone has been shown to be an unreliable predictor of physiological functioning, due to individual differences in biological aging (Kotschy et al., 2024). Summarizing complex epigenetic information, these clocks offer a quantifiable measure of biological age that can be more predictive of health outcomes than chronological age, providing a valuable metric for both biological and social-science research. However, the development and implementation of various epigenetic clocks presents a major challenge (Bell et al., 2019). While aimed at capturing the same latent concept of epigenetic aging, the algorithms used to define different epigenetic clocks diverge in terms of methodology and selection of relevant epigenetic sites – leading to inconsistencies and measurement error.

Here, we provide an illustration of the use of epigenetic clocks in economics research, leveraging epigenetic and behavioral data from a large cohort of children in the UK, the Avon Longitudinal Study of Parents and Children (ALSPAC). Our first contribution is to propose a methodological

framework that combines several existing clocks into a single metric, which we interpret as an improved and more robust measure of latent epigenetic age: the Multi EpiGenetic Age (MEGA) clock. We do so by integrating four well-established clocks (Horvath, 2013; Hannum et al., 2013; Levine et al., 2018; Lu et al., 2019), each trained to predict different aspects of aging and health. By leveraging different methods, such as exploratory factor analysis and Structural Equation Modeling (SEM), our method is effective in reducing measurement error inherent to individual clocks and increasing efficiency, providing a more accurate measure of epigenetic age. With the increasing availability of pre-computed epigenetic clocks in large individual-level datasets, this combined metric is particularly easy to implement without in-depth knowledge of the biology behind the single clocks. It also retains the intuitive age-scale of epigenetic clocks, making it directly comparable to chronological age.

Individuals whose epigenetic age is higher than their chronological age are aging at a faster speed biologically than they are chronologically: this is age acceleration. Our second contribution consists of two novel empirical applications that illustrate how the MEGA clock can be used to study the determinants and consequences of accelerated epigenetic aging. First, we examine the impact of longitudinal exposure to child abuse on epigenetic age acceleration in late adolescence. Prior research, such as the ALSPAC study by Lawn et al. (2018), has shown that child abuse is associated with modifications across the whole epigenome. We provide novel evidence to specifically address whether the epigenetic modifications associated with child abuse translate into faster epigenetic aging. We do so by investigating the longitudinal relationship between child abuse and epigenetic aging, considering different abuse trajectories and critical periods of exposure. Our results show that exposure to child maltreatment before adolescence is associated with half a year of accelerated epigenetic aging, an effect comparable to about half the magnitude of the age-acceleration premium for being female.

These effects are likely to be consequential for young people's health and wellbeing. Faster epigenetic aging has been shown to be predictive of numerous outcomes, such as mortality, disease incidence, and cognitive performance in children. In our second empirical application, we contribute to this literature by investigating the association between age acceleration and human capital (Heckman et al., 2006; Cunha et al., 2010). The biological processes linked to accelerated epigenetic aging – which include, among others, the regulation of the immune system, lipid

function, and neuronal pathways (Han et al., 2018; Lu et al., 2019) – may well be detrimental to the formation of cognitive and non-cognitive skills. In particular, we include the MEGA clock as an input in the human capital production function, hypothesizing that adolescents who are age-accelerated display worse cognitive and non-cognitive outcomes in early adulthood – after controlling for family background, past cognitive and non-cognitive skills and past health. Our results suggest that epigenetic age acceleration in adolescence predicts fewer years of education and worse mental-health and employment outcomes, above and beyond traditional health biomarkers (e.g. BMI) and health behaviors (e.g. smoking and drinking).

We make several important contributions. First, we advance the literature on epigenetic clocks, through a methodological innovation aimed at reducing measurement error and minimizing the probability of type-2 error, thereby enhancing the reliability and efficiency of epigenetic age measurement. There are several current limitations of the literature employing epigenetic clocks, some of which we overcome with the MEGA clock. These include the arbitrary choice of one clock over others, the limited replicability of results across different clocks and limited statistical power. To the best of our knowledge, this is the first study to build an integrated epigenetic clock based on other clocks. Recently, Martinez et al. (2024) have proposed a complementary, yet distinct, approach: using a battery of biomarkers, including epigenetic clocks, and confirmatory factor analysis, they identify three factors interpretable as epigenetic age, systemic biological age, and immune age.

Second, our work fits in the literature on early life adversity and epigenetic clocks, as exemplified by systematic reviews like those by Cecil et al. (2020) and Rubens et al. (2023). Using data from ALSPAC mothers, Lawn et al. (2018) find that sexual abuse experienced by the age of 17 is associated with a 3.4 years higher epigenetic age in adulthood (Horvath clock), after adjusting for childhood and adulthood socio-economic positioning. Marini et al. (2020) look at early childhood DNAm in ALSPAC children and find that sexual and physical abuse, especially during early and middle-childhood sensitive periods, are associated with accelerated epigenetic aging at age 7. The effect is driven by girls and is only found when using the Hannum rather than the Horvath's clock. Unlike previous studies, we incorporate dynamic trajectories and measures taken across different developmental periods, adding a temporal dimension to our analysis. While a similar longitudinal approach can be found in Lussier et al. (2023), these authors examined broad changes to the

epigenome associated with abuse; our focus on the portion of the epigenome that predicts aging, disease and mortality (as captured by the MEGA clock) contributes to a more nuanced understanding of the epigenetic footprint of child adversity.

Through our second empirical application, we additionally contribute to studying the consequences of epigenetic age acceleration. Age acceleration has been shown to be predictive of all-cause mortality (Faul et al., 2023; Lu et al., 2019; Perna et al., 2016) and the occurrence of diseases such as cancer (Horvath, 2013; Dugué et al., 2018; Perna et al., 2016), Alzheimer's disease (McCartney et al., 2018), and healthcare utilization (Davillas and Jones, 2024). In children, age acceleration has been shown to display negative association with test-scores (Niccodemi et al., 2022), cognitive functioning (Raffington et al., 2023a), and to increase social disparities in mental health (Raffington et al., 2023b). Our focus on young adults contributes to the understanding of the early adulthood consequences of accelerated epigenetic aging.

From an economics standpoint, our study contributes to the literature on the adverse health consequences of child abuse (Currie and Spatz Widom, 2010; Fletcher, 2009; Henkaus, 2022; Suglia et al., 2014). On top of using a biomarker of aging that is still new to the health economics literature (epigenetic age), we additionally incorporate information from various raters (i.e., mothers, fathers, and children) to enhance the accuracy of measuring sensitive constructs like child abuse. Finally, our life course approach maps out the experiences of childhood and adolescence, providing a comprehensive view of how early-life conditions influence biological aging and subsequent outcomes. This holistic perspective bridges gaps in this literature (e.g. Korous et al., 2023; Petrovic et al., 2023), offering new insights into the interplay between environmental factors and biological processes over time.

The remainder of the paper is organized as follows. Section 2 provides description of the epigenetic clocks, with a review of the causes and consequences of accelerated epigenetic aging. Section 3 provides the methodological framework of the MEGA clocks, while Section 4 introduces the ALSPAC data and estimation sample. Section 5 first describes the empirical features of the MEGA clocks and then presents the empirical strategy and results for each of the two empirical applications. Last, Section 6 advances some hypotheses of mechanisms, and Section 7 concludes.

2. Epigenetic clocks

All humans age, but the rate at which we do differs considerably from person to person. This deviation between the "rate of ageing" and chronological age is referred to as "biological age". Biological aging is associated with progressive loss of function at the cellular, tissue, and organ levels, further promoting the general decline in physical functioning and cognitive performance (López-Otín et al., 2013). One well-studied mechanism is that of epigenetics, namely the ensemble of reversible chemical and structural alterations to the genome that can lead to long-term changes in gene activity without altering the underlying DNA sequence itself (Klengel et al., 2014). The most studied form of epigenetics is DNA methylation (DNAm).[1] This process happens when a methyl group binds to the DNA strand at a CpG site (a region in DNA where a cytosine nucleotide is adjacent to a guanine nucleotide, with one phosphate in between), leading to inhibited gene expression.

The biological age of a person is a more accurate measure of an individual's functional capacity compared to their actual chronological age, and studies have found evidence of age-related hypo- or hyper-methylation within specific CpG sites or islands. This has laid the foundation for the development of epigenetic biomarkers of aging, also referred to as epigenetic clocks. Epigenetic clocks use supervised machine-learning algorithms, such as penalized elastic net regression, to estimate biological age based on read-outs from tens or hundreds of DNAm sites across the genome.

The first two widely used epigenetic clocks, Horvath's DNAmAge clock and Hannum's clock, were developed in 2013 to predict chronological age. The Horvath clock, the first epigenetic age predictor, was based on a programmed algorithm to identify age-related signals and was based on the combined methylation status of 353 CpG sites. This was developed as a multi-organ clock that predicted age from embryo to old age, from data on many tissues and organs (e.g. whole blood, cerebellum, colon, kidney, liver, lung). The Hannum clock was developed on blood-based DNAm; using elastic net regression to predict individual chronological age, the clock 71 identified CpG

[1] While other forms of epigenetic changes exist (e.g. histone modification), current biotechnologies have become increasingly cost-efficient in the stabilization and extraction of DNAm from human tissues. As a result, there is a wealth of data and well-characterized tools available for studying DNAm.

sites that can be used to accurately predict age (Hannum et al., 2013). Compared with Horvath's epigenetic clock, epigenetic age estimated by Hannum's epigenetic clock shows higher accuracy in the prediction of blood samples from adults. Since the model was only trained in whole-blood samples from adults, estimates of Hannum epigenetic age have been shown to be biased when applied in non-blood tissues (Simpkin et al., 2016, Simpkin et al., 2017) and in children (Hannum et al., 2013).

It has been argued that these "first-generation" clocks may exclude CpGs whose methylation patterns may reflect variations in biological age, because the algorithms were trained against chronological age. For example, the algorithm underlying these two clocks was programmed to detect age-dependent patterns and could perfectly predict chronological age. However, the better the predicted chronological age was, the worse the predicted mortality rate was (Zhang Q et al, 2019).

In recent years, a second generation of DNAm-based biomarkers for aging has been developed, capturing CpGs associated with the functional stage – along with chronological age. In 2018 Levine developed the PhenoAge clock, which was trained to predict not only chronological age but other aging-related indicators (e.g. blood glucose, liver and kidney markers of function).[2] This clock, while being a good predictor of chronological age, has been extensively shown to predict age-related health outcomes better than the first-generation clocks and allows to differentiate morbidity and mortality risks of individuals with the same chronological age (Levine et al., 2018). Later, Horvath developed the GrimAge which superseded the previous clocks in predicting both age-related disease as well as mortality. GrimAge is a linear combination of DNAm-based surrogate biomarkers for health-related plasma proteins, smoking pack-years, sex, and chronological age. It was a stronger predictor of lifespan, age-related conditions, disease, and mortality risk compared to the widely used Horvath's DNAmAge (Lu et al., 2019).

In different epigenetic clocks, the number of age-related CpGs forming the predictor varies (Horvath: 353 CpGs, Hannum: 71 CpGs, PhenoAge: 513 CpGs, GrimAge: 1030 CpGs), as do the CpGs included in the model. For instance, only 41 of the 513 CpGs that construct PhenoAge are

[2] The model combined 10 clinical characteristics, including chronological age, albumin, creatinine, glucose, C-reactive protein levels, lymphocyte percentage, mean cell volume, red blood cell distribution width, alkaline phosphatase, and white blood cell count. Based on 513 age-related CPGs on 3 chips (27 K, 450 K, 850 K), the DNAm PhenoAge has achieved greater applicability across chip platforms than other clocks.

the same as Horvath's clock; PhenoAge and the Horvath's clock share as few as 5 CpGs with Hannum's clock (Levine et al., 2018).[3]

Age acceleration is computed as the discrepancy between chronological age and epigenetic age. The most common approach to do so is to measure age acceleration as residuals from a regression of epigenetic age on chronological age. This is the approach we adopt in this paper: our empirical specifications in Section 3.3 always keep chronological age constant, so that epigenetic age is interpretable as epigenetic age acceleration.

3. From several clocks to one latent construct: the MEGA clock

Different epigenetic clocks capture different aspects of the biological processes linked to aging, disease onset and all-cause mortality. Drawing from the social-sciences literature on the measurement of latent factors, here, we propose a novel way to harness the information coming from different epigenetic clocks into a unique measure of epigenetic aging. In the current study, we focus on two first-generation clocks (the Horvath and the Hannum clocks), trained to predict chronological age, and two second-generation clocks (the PhenoAge and GrimAge clocks), trained to predict lifespan/functional stage. The rationale for focusing on these four specific clocks is twofold. On the one hand, their robustness and replicability have been shown extensively across a variety of samples and tissues (Lu et al., 2019; Maddock et al., 2020; McCrory et al., 2021). On the other hand, they all rely on genome-wide DNAm data derived from Illumina arrays and are estimated using similar methods (i.e. penalized elastic net algorithms). If different epigenetic clocks truly capture different facets of the molecular physiological determinants of aging and diverse aspects of health, then combining the information across these four clocks will reduce measurement error, leading to a more robust, holistic measure of DNAm age: the Multi EpiGenetic Age (MEGA) clock.

We here define three versions of the MEGA clock, each varying in the strength of their underlying assumptions. While in the empirical applications we will rely on the four clocks defined above, the procedures described below are generalizable to any set of epigenetic clocks. The first,

[3] It is suggested that models with a large number of CpG inclusions are more robust and accurate than those with fewer CpG inclusions (Li et al., 2015; Lin et al., 2016). However, the epigenetic clocks measure the comprehensive properties of the methylome, and the most robust model can be obtained by including a moderate number of age CpGs (Horvath and Raj, 2018).

MEGA$_{WGT}$, is a weighted index approach based on Anderson (2008). Here, clocks are combined as a weighted sum, with weights being the row-sums of the inverse variance-covariance matrix of the clocks. Let $\{C_1, \dots, C_K\}$ be a set of epigenetic clocks, all expressed in the same unit (years of age) and M their variance-covariance matrix. Let I_K be a vector of ones, of length *K*. The *MEGA*$_{WGT}$ clock is defined as:

$$MEGA_{WGT} = \frac{\sum_{k=1}^{K} w_{k,1} C_k}{\sum_{k=1}^{K} w_{k,1}} \quad (1)$$

Where $w_{k,1}$ is the *k*-th element of vector $w = M^{-1} I_K$. As described in Anderson (2008), this weighting procedure is analogous to the joint estimation of seemingly unrelated regression models of all outcomes on the treatment indicator, constraining all coefficients to be equal, and corresponds to an efficient Generalized Least Squares (GLS) estimator.

The second, *MEGA*$_{FA}$, is instead based on exploratory Factor Analysis (FA), commonly used to identify a latent construct based on a set of measurements. It is a data-driven method that aims to explain the observed variability in the data by grouping variables that tend to co-vary and extracting a smaller number of latent factors that account for the patterns in the observed variables. From a theoretical standpoint, since $\{C_1, \dots, C_K\}$ are selected to capture the same latent concept of epigenetic age, we expect only one factor to be retained by the exploratory FA. We interpret this factor as the *MEGA*$_{FA}$, a more flexible version of the *MEGA*$_{WGT}$ that allows weights to reflect not only the correlation structure of the single clocks, but also their factor loadings. Let L_K be the vector of factor loadings for each of the *K* clocks. Then *MEGA*$_{FA}$ is defined similarly to *MEGA*$_{WGT}$ as:

$$MEGA_{FA} = \frac{\sum_{k=1}^{K} u_{k,1} C_k}{\sum_{k=1}^{K} u_{k,1}} \quad (2)$$

Where $u_{k,1}$ is the *k*-th element of vector $u = M^{-1} L_K$.

Last, *MEGA*$_{SEM}$, is based on Structural Equation Modelling (SEM) and relies on the strongest assumptions. The advantage of SEM is that it allows to simultaneously model both observed and latent variables and their direct relationships, as well as error terms. Informed by the FA, we then estimate the *MEGA*$_{SEM}$ clock as the latent factor of a measurement equation based on the individual

clocks, while simultaneously modelling its relationship with child abuse in the behavioral (or structural) equation. The measurement equation can be illustrated as follows:

$$C_k = \lambda_k \, EA^* + \varepsilon_k, \qquad k = 1, \dots, K \tag{3}$$

Where C_k, the *k*-th epigenetic clock, can be seen as an indicator of latent epigenetic age, denoted EA^*. λ_k is the factor loading parameter and ε_k is an stochastic error term.

In addition, we assume that the latent epigenetic age is linearly dependent on a vector of observable covariates, $X = \{X_1, \dots, X_P\}$. We can therefore specify the following behavioral equation:

$$EA^* = X'\gamma + \omega \tag{4}$$

Where γ is a $P \times 1$ vector of parameters and ω is an idiosyncratic error term. Plugging Equation (4) into Equation (3), we are thus able to estimate a system of *k* reduced form regressions of the kind:

$$C_k = \lambda_k \, X'\gamma + \lambda_k \omega + \varepsilon_k, \qquad k = 1, \dots, K \tag{5}$$

Conditional on the behavioral equation being correctly specified, SEM is generally more efficient because it use all the data at once to estimate parameters. This simultaneous estimation minimizes potential information loss and takes into account the full covariance structure of the data.

The estimation of the three methods outlined above with commonly used statistical software (e.g. Stata, R) automatically entails the standardization of the variables used to build the latent factor or, in the case of $MEGA_{SEM}$, the expression of the latent factor in the same scale of the predicted value of the first clock included in the measurement equation.[4] However, one of the advantages of epigenetic clocks is their interpretability in years of age, a metric that is both simple and of great public-policy relevance. In order to bring the MEGA clocks back to a unit of measure that can be interpreted in terms of years of age, we adopt a 'de-standardization' procedure, by manually computing the weighted averages in Equations (1) and (2) leaving the clocks $\{C_1, \dots, C_K\}$ in their natural scale, years of age.

[4] Here, we use Stata v. 17, and commands "sem" for $MEGA_{SEM}$, "factor" for $MEGA_{FA}$ and "egen weightave" for $MEGA_{WGT}$.

By combining information from different clocks, the approach we propose offers the advantage of reducing the measurement error linked to each clock individually, thus allowing us to capture the process of epigenetic aging more precisely. Given that epigenetic data availability is often afflicted by relatively small sample sizes, minimizing measurement error becomes particularly advantageous. This reduction has the potential to decrease the probability of type-two error, by mitigating the risk of attenuation bias in the estimation of the coefficients of interest.

4. The Avon Longitudinal Study of Parents and Children

4.1. Data Overview

The data for the proposed work originate from the Avon Longitudinal Study of Parents and Children (ALSPAC), also known as "Children of the 90s", an English birth cohort study designed to explore the effects of genetic, environmental, and social factors on the health and development of children and their families. The study initially enrolled over 14,000 pregnant women residing in the county of Avon, UK, with expected delivery dates between April 1991 and December 1992, resulting in 14,062 live births and 13,988 children surviving to one year of age. In the late 1990s, additional eligible mothers and children who had not joined the initial waves were recruited, bringing the total sample size to 15,447 pregnancies and 14,901 children alive at one year of age. This included 14,833 unique mothers as of September 2021, following further phases of recruitment.

ALSPAC employs a longitudinal design with multiple follow-up visits at key developmental stages. Data collection encompasses questionnaires, clinical assessments, biological sampling, and linkage to administrative records. This comprehensive approach captures a wide array of information, including physical and mental health, cognitive abilities, socioeconomic factors, and environmental exposures (Boyd *et al.*, 2013; Fraser *et al.*, 2013; Northstone *et al.*, 2019, 2023). Biological data collection includes genetic and epigenetic information, with blood samples taken at various time points. Longitudinal DNAm data for 1,022 children and their mothers have been collected using Illumina Infinium 450k methylation arrays. For children, DNAm was extracted from cord blood samples at birth and from peripheral blood samples at child ages 7 and 15-19 years. See Relton *et al.* (2015) for a detailed description of the Accessible Resource for Integrated Epigenomic Studies (ARIES), the sub-cohort of ALSPAC with DNA methylation data.

Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol (Harris *et al*., 2009). REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies. The ALSPAC study website contains detailed information on all the available data, through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data/).[5]

4.2. Parameterizing Child Abuse and Epigenetic Age

Of particular interest for this paper are two features of the ALSPAC dataset: the longitudinal availability of measures of child abuse and the collection of DNAm data. Based on the former, we are able to compute comprehensive measures of exposure to child abuse in different developmental stages, based on ratings from parents and children. The latter instead allows us to build epigenetic clocks, in order to test whether experiences of abuse throughout childhood correlate with accelerated epigenetic aging in late adolescence.

Child abuse. For child abuse, ALSPAC prospectively collected questions on child cruelty from mothers (M) and their partners (P).[6] Caregivers were asked to report physical and emotional cruelty towards the child coming from themselves or their partners several times throughout the data collection period.[7] In order to make carer-reported data more comparable to self-reported data on child abuse (described in the paragraph below), we define two developmental periods: one ranging from age 0 to 10 and the other from age 11 to 18. We then combine the carer-reported cruelty across periods and, for each rater r and period t, with $r \in \{M, P\}$ and $t \in \{(0-10), (11-18)\}$, we define exposure to cruelty as a binary variable equal to one if rater r reported any instances of child physical or emotional cruelty across the given period t and zero otherwise.

[5] Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004).

[6] Similar question on whether the child was exposed to sex abuse from anyone were collected too. However, due to the sensitive nature of these questions for parents, we have opted not to include them in this analysis.

[7] Mothers and their partners were asked to report child cruelty in the periods corresponding to the following child ages: 8 months (since birth), 1.75 years (since age 8 months), 2.75 years (since age 18 months), 4 years (since age 2.5), 5 years (in past year), 6 years (since age 5), 9 years (since age 8), 9-10 years, 11 years, and 17-18 years. While we have reports from mothers in all periods mentioned above, we only observe reports of child cruelty from the mothers' partners at child ages 6 years, 9-10 years and 11 years.

Once cohort children reached adulthood (22+ years), they were also retrospectively asked to report any instances of physical or sexual abuse. These retrospective questions asked for the frequency adults in the family were violent towards the study child in two time windows: before the age of 11 and between the ages of 11 and 17. These questions, rated on a 5-point Likert scale from 1 'Never' to 5 'Very often', are shown in more detail in the first six rows of Table A1, together with the value used as a cutoff for dichotomization. We build a measure of child self-reported cruelty in each of the two developmental periods that is equal to one if at least one of the six dichotomized self-reported cruelty variables is equal to one, and zero otherwise. Study children were also asked about sex abuse over the same two time periods (see last two rows of Table A1 for more details on the questions). Similar to before, we build one binary indicator of self-reported sex abuse for each developmental period, equal to one if the study child reported having experienced any sexual abuse at least once and zero otherwise.

Figure 1 shows the correlation between child and parent reported measures of child abuse for all ALSPAC children for whom the variables above are non-missing. While there are strong intertemporal and cross-dimensional correlations between reports from the same rater, cross-rater correlations are positive but lower in magnitude.[8] This is consistent with the literature on parental reporting versus child self-reporting of adverse childhood experiences, which highlights substantial discrepancies in reporting of instances of abuse. Caregiver reports often understate the severity and frequency of adverse events compared to children's self-reports, whether the caregiver is reporting their own or their partner's behavior (Fisher et al., 2011). Sibling corroboration indicates that self-reports are reliable. Newbury et al. (2018) and Baldwin et al. (2019) show that prospective parental reports and retrospective self-reports identify largely non-overlapping groups of maltreated individuals, with self-reports showing stronger associations with psychiatric problems. Using ALSPAC data, Houtepen et al. (2018) find that critical events like sexual abuse are underreported by parents, suggesting that children's self-reports might hold additional value. Soares et al. (2021) exclusively use retrospective self-reports due to the limitations of prospective parental data, while Warren et al. (2019) combine both, finding self-reports more significant in cases of sexual abuse. In light of the findings from the literature in developmental psychology described above, here, we adopt a comprehensive approach including both self-reports and

[8] The correlation table looks similar when restricting the sample to children with available DNAm information, which is how we define the estimation sample used to produce the main results of this paper (see Figure A1).

prospective parental reports to capture the full spectrum of childhood maltreatment, accounting for potential underreporting by parents and minimizing measurement error.

DNAm. Based on DNAm data, we calculate the epigenetic ages from the following four epigenetic clocks: Horvath (Horvath, 2013), Hannum (Hannum et al., 2013), PhenoAge (Levine et al., 2018) and GrimAge (Lu et al., 2019).[9] We first combine these clocks, in a novel approach to the measurement of epigenetic age (described in Section 3), and then use them separately, to test the validity of our approach.

4.3. Estimation Sample

The final estimation sample consists of 448 observations with available information on DNAm and measures of child abuse. Table 1 shows the prevalence of abuse in the estimation sample in two developmental periods (age 0-10 and age 11-18), as rated by the mother (header 'M'), the mother' partner ('P'), the child ('C'), and all possible combination of these three raters. Child abuse is divided between instances of child cruelty (emotional and physical), sex abuse and a combination of the two ('Any child abuse'). As we only have access to self-reported data on childhood sex abuse, in the 'Sex abuse' rows all cells but those in column (C) are empty.[10] Instances of child cruelty between child ages 0 and 10 range between 2% to 23% in the sample, depending on the rater – with children reporting higher prevalence than parents. When combining all raters together (our preferred measure), the prevalence of child cruelty goes up to 33.7%. Combining this with the 3.6% self-reported cases of sex abuse leads to up to 34.8% of children in the sample having experienced some form of emotional and/or physical cruelty or sex abuse before turning 11 years old.[11] When looking at abuse during adolescence (from age 11 onwards), we find again that children report more instances of abuse as compared to their parents, with the total number of those exposed to any form of abuse according to any rater reaching almost 20%. There is a relatively large persistence of abuse over time: 15.6% of children in the sample have experienced some form of abuse in both developmental periods, suggesting that almost half of

[9] As standard in the literature, we compute GrimAge using the PCGrimAge algorithm (Higgins-Chen et al., 2022).
[10] 'Any child abuse' is thus computed in each column using the 'Child cruelty' measure from the (group of) rater(s) indicated in the column and child-reported sex abuse.
[11] The relatively high incidence of child abuse could be explained by the socio-economic gradient in parents' reports. High-SES parents, which are disproportionately represented in the ALSPAC cohort, tend to report being cruel to their children more often than low-SES parents. In order to limit confounding coming from this source, all regressions in Section 5 control for maternal education and paternal social class.

those experiencing abuse in early childhood (0-10 years old) are also victim of abuse in adolescence.

Statistics describing the socio-demographic characteristics of the estimation sample, compared to the full ALSPAC sample, are reported in Table A2. As documented by Relton et al. (2015), children in the epigenetic subsample of the data come from positively selected families in terms of socio-economic characteristics. Standard *t*-tests of differences in means between our estimation sample and the general ALSPAC sample (reported in column 3 of Table A2) reveal that children in our sample are more likely to have slightly older and more educated mothers, and fathers with a higher social class, as compared to the average ALSPAC child. On top of this, our sample has a small gender unbalance (62% of girls as compared to boys) and is made up of a larger share of first-born children (50% against 44% on average).

5. Main Results

5.1. Features of the MEGA clocks

While computing $MEGA_{WGT}$ is fairly straightforward, $MEGA_{FA}$ relies on a theoretical assumption: only one factor, which we interpret as a measure of latent epigenetic age, retained from the exploratory factor analysis. When performing factor analysis in our estimation sample using the four epigenetic clocks described above (Horvath, Hannum, PhenoAge and GrimAge), only one factor appears to explain more variance in the data as compared to each single clock individually, i.e. only one factor satisfies the Kaiser criterion of having eigenvalue greater than one (Kaiser, 1960).[12] As shown in Table A3, all four clocks display factor loadings larger than 0.4, a threshold commonly used in the literature to retain items (Stevens, 2002), and all have more than 50% unique variance (i.e. variance that is not explained by the remaining clocks).

The distributions of the three MEGA clocks, together with the four original clocks and chronological age, are displayed in Figure 2. Descriptive statistics for the same variables can be found in Table 2. Child age in the estimation sample ranges from 14.6 to 19.3, following a bimodal distribution around ages 15 and 17 – the timing of the two clinical assessments collecting DNAm data in ALSPAC. The four traditional clocks are on average more dispersed and less centered around chronological age as compared to the MEGA clocks. When looking at their performance

[12] The selection of only one factor is unambiguous, as the second-largest factor has eigenvalue equal to 0.10.

in predicting chronological age in Figure A2, the Hannum and Horvath clock appear to be the most accurate, with the linear fit following closely the 45-degree line. GrimAge displays the largest shift in the intercept (GrimAge estimates of epigenetic age are on average 17.4 years higher than chronological age), but its level of dispersion is the smallest and its slope is the second-closest to one after Horvath. When looking at the predictive performance of the MEGA clocks against chronological age in Figure A3, all three methods exhibit a lower degree of dispersion as compared to the standard clocks and have slopes that are very close to 1. Consistently, $MEGA_{SEM}$ and $MEGA_{FA}$ display the highest correlation with age in our sample, closely followed by GrimAge and $MEGA_{WGT}$ (Figure A4). All in all, the MEGA clocks appear to be better predictors of chronological age than the traditional clocks.

5.2. Application 1: The Consequences of Child Abuse for Biological Aging

5.2.1. Estimating Equation

Next, we test the associations between measures of child abuse over time and the MEGA clocks defined, respectively, in sections 3.1 and 3.2. We do so by estimating the following linear regression model:

$$MEGA_{i,t} = \beta_0 + \beta_1 Abuse_{i,t} + \beta_2 Abuse_{i,t-1} + \gamma X_i + \varepsilon_{i,t} \quad (6)$$

Where $MEGA_{i,t}$ is one of the MEGA clocks for child *i* measured in period *t*, corresponding to late adolescence. $Abuse_{i,t}$ is a binary measure of exposure to child cruelty or child sexual abuse between child ages 11 and 18, as reported by either the mother, the mother's partner or the child herself. $Abuse_{i,t-1}$ is defined similarly for the time period going from the child's birth to age 10. Last, X_i is a vector of individual controls. Importantly, X_i includes the child's age at the time of the DNAm assessment used to build the MEGA clocks, meaning that we can interpret the outcome as a measure of age acceleration. All other controls are measured around the time of the child's birth. They include the mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order.

From the literature on the epigenetic effects of childhood adverse life experiences, we expect coefficients β_1 and β_2 to be positive: experiencing child cruelty and sex abuse likely correlates with increased age acceleration. However, little is known on the longitudinal exposure to abuse over the course of childhood. It is thus unclear *a priori* whether recent abuse will be more strongly

associated with age acceleration in late adolescence ($\beta_1 < \beta_2$) or if early childhood adversity will exhibit a stronger, long-lasting scarring effect ($\beta_1 > \beta_2$).

The identification assumptions to interpret β_1 and β_2 as causal effects would be rather strong, that is an absence of correlation between child abuse and the error term. While we do control for a wide set of individual and family control in X_i, there are likely other unobserved confounders that we are unable to control for. Estimates of β_1 and β_2 should thus be interpreted as conditional correlations, rather than causal effects.

5.2.2. Results

Table 3 reports the OLS estimates of the coefficients in Equation (3). Different versions of the MEGA clock are used across columns: $MEGA_{SEM}$ in column 1, $MEGA_{FA}$ in column 2 and $MEGA_{WGT}$ in column 3. From the table, results appear to be very similar across the different versions of the clock, indicating that experiencing any form of child abuse (sexual or cruelty) between ages 0 and 10 is associated with over half a year of accelerated epigenetic age. Abuse experienced after age 11 does not appear to be significantly associated with age acceleration – the coefficient being mostly negative and small in magnitude. The negative sign here is due to the strong intertemporal correlation of our measure of abuse. When we include only one of the two indicators for child abuse in the regression, the coefficients are always positive (albeit only significant for the 0-10 age range).

The half-a-year acceleration in epigenetic aging is comparable in magnitude to about half of the epigenetic-age premium for being a girl. Our effect size is within the estimated range of other studies that have looked at the association between early life adversity and age acceleration: Marini et al. (2020) estimate 1 to 2 months of accelerated aging at age 7 for ALSPAC children exposed to sexual and physical abuse, a number that our estimates suggest might compound over time. This compounding effect is consistent with the effect sizes found Lawn et al. (2018), who show that childhood exposure to sexual and physical abuse correlates with 2.7 to 3.4 higher epigenetic age as adults (29 to 47 years old).

As indicated in Table 4, when disaggregating child abuse into child cruelty and sex abuse, only child cruelty between ages 0 and 10 attracts a statistically significant estimate at the 5% level. The positive coefficient is similar in size to those displayed in Table 3, amounting to about half a year

greater age acceleration for children exposed to child cruelty by age 11. Estimated coefficients for sex abuse are positive for both periods, albeit statistically insignificant at conventional thresholds. Results from Table 3 appear thus to be driven by child cruelty more than sex abuse. However, as child cruelty constitutes the largest fraction of the overall abuse measure (as summarized in Table 1), we cannot exclude that we are underpowered to estimate a statistically significant coefficient for sex abuse – which only 3.6% of the sample reports having experienced between ages 0 and 10.

5.3. Application 2: Biological Aging and Human Capital

5.3.1. Estimating Equation

Our second empirical application deals with estimating the association between accelerated epigenetic aging in late adolescence and early-adulthood cognitive and non-cognitive outcomes. We do so in a similar fashion as equation (1), using the following linear regression model:

$$Y_{t+1} = \delta_0 + \delta_1 MEGA_{i,t} + \delta_2 Health_{i,t} + \theta X_i + \varepsilon_{i,t} \quad (7)$$

Where Y_{t+1} is one of the four outcomes measured in early adulthood (period *t+1*). All outcomes are binary and indicate negative outcomes, measured as far in time as possible within the ALSPAC data handout. Here, we consider two outcomes that are linked to the cognitive sphere, namely not having attained a university degree by age 26 and being neither in employment, education or training (NEET) at age 25. The remaining two outcomes relate more to the non-cognitive sphere and mental health: the first is based on being above the diagnostic threshold of the self-assessed Short Moods and Feelings Questionnaire (SMFQ) at age 25 (Angold et al., 1995), a widely used psychometric scale of mental health used as a screening tool for depressive symptoms; the second is having been diagnosed with depression by age 22. $MEGA_{i,t}$ is again one of the MEGA clocks for child *i* measured in period *t*, namely late adolescence. X_i is the same vector of controls defined above, while $Health_{i,t}$ is a vector of individual health and behavioral outcomes measured at age 15: BMI, being a smoker and drinking.

Similar to Application 1, here too we can only interpret δ_1 as a conditional association rather than the causal effect of one extra year of epigenetic age acceleration on early-adulthood outcomes. In an augmented version of equation (7), we additionally control for factors for cognitive and non-cognitive skills, measured at the same time as the MEGA clock and computed in a similar way, based on age-16 cognitive and non-cognitive assessments. The cognitive skills factor is based on

Mathematics, English, Science and aggregate test-scores, from linked administrative data on GCSE (or equivalent) exams. The non-cognitive skills factor, on the other hand, is based on child self-reported SMFQ and on four subscales of the mother-reported Strengths and Difficulties Questionnaire (SDQ; Goodman et al., 2000).[13]

5.3.2. Results

Results from equation (7) are illustrated in Figure 3. The figure shows that one extra year of epigenetic age acceleration is associated with a 3 percentage points (pp) increase in the probability of not having a university degree by age 26, a 1 pp increase in the probability of being NEET at 25, a 1.2 pp increase in the probability of being above the diagnostic SMFQ threshold at age 25 and between a 1 and a 2 pp increase in the probability of being diagnosed with depression by age 22. These effect sizes range from 8 to 16 percent of the mean value of the outcomes, providing evidence that a moderate worsening of economic and mental-health outcomes can be observed already in early adulthood for those with higher age acceleration in childhood. Results on educational attainment are consistent with Mareckova et al. (2023), who find evidence of a small negative association between Horvath age acceleration and IQ in a sample of young adult women.

Tables 5 and 6 show that age acceleration holds a predictive power over these outcomes that goes above and beyond the association between negative health outcomes or risky behaviors in adolescence and adult outcomes. The tables show estimates for coefficients δ_1 and δ_2 from Equation (7), both in a simplified version that does not control for adolescent health (columns 1, 3, and 5) and for the full model specification (columns 2, 4, and 6). Results from the tables reveal that the coefficients attached to the MEGA clocks are robust to the introduction of the *Health* control vector, suggesting that the association between epigenetic age acceleration and early-adulthood outcomes is not confounded by these other observables health outcomes.

In order to benchmark the effect sizes of the associations between the MEGA clocks and early-adulthood outcomes, we additionally include two additional latent factors to the measurement equation, measuring cognitive and non-cognitive skills respectively. Results are shown in Figure

[13] The SDQ is a 25-items questionnaire developed by psychologists that is used as a screening tool for socio-emotional and behavioral problems in children and adolescents. The four subscales we use here, each measured on a 0-10 discrete scale, are those that make up the composite 'total SDQ' score: emotional symptoms, hyperactivity/inattention, peer relationship problems, and conduct problems. Higher values correspond to greater problems in each of the areas indicated by the subscale title.

A5, where the MEGA clock weakly, but consistently predicts worse early-adulthood outcomes, consistent with Figure 3.[14] On the contrary, the cognitive and non-cognitive factors, built similar to the MEGA clocks as either latent factors from SEM, factor analysis or covariance weighting, predict better early-adulthood outcomes. The relationships are, as expected, stronger between cognitive skills and educational attainment and between non-cognitive skills and mental health outcomes. While these effect sizes are on average larger in magnitude than those attached to the MEGA clocks, the latter remain in the ballpark of the point estimates linked to non-cognitive skills – suggesting that these modest (and somewhat imprecise) effect sizes might actually be economically relevant.

5.4. Sensitivity tests

We first test whether the results from Table 3 are similar when using traditional clocks instead of the MEGA clock as dependent variables in Equation (6). Figure 4 reports estimates of coefficients β_1 and β_2 for each of the outcomes indicated in the legend. First, we show results for each of the single clocks used to build the MEGA clock (blue round markers) and, second, we report the coefficients from Table 3 for convenience (red diamond markers). Estimates of child abuse before age 11 using the traditional clocks display on average larger magnitudes than those estimated with the MEGA clocks, with point estimates ranging from 0.52 to 1.00. The MEGA clocks produce results that are the closest to the GrimAge clock (the one they correlate with the most, as shown in Figure A4), but have smaller standard errors on average, suggesting that – all else equal – our aggregating procedures do indeed reduce measurement error as compared to using any one of the single clocks separately. In Figure A6, we replicate the same exercise but disaggregating child abuse into its two components: cruelty and sex abuse (similar to Table 3). Results are again consistent with those from Figure 4, with the MEGA clocks displaying on average smaller confidence intervals than the traditional clocks and point estimates aligned with the GrimAge clock.

Due to the ease of interpretation of the dependent variable in its natural scale (years of age), we have so far left clocks in their natural scale. However, for a subset of the MEGA clock procedures, this implied a manual de-standardization. Due to standardization entering the procedures at

[14] The MEGA clocks and the cognitive and non-cognitive factors are standardized in the figure, to enhance comparability.

different moments (see Appendix B for more details), the variance captured by the standardized MEGA clocks slightly differs from that of the unstandardized ones. In order to test whether our results depend on the choice of scale of the dependent variables, in Tables A4 and A5 respectively we replicate Tables 3 and 4 with dependent variables standardized to have mean zero and standard deviation one. Standardized results are robust across specifications and clocks. Specifically, results in Table A4 indicate that the half-a-year age acceleration penalty associated with child abuse before age 11 can be alternatively read as a 22 to 25% of a standard deviation increase in age acceleration.

In addition, in order to test whether results are sensitive to the exclusion of either one clock used in the MEGA algorithms, we compute leave-one-out versions of the MEGA clocks. Exploratory factor analysis confirms a uni-factor model for any combination of three out of the four clocks. Results for the first empirical application, displayed in Figure A7, are robust to the exclusion of either one clock. As expected, this comes at the expenses of precision, the loss of which is greater when excluding the GrimAge clock. All point estimates for early-childhood abuse are remarkably stable and they remain positive and statistically different from zero at least at the 10% level. Similarly, Figure A8 applies this leave-one-out computational strategy to the second empirical exercise. Here too the point estimates from even-numbered columns of Tables 5 and 6 remain quite robust when excluding either one of the Hannum, Horvath or PhenoAge clocks. Interestingly, point estimates converge to zero when excluding the GrimAge clock from the MEGA, indicating that most of the associations between epigenetic age acceleration and early-adulthood outcomes that we observe run through this clock.

We then turn to our definition of the exposure of interest in our first empirical application, child abuse. As argued in section 4.1, our preferred measure of abuse includes information from all available raters, namely mothers, their partners and the child herself. Table A6 shows that our results in Table 3 are not driven by this measurement choice: the coefficients attached to abuse between ages 0 and 10 are stable across raters and combination of raters, roughly ranging from 0.3 to 0.5. The use of reports from all raters (last column of Table A6) comes with the lowest standard errors across all methods, suggesting that harnessing these different sources of information can help reducing the measurement error linked to the under-reporting of sensitive constructs.

6. Mechanisms

Why do children who experience abuse in childhood age faster? One possible biological mechanism involves changes in the structure of blood cells such as leukocytes, which are predicting of epigenetic aging (Lima et al., 2022). Different blood cell types have distinct methylation profiles, which is a reason why blood cell counts are typically partialled out in epigenetic studies. However, it can be argued that these counts are 'bad controls' (Angrist and Pischke, 2009), as they may mediate the relationship between environmental exposures and the epigenome. Child abuse, in particular, has been consistently linked to changes in immune cell proportions, driven by increased inflammatory activity (D'Elia et al., 2018; Renna et al., 2021).

To test whether cell counts mediate the relationship between child abuse and epigenetic aging in our sample, we compute standard blood cell counts for peripheral blood samples from the age 15-19 DNAm data in ALSPAC,[15] using the Houseman et al. (2012) method. In our estimation sample, children exposed to abuse at least once display substantially larger associations between blood cell counts and age acceleration (Figure A9). When controlling for blood cell counts in our main regressions, the magnitude of the estimated coefficients decreases by approximately one-third (Table A7). While none of the estimates in Table A7 are significantly different from their counterparts in Table 3, this result is suggestive of the fact that blood cell counts may play a mediating role in the association between childhood abuse and age acceleration.

Due to the limited availability of other biological biomarkers in our sample, directly assessing the presence of other biological channels is challenging. Nevertheless, we can draw from literature in molecular epigenetics to highlight plausible mechanisms that explain the associations we find between epigenetic aging, childhood exposures and early-adulthood outcomes.

Related to our first empirical application, differences in how people respond to stress and the symptoms they experience can shape the epigenetic response to traumatic childhood experiences. Biological and behavioral mechanisms are both likely to be in place. Beyond the changes in cell proportions discussed above, stress response is linked to the functioning of the hypothalamic-pituitary-adrenal axis and cortisol production, which are in turn associated with higher epigenetic aging (Dammering et al., 2021; Suarez et al., 2018). Psychological resilience can additionally

[15] The cell types considered here are B-lymphocytes (Bcell), CD4+ T-lymphocytes (CD4T), CD8+ T-lymphocytes (CD8T), granulocytes (Gran), monocytes (Mono), and natural killer (NK).

moderate stress-related health impacts; individuals with stronger emotion regulation and self-control tend to be better protected from stress-induced epigenetic aging (Harvanek et al., 2021; Harvanek et al., 2023). Research on posttraumatic stress disorder (PTSD) suggests that specific symptom groups, such as emotional withdrawal, sleep problems, and cognitive dysfunction, are more closely tied to changes in epigenetic aging than overall PTSD severity (Katrinli et al., 2020; Na et al., 2022). In addition, risky behaviors and lifestyle changes can also play a role in how stress affects epigenetic aging (Schmitz et al., 2022; Jung et al., 2023). Biological and behavioral stress responses have been hypothesized to affect other hallmarks of aging, such as circadian rhythms, immune system functioning and nutrition, which in turn imprint onto the epigenome (Harvanek et al., 2023).

Linked to our second empirical application, we review the literature to investigate the biological rationale through which more accelerated epigenetic aging predicts worse cognitive and non-cognitive outcomes in adulthood. The machine learning approaches used to develop epigenetic clocks are fundamentally agnostic to the underlying biology, so the biological gene functions and processes (biological annotation) reflects multiple, interconnected mechanisms rather than a single one. The biological annotation of the CpG sites used in the Horvath, Hannum, PhenoAge and GrimAge clocks have been shown to regulate biological processes such as cell death and survival (Horvath, 2013), cellular growth and proliferation (Horvath, 2013; Hannum et al., 2013), locomotion (Hannum et al., 2013), inflammation (Levine et al., 2018), antiviral response (Levine et al., 2018) and DNA damage recognition and repair (Levine et al., 2018). These CpGs are over-represented in gene sets involved in the immune system, lipid function and adipocytes communication (Lu et al., 2019).

Although these aging traits may not have been clinically present yet in the young adults from the ALSPAC sample, the underlying biological processes could already be detectable and correlate with early signs of cognitive decline (Felt et al., 2023). While DNAm has been shown to contribute to learning processes and memory, through the regulation of the central nervous system, there is scarce evidence on the link between epigenetic clocks and the biological processes involved in cognitive and non-cognitive skills maintenance and production. One example is the work of Han et al. (2018), who find that three out of ten most involved gene regions regulate neuronal pathways (e.g. neurogenesis, neuron differentiation, neuron death).

7. Conclusion

This study illustrates the potential of epigenetic clocks as valuable tools in economic research for understanding the biological underpinnings of socioeconomic influences on health and behavior. We developed the Multi EpiGenetic Age (MEGA) clock, a novel metric that combines four established epigenetic clocks to improve the accuracy and reliability of epigenetic age measurements. Our empirical applications using data from a cohort of children in the UK reveal important findings: first, early-life exposure to child abuse is associated with accelerated epigenetic aging in adolescence, highlighting the long-term biological impact of early-life adversity; second, accelerated epigenetic aging in adolescence predicts worse cognitive and non-cognitive outcomes in early adulthood, suggesting broader implications for health and productivity.

These findings emphasize the importance of considering biological measures alongside traditional socioeconomic indicators in economic research. The MEGA clock offers an easily interpretable and reliable tool for understanding how environmental factors shape biological aging processes, providing a pathway to more targeted and effective policy interventions. Future research should continue to refine these measures and explore their applications across diverse populations and contexts. By bridging the gap between biology and economics, our work contributes to a more comprehensive understanding of human development and the intricate interplay between nature and nurture.

References

Abdellaoui, Abdel, Loic Yengo, Karin J. H. Verweij, and Peter M. Visscher. “15 Years of GWAS Discovery: Realizing the Promise.” The American Journal of Human Genetics 110, no. 2 (February 2, 2023): 179–94. https://doi.org/10.1016/j.ajhg.2022.12.011.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association,* 103 (484), 1481–95.

Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescent. *International Journal of Methods in Psychiatric Research*, 5, 237–249.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist’s companion*. Princeton University Press.

Baldwin, J. R., Reuben, A., Newbury, J. B., & Danese, A. (2019). Agreement between prospective and retrospective measures of childhood maltreatment: a systematic review and meta-analysis. *JAMA psychiatry*, 76(6), 584-593.

Beauchamp, J. P., Cesarini, D., & Johannesson, M. (2017). The psychometric and empirical properties of measures of risk preferences. *Journal of Risk and Uncertainty*, 54, 203-237.

Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., ... & Rakyan, V. K. (2019). DNA methylation aging clocks: challenges and recommendations. *Genome Biology*, 20, 1-24.

Benjamin, Daniel J, Christopher F Chabris, Edward L Glaeser, Vilmundur Gudnason, Tamara B Harris, David I Laibson, Lenore J Launer, and Shaun M Purcell. “Genoeconomics.” In Biosocial Surveys, edited by Maxine Weinstein, James W Vaupel, and Kenneth W Wachter, 304–35. Washington, D.C.: The National Academies Press, 2007.

Benjamin, Daniel J., David Cesarini, Patrick Turley, and Alexander Strudwick Young. “Social-Science Genomics: Progress, Challenges, and Future Directions.” Working Paper. Working Paper Series. National Bureau of Economic Research, May 2024. https://doi.org/10.3386/w32404.

Björklund, A., Lindahl, M., & Plug, E. (2006). The origins of intergenerational associations: Lessons from Swedish adoption data. *Quarterly Journal of Economics*, 121, 999-1028.

Boyd, A., et al. (2013). Cohort profile: The “Children of the 90s”-the index offspring of the Avon longitudinal study of parents and children. *International Journal of Epidemiology*, 42, 111–127.

Cecil, C. A., Zhang, Y., & Nolte, T. (2020). Childhood maltreatment and DNA methylation: A systematic review. *Neuroscience & Biobehavioral Reviews*, 112, 392-409.

Cesarini, D., Johannesson, M., Lichtenstein, P., Sandewall, Ö., & Wallace, B. (2010). Genetic variation in financial decision-making. *Journal of Finance*, 65, 1725-1754.

Cole, Steven W. "Social Regulation of Human Gene Expression." Current Directions in Psychological Science 18, no. 3 (June 1, 2009): 132–37. https://doi.org/10.1111/j.1467-8721.2009.01623.x.

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3), 883-931.

Currie, J., & Spatz Widom, C. (2010). Long-term consequences of child abuse and neglect on adult economic well-being. *Child Maltreatment*, 15, 111-120.

D'Elia, A. T., Matsuzaka, C. T., Neto, J. B., Mello, M. F., Juruena, M. F., & Mello, A. F. (2018). Childhood sexual abuse and indicators of immune activity: A systematic review. *Frontiers in Psychiatry*, 9, 354.

Dammering, F., Martins, J., Dittrich, K., Czamara, D., Rex-Haffner, M., Overfeld, J., ... & Heim, C. (2021). The pediatric buccal epigenetic clock identifies significant ageing acceleration in children with internalizing disorder and maltreatment exposure. *Neurobiology of Stress*, 15, 100394.

Davillas, A., & Jones, A. M. (2024*). Biological Age and Predicting Future Health Care Utilisation.* Institute of Labor Economics (IZA) Discussion Paper No. 17159.

Dugué, P. A., Bassett, J. K., Joo, J. E., Jung, C. H., Ming Wong, E. E., Moreno-Betancur, M., ... & Milne, R. L. (2018). DNA methylation-based biological aging and cancer risk and survival: Pooled analysis of seven prospective studies. *International Journal of Cancer*, 142(8), 1611-1619.

Faul, J. D., Kim, J. K., Levine, M. E., Thyagarajan, B., Weir, D. R., & Crimmins, E. M. (2023). Epigenetic-based age acceleration in a representative sample of older Americans: Associations with aging-related morbidity and mortality. *Proceedings of the National Academy of Sciences*, 120(9), e2215840120.

Felt, J. M., Yusupov, N., Harrington, K. D., Fietz, J., Sliwinski, M. J., Ram, N., ... & BeCOME Working Group. (2023). Epigenetic age acceleration as a biomarker for impaired cognitive abilities in adulthood following early life adversity and psychiatric disorders. *Neurobiology of Stress*, 27, 100577.

Fletcher, J. M. (2009). Childhood mistreatment and adolescent and young adult depression. *Social Science and Medicine*, 68, 799–806.

Fisher, H.L., Bunn, A., Jacobs, C., Moran, P., Bifulco, A. (2011). Concordance between mother and offspring retrospective reports of childhood adversity. *Child Abuse & Neglect*, 35 (2), 117-122.

Fraser, A., et al. (2013). Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology*, 42, 97–110.

Goldberger, Arthur Stanley. *Twin methods: A skeptical view*. Social Systems Research Institute, University of Wisconsin-Madison, 1976.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177(6), 534-539.

Hannum, G., Guinney, J., Zhao, L., Zhang, L. I., Hughes, G., Sadda, S., ... & Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 49(2), 359-367.

Harris, P. A., et al. (2009), Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42, 377-381.

Harvanek, Z. M., Fogelman, N., Xu, K., & Sinha, R. (2021). Psychological and biological resilience modulates the effects of stress on epigenetic aging. Translational Psychiatry, 11(1), 601.

Harvanek, Z. M., Boks, M. P., Vinkers, C. H., & Higgins-Chen, A. T. (2023). The cutting edge of epigenetic clocks: In search of mechanisms linking aging and mental health. *Biological Psychiatry*, 94(9), 694-705.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411-482.

Henkhaus, L. E. (2022). The lasting consequences of childhood sexual abuse on human capital and economic well-being. *Health Economics*, 31, 1954-1972.

Higgins-Chen, A. T., Thrush, K. L., Wang, Y., Minteer, C. J., Kuo, P. L., Wang, M., ... & Levine, M. E. (2022). A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. *Nature Aging*, 2(7), 644-661.

Hoffmann, A., & Spengler, D. (2012). DNA memories of early social life. *Neuroscience*, 4, 64-75.

Houmark, M. A., Ronda, V., & Rosholm, M. (2024). The nurture of nature and the nature of nurture: How genes and investments interact in the formation of skills. *American Economic Review*, 114, 385-425.

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., ... & Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, 13, 1-16.

Houtepen, L. C., Heron, J., Suderman, M. J., Tilling, K., & Howe, L. D. (2018). Adverse childhood experiences in the children of the Avon Longitudinal Study of Parents and Children (ALSPAC). *Wellcome Open Research*, 3, 106.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, 14, 1-20.

Jung, J., McCartney, D. L., Wagner, J., Yoo, J., Bell, A. S., Mavromatis, L. A., ... & Lohoff, F. W. (2023). Additive effects of stress and alcohol exposure on accelerated epigenetic aging in alcohol use disorder. *Biological Psychiatry*, 93(4), 331-341.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151.

Katrinli, S., Stevens, J., Wani, A. H., Lori, A., Kilaru, V., van Rooij, S. J., ... & Smith, A. K. (2020). Evaluating the impact of trauma and PTSD on epigenetic prediction of lifespan and neural integrity. *Neuropsychopharmacology*, 45(10), 1609-1616.

Klengel, T., Pape, J., Binder, E. B., & Mehta, D. (2014). The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology*, 80, 115-132.

Korous, K. M., Surachman, A., Rogers, C. R., & Cuevas, A. G. (2023). Parental education and epigenetic aging in middle-aged and older adults in the United States: a life course perspective. *Social Science and Medicine*, 333, 116173.

Kotschy, R., Bloom, D. E., & Scott, A. J. (2024). *On the limits of chronological age*. NBER Working Paper no. 33124.

Lawn, R. B., Anderson, E. L., Suderman, M., Simpkin, A. J., Gaunt, T. R., Teschendorff, A. E., ... & Howe, L. D. (2018). Psychosocial adversity and socioeconomic position during childhood and epigenetic age: analysis of two prospective cohort studies. *Human Molecular Genetics*, 27, 1301-1308.

Leibowitz, Arleen. “Home Investments in Children.” Journal of Political Economy 82, no. 2, Part 2 (1974): S111–31.

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., ... & Horvath, S. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*, 10(4), 573.

Lima, C. N., Suchting, R., Scaini, G., Cuellar, V. A., Del Favero-Campbell, A., Walss-Bass, C., ... & Fries, G. R. (2022). Epigenetic GrimAge acceleration and cognitive impairment in bipolar disorder. *European Neuropsychopharmacology*, 62, 10-21.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell*, 153, 1194–1217. doi: 10.1016/j.cell.2013.05.039.

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., ... & Horvath, S. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*, 11(2), 303.

Lussier, A. A., Zhu, Y., Smith, B. J., Cerutti, J., Fisher, J., Melton, P., ... & Dunn, E. C. (2023). The timing of childhood adversity associates with epigenetic patterns across childhood and adolescence: results from a prospective, longitudinal study. *The Lancet Child & Adolescent Health*, 7, 532.

Maddock, J., Castillo-Fernandez, J., Wong, A., Cooper, R., Richards, M., Ong, K. K., ... & Hardy, R. (2020). DNA methylation age and physical and cognitive aging. *The Journals of Gerontology: Series A*, 75(3), 504-511.

Mareckova, K., Pacinkova, A., Marecek, R., Sebejova, L., Izakovicova Holla, L., Klanova, J., ... & Nikolova, Y. S. (2023). Longitudinal study of epigenetic aging and its relationship with brain aging and cognitive skills in young adulthood. *Frontiers in Aging Neuroscience*, 15, 1215957.

Marini, S., Davis, K. A., Soare, T. W., Zhu, Y., Suderman, M. J., Simpkin, A. J., ... & Dunn, E. C. (2020). Adversity exposure during sensitive periods predicts accelerated epigenetic aging in children. *Psychoneuroendocrinology*, 113, 104484.

Martinez, R.A.M., Green Howard, A., Fernández-Rhodes, L., Maselko, J., Pence, R.W., Dhingra,R., Aiello, A.E. (2024). *Addressing conceptual and measurement disagreement of biological age: A confirmatory factor analysis*. Mimeo.

McCartney, D. L., Stevenson, A. J., Walker, R. M., Gibson, J., Morris, S. W., Campbell, A., ... & Marioni, R. E. (2018). Investigating the relationship between DNA methylation age acceleration and risk factors for Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, 429-437.

McCrory, C., Fiorito, G., Hernandez, B., Polidoro, S., O'Halloran, A. M., Hever, A., ... & Kenny, R. A. (2021). GrimAge outperforms other epigenetic clocks in the prediction of age-related clinical phenotypes and all-cause mortality. *The Journals of Gerontology: Series A*, 76(5), 741-749.

Na, P. J., Montalvo-Ortiz, J. L., Nagamatsu, S. T., Southwick, S. M., Krystal, J. H., Gelernter, J., & Pietrzak, R. H. (2022). Association of symptoms of posttraumatic stress disorder and GrimAge, an epigenetic marker of mortality risk, in US military veterans. *Journal of Clinical Psychiatry*, 83(4), 41671.

Newbury, J.B., Arseneault, L., Moffitt T,E., Caspi, A. Danese A., Baldwin J.R., Fisher, H.L. 2018. Measuring childhood maltreatment to predict early-adult psychopathology: Comparison of prospective informant-reports and retrospective self-reports. *Journal of Psychiatric Research*, 96, 57-64. https://doi.org/ 10.1016/j.jpsychires.2017.09.020

Niccodemi, G., Menta, G., Turner, J., & D'Ambrosio, C. (2022). Pace of aging, family environment and cognitive skills in children and adolescents. SSM-Population Health, 20, 101280.

Northstone, K., Lewcock, M., Groom, A., Boyd, A., Macleod, J., Timpson, N., & Wells, N. (2019). The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Research*, 4, 15132.1.

Northstone, K., Shlomo, Y. B., Teyhan, A., Hill, A., Groom, A., Mumme, M., ... & Golding, J. (2023). The Avon Longitudinal Study of Parents and Children ALSPAC G0 Partners: A cohort profile. *Wellcome Open Research*, 8, 18782.2.

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., ... & Young, A. I. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, 54, 437-449.

Papageorge, N. W., & Thom, K. (2020). Genes, education, and labor market outcomes: evidence from the health and retirement study. *Journal of the European Economic Association*, 18, 1351-1399.

Perna, L., Zhang, Y., Mons, U., Holleczek, B., Saum, K. U., & Brenner, H. (2016). Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clinical Epigenetics*, 8, 1-7.

Petrovic, D., Carmeli, C., Sandoval, J. L., Bodinier, B., Chadeau-Hyam, M., Schrempft, S., ... & Stringhini, S. (2023). Life-course socioeconomic factors are associated with markers of epigenetic aging in a population-based study. *Psychoneuroendocrinology*, 147, 105976.

Raffington, L., Tanksley, P. T., Sabhlok, A., Vinnik, L., Mallard, T., King, L. S., ... & Tucker-Drob, E. M. (2023a). Socially stratified epigenetic profiles are associated with cognitive functioning in children and adolescents. *Psychological Science*, 34, 170-185.

Raffington, L., Tanksley, P. T., Vinnik, L., Sabhlok, A., Patterson, M. W., Mallard, T., ... & Harden, K.P. (2023b). Associations of DNA-Methylation Measures of Biological Aging With Social Disparities in Child and Adolescent Mental Health. *Clinical Psychological Science*, 21677026231186802.

Renna, M. E., Peng, J., Shrout, M. R., Madison, A. A., Andridge, R., Alfano, C. M., ... & Kiecolt-Glaser, J. K. (2021). Childhood abuse histories predict steeper inflammatory trajectories across time. *Brain, Behavior, and Immunity*, 91, 541-545.

Relton, C. L., Gaunt, T., McArdle, W., Ho, K., Duggirala, A., Shihab, H., ... & Davey Smith, G. (2015). Data resource profile: accessible resource for integrated epigenomic studies (ARIES). *International Journal of Epidemiology*, 44, 1181–1190.

Rubens, M., Bruenig, D., Adams, J. A., Suresh, S. M., Sathyanarayanan, A., Haslam, D., ... & Mehta, D. (2023). Childhood maltreatment and DNA methylation: a systematic review. *Neuroscience & Biobehavioral Reviews*, 147, 105079.

Schmitz, L. L., Zhao, W., Ratliff, S. M., Goodwin, J., Miao, J., Lu, Q., ... & Smith, J. A. (2022). The socioeconomic gradient in epigenetic ageing clocks: evidence from the multi-ethnic study of atherosclerosis and the health and retirement study. *Epigenetics*, 17(6), 589-611.

Soares, A.G., Zimmerman, A., Zammit, S., Karl, A., Halligan, S.L., & Fraser, A. (2021). Abuse in Childhood and Cardiometabolic Health in Early Adulthood: Evidence From the Avon Longitudinal Study of Parents and Children. *Journal of the American Heart Association*, 10, e021701.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences (Vol. 4)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Suarez, A., Lahti, J., Czamara, D., Lahti-Pulkkinen, M., Girchenko, P., Andersson, S., ... & Raikkonen, K. (2018). The epigenetic clock and pubertal, neuroendocrine, psychiatric, and cognitive outcomes in adolescents. *Clinical Epigenetics*, 10, 1-12.

Suglia, S. F., Clark, C. J., Boynton-Jarrett, R., Kressin, N. R., & Koenen, K. C. (2014). Child maltreatment and hypertension in young adulthood. *BMC Public Health*, 14, 1149.Szyf, Moshe. "The Early Life Environment and the Epigenome." Biochimica et Biophysica Acta 1790, no. 9 (September 2009): 878–85. https://doi.org/10.1016/j.bbagen.2009.01.009.

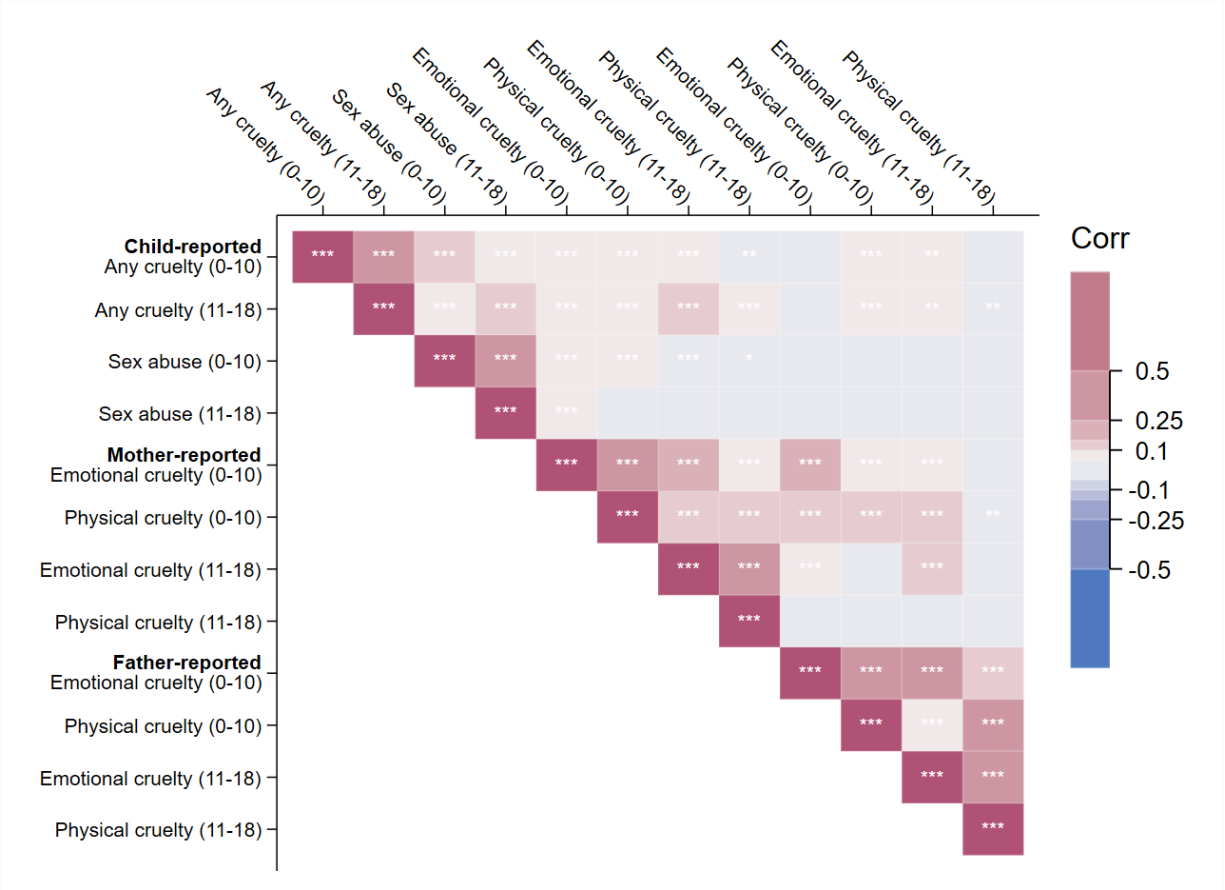
Taubman, Paul. (1976). Earnings, Education, Genetics, and Environment. *Journal of Human Resources* 11, no. 4: 447–61.

Warren, A.-S., Goldsmith, K. A., & Rimes, K. A. (2019). Childhood gender-typed behavior and emotional or peer problems: A prospective birth-cohort study. *Journal of Child Psychology and Psychiatry*, 60(8), 888–896. https://doi.org/10.1111/jcpp.13051.

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., ... & Lee, J. Y. (2022). A saturated map of common genetic variants associated with human height. *Nature*, 610, 704-712.

Figures and Tables

Figure 1: Cross-rater, intertemporal correlations of child cruelty and sex abuse



Notes: the correlation matrix is based on the sample of 3937 children in ALSPAC for whom all measures above are available.

Figure 2: Distribution of epigenetic age by clock and chronological age

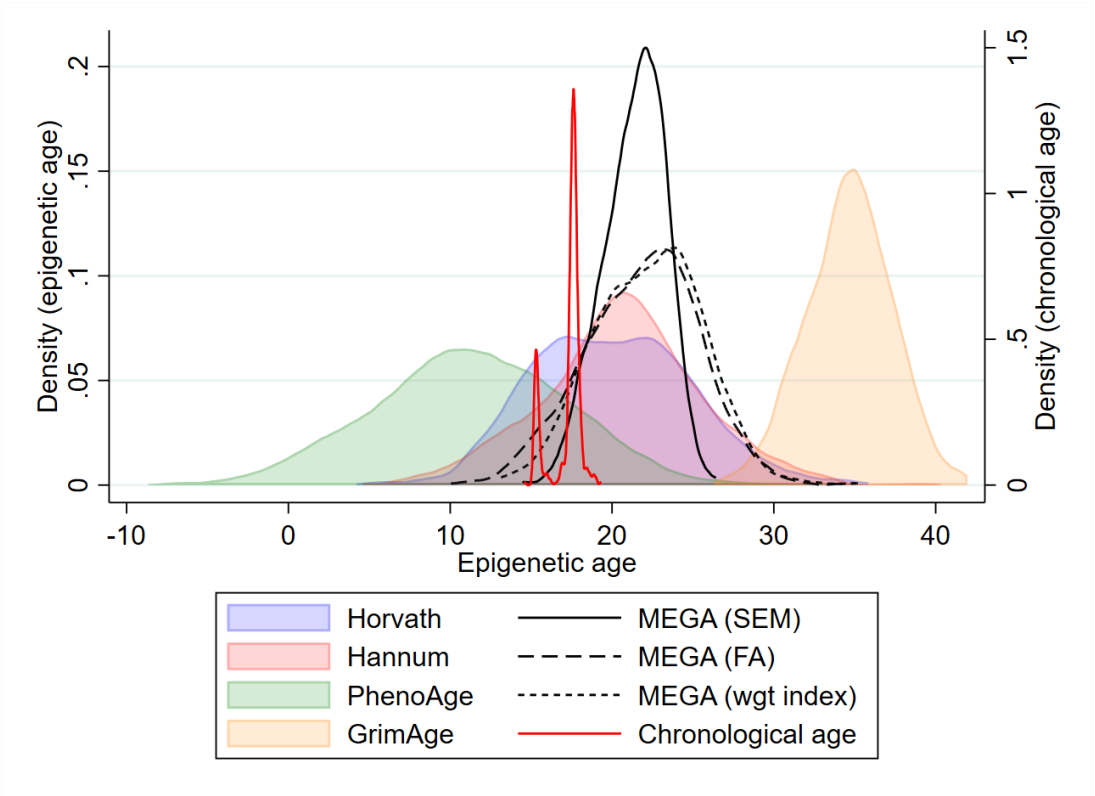
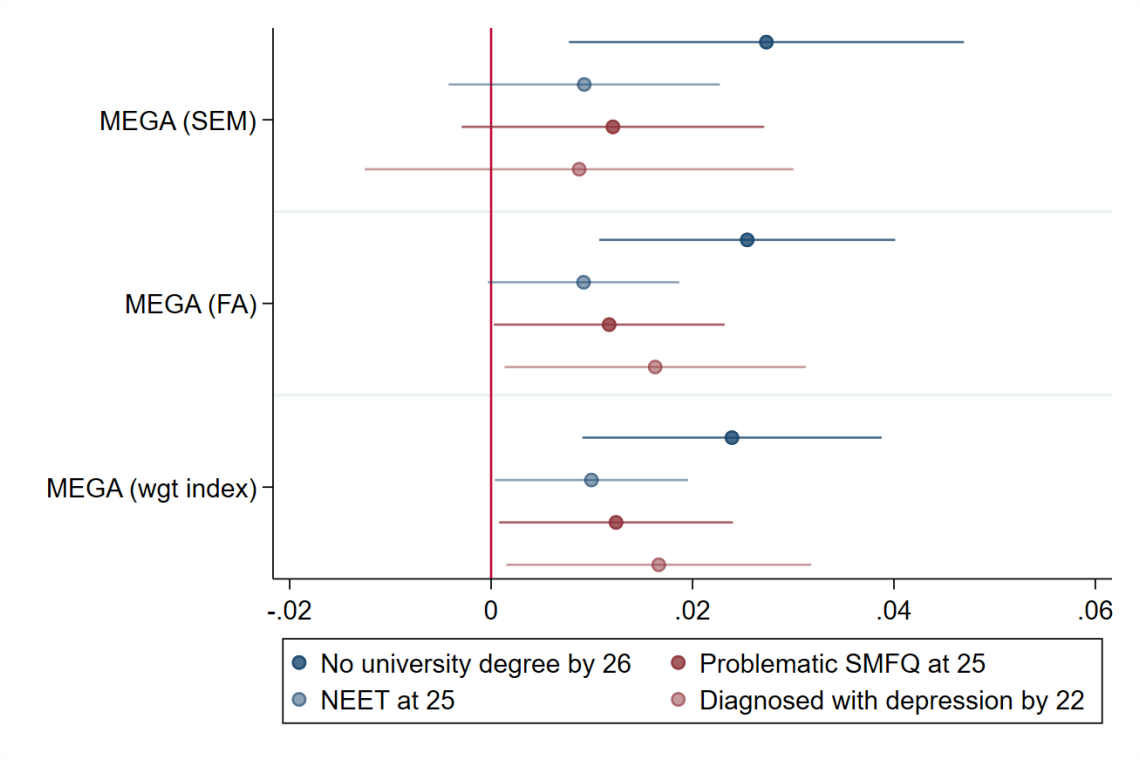
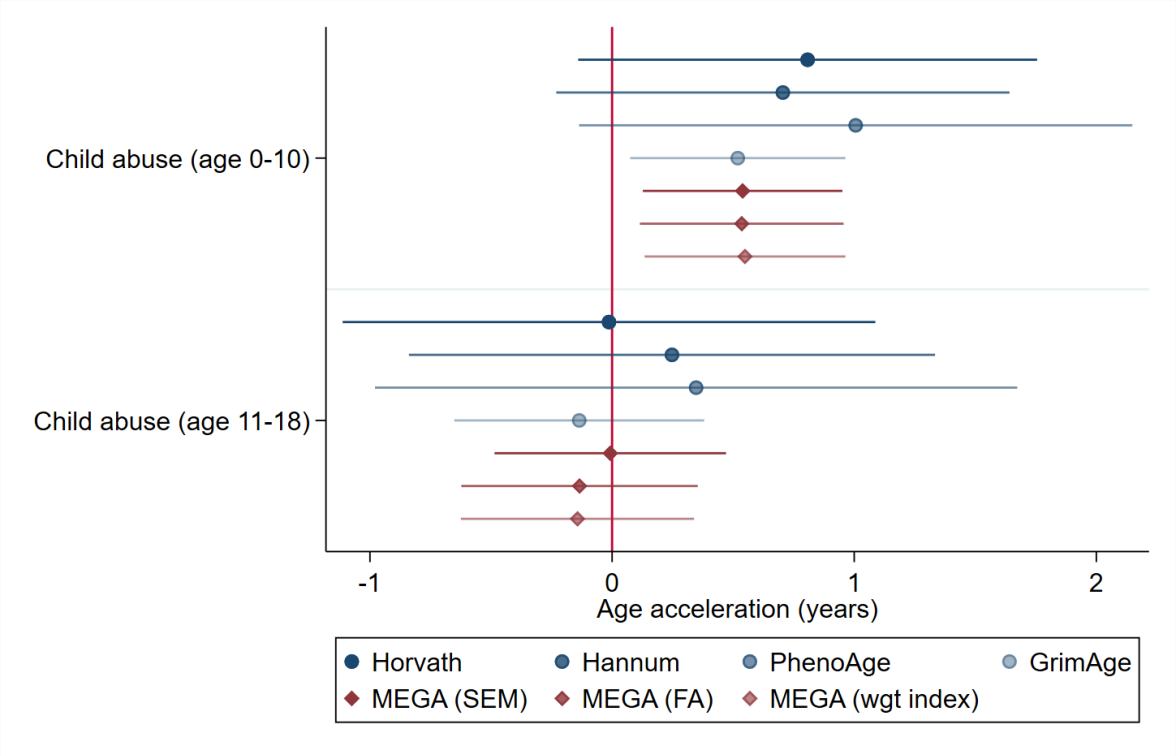


Figure 3: Age acceleration and early-adulthood outcomes



Notes: The figure displays point estimates for the MEGA clocks from columns (2), (4), and (6) of Tables 5 and 6. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. Health outcomes (BMI, smoking and drinking) are additionally controlled for. Spikes are for 90% confidence intervals.

Figure 4: Child abuse and age acceleration across traditional clocks and MEGA clocks



Notes: The figure replicates results from Table 2 using the single clocks as dependent variables instead of the MEGA clock. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. Spikes are for 90% confidence intervals.

Table 1: Prevalence of abuse in the estimation sample

	M (1)	P (2)	C (3)	MP (4)	CM (5)	CP (6)	CMP (7)
Age 0-10							
Child cruelty	13.8%	2.0%	23.0%	15.3%	33.3%	23.9%	33.7%
Sex abuse			3.6%				
Any child abuse	15.5%	5.6%	24.6%	17.9%	34.4%	25.4%	34.8%
Age 11-18							
Child cruelty	4.7%	0.4%	12.1%	4.9%	15.6%	12.5%	15.8%
Sex abuse			5.4%				
Any child abuse	9.4%	5.8%	16.5%	9.6%	19.4%	17.0%	19.6%

Notes: The table reports the prevalence of abuse in the estimation sample of 448 observations. Letters in the column headers indicate the person who reported the measure of abuse: 'M' is for mothers, 'P' is for the mother's partner, and 'C' is for the child.

Table 2: Descriptive statistics of chronological age and epigenetic age

	Mean	Std. Dev.	Min	Max
Chronological age	17.23	0.95	14.6	19.3
Traditional clocks				
Horvath	19.97	4.95	4.2	35.8
Hannum	20.62	5.03	4.6	40.3
PhenoAge	11.29	6.15	-8.7	34.7
GrimAge	34.64	2.67	26.3	42.0
MEGA clocks				
$MEGA_{SEM}$ (GrimAge)	21.40	1.90	14.5	26.4
$MEGA_{SEM}$ (Horvath)	16.45	1.46	11.1	20.3
$MEGA_{SEM}$ (Hannum)	33.27	2.95	22.5	41.1
$MEGA_{SEM}$ (PhenoAge)	34.54	3.07	23.3	42.7
$MEGA_{FA}$	33.78	2.55	26.3	40.8
$MEGA_{WGT}$	33.19	2.50	26.6	40.6

Notes: Descriptive statistics refer to the estimation sample of 448 observations.

Table 3: Child abuse and age acceleration from the MEGA clock

	SEM (1)	FA (2)	Weighted index (3)
Any child abuse (0-10)	0.539**	0.536**	0.549**
	(0.251)	(0.255)	(0.252)
Any child abuse (11-18)	-0.007	-0.134	-0.143
	(0.291)	(0.296)	(0.292)
Observations	448	448	448
Adjusted R-squared	.	0.272	0.264

Notes: Standard errors in parentheses. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Child abuse and age acceleration from the MEGA clock: disaggregation

	SEM (1)	FA (2)	Weighted index (3)
Any child cruelty (0-10)	0.546**	0.543**	0.559**
	(0.254)	(0.259)	(0.255)
Any sex abuse (0-10)	0.253	0.270	0.267
	(0.568)	(0.580)	(0.572)
Any child cruelty (11-18)	-0.054	-0.228	-0.214
	(0.314)	(0.319)	(0.315)
Any sex abuse (11-18)	0.281	0.377	0.358
	(0.468)	(0.477)	(0.471)
Observations	448	448	448
Adjusted R-squared	.	0.272	0.263

Notes: Standard errors in parentheses. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Age acceleration and early-adulthood cognitive outcomes

	SEM		FA		Weighted Index	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. No University degree by 26						
MEGA	0.028**	0.027**	0.027***	0.025***	0.026***	0.027***
	(0.012)	(0.012)	(0.009)	(0.009)	(0.009)	(0.009)
BMI at age 15		-0.006		-0.006		-0.006
		(0.007)		(0.007)		(0.007)
Smoking at age 15		0.283***		0.286***		0.321***
		(0.083)		(0.086)		(0.088)
Drinking at age 15		-0.025		-0.026		-0.028
		(0.041)		(0.042)		(0.042)
Observations	393	393	393	393	393	393
Adjusted R-squared			0.128	0.148	0.126	0.145
Panel B. NEET at 25						
MEGA	0.011	0.009	0.010*	0.009	0.010*	0.010*
	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
BMI at age 15		0.005		0.006		0.006
		(0.005)		(0.004)		(0.004)
Smoking at age 15		0.043		-0.005		-0.006
		(0.052)		(0.054)		(0.054)
Drinking at age 15		-0.036		-0.032		-0.033
		(0.027)		(0.027)		(0.027)
Observations	448	448	448	448	448	448
Adjusted R-squared			-0.004	0.025	-0.004	0.026

Notes: Standard errors in parentheses. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

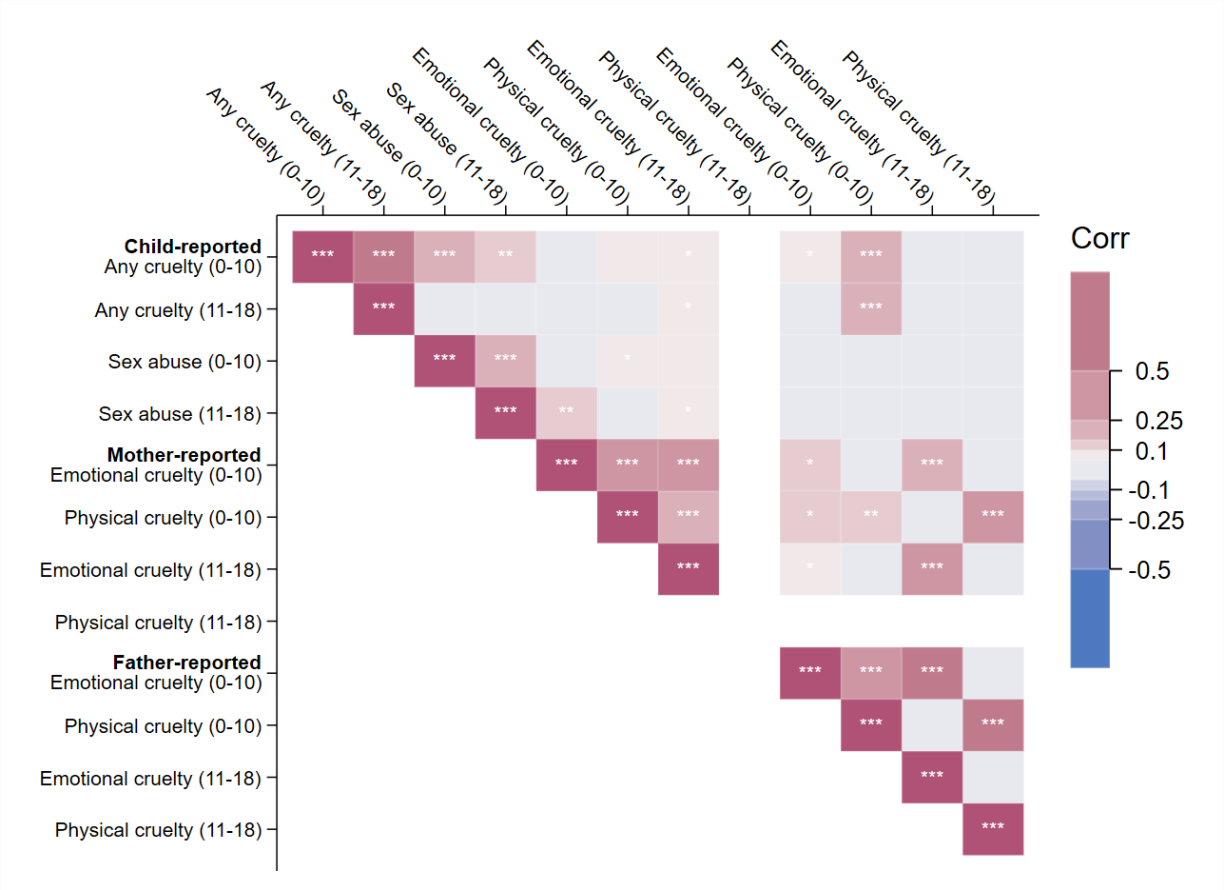
Table 6: Age acceleration and early-adulthood non-cognitive outcomes

	SEM		FA		Weighted Index	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Problematic SMFQ at 25						
MEGA	0.011	0.012	0.011*	0.012*	0.012*	0.012*
	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
BMI at age 15		-0.004		-0.004		-0.004
		(0.005)		(0.005)		(0.005)
Smoking at age 15		0.005		0.014		0.013
		(0.068)		(0.073)		(0.073)
Drinking at age 15		-0.016		-0.016		-0.017
		(0.031)		(0.032)		(0.032)
Observations	338	338	338	338	338	338
Adjusted R-squared			0.014	0.002	0.015	0.003
Panel B. Diagnosed with depression by 22						
MEGA	0.011	0.009	0.017*	0.016*	0.017*	0.017*
	(0.013)	(0.013)	(0.009)	(0.009)	(0.009)	(0.009)
BMI at age 15		0.006		0.006		0.006
		(0.007)		(0.007)		(0.007)
Smoking at age 15		0.062		0.021		0.020
		(0.082)		(0.085)		(0.085)
Drinking at age 15		0.037		0.041		0.040
		(0.042)		(0.042)		(0.042)
Observations	445	445	445	445	445	445
Adjusted R-squared			0.001	0.008	0.001	0.008

Notes: Standard errors in parentheses. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

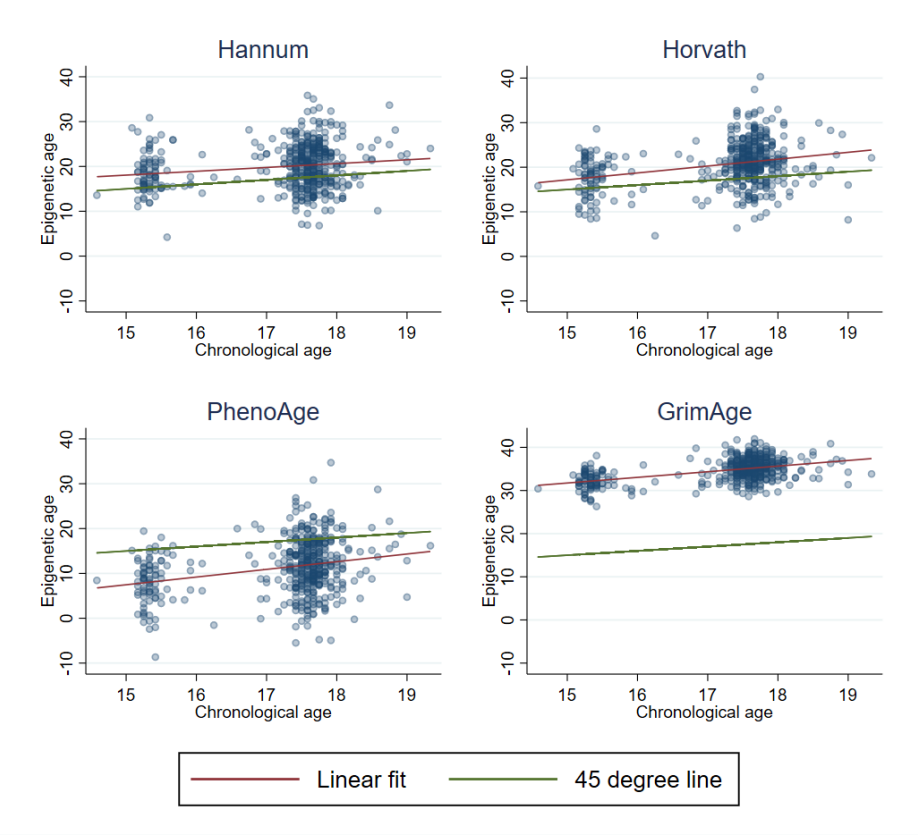
Appendix A: Supplementary Figures and Tables

Figure A1: Cross-rater, intertemporal correlations of child cruelty and sex abuse



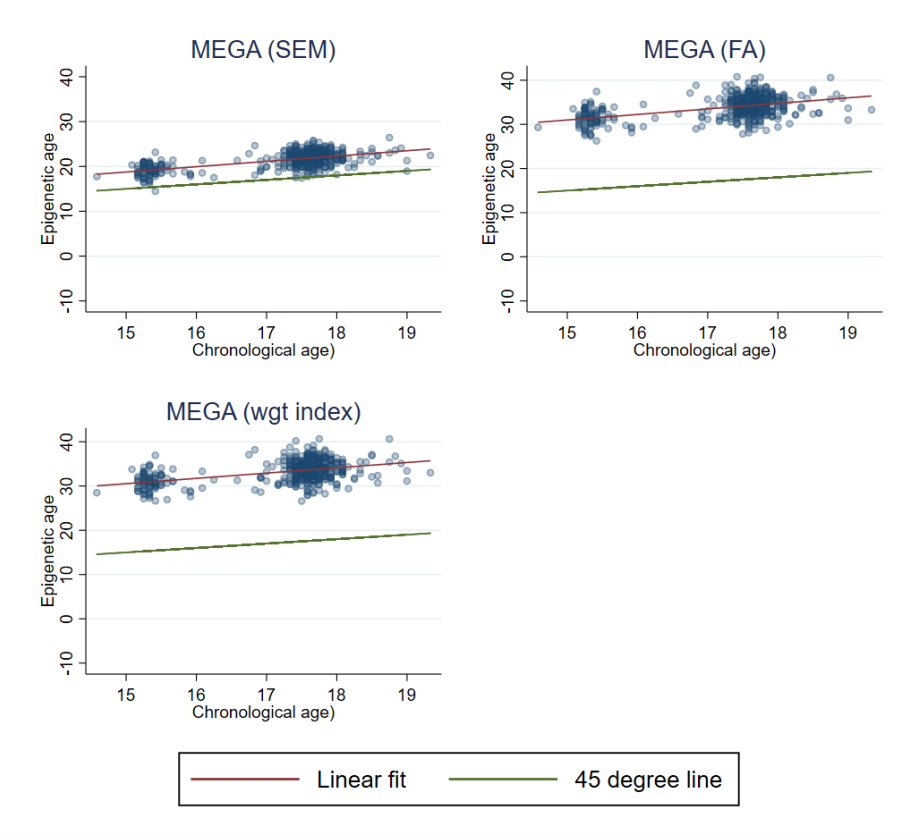
Notes: The graph replicates Figure 1 in the estimation sample of 448 participants. Mother reported physical cruelty is blank as there are no cases of mother-reported physical cruelty between child ages 11 and 18 in the estimation sample.

Figure A2: Chronological age and epigenetic age in four epigenetic clocks



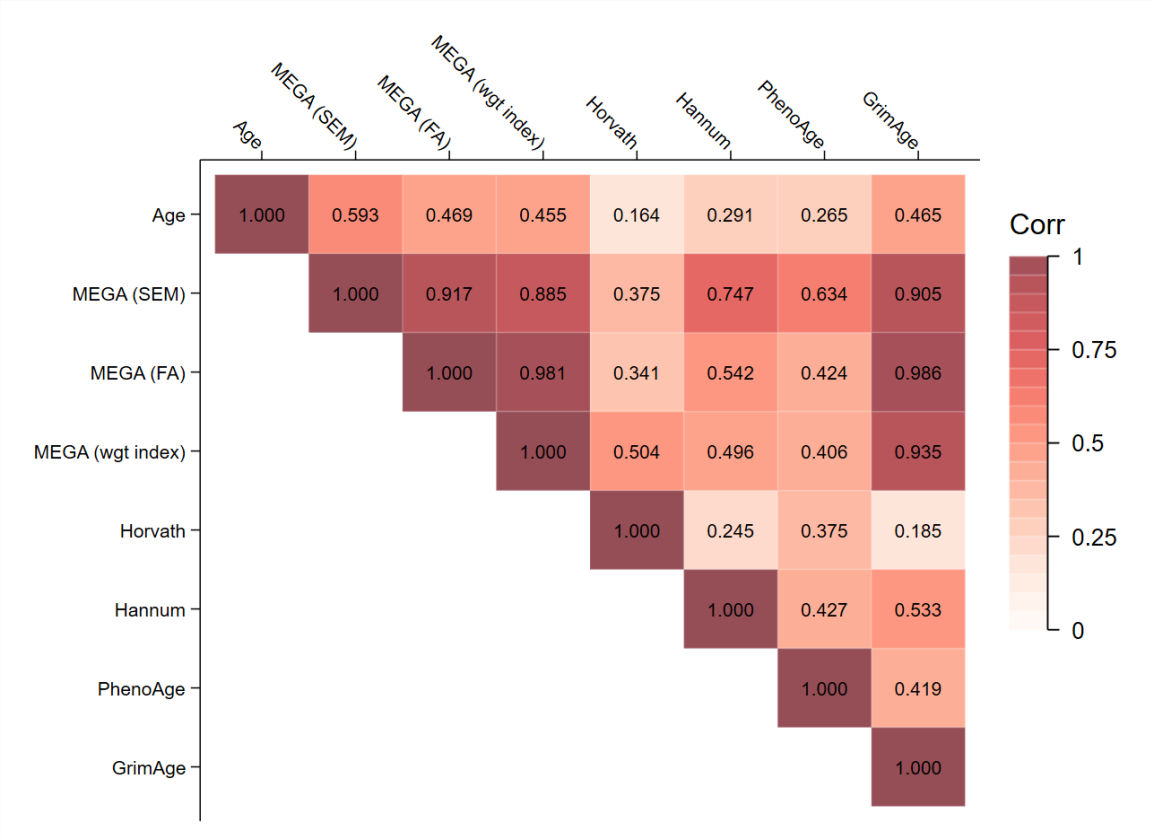
Notes: N=448.

Figure A3: Chronological age and epigenetic age in the MEGA clocks



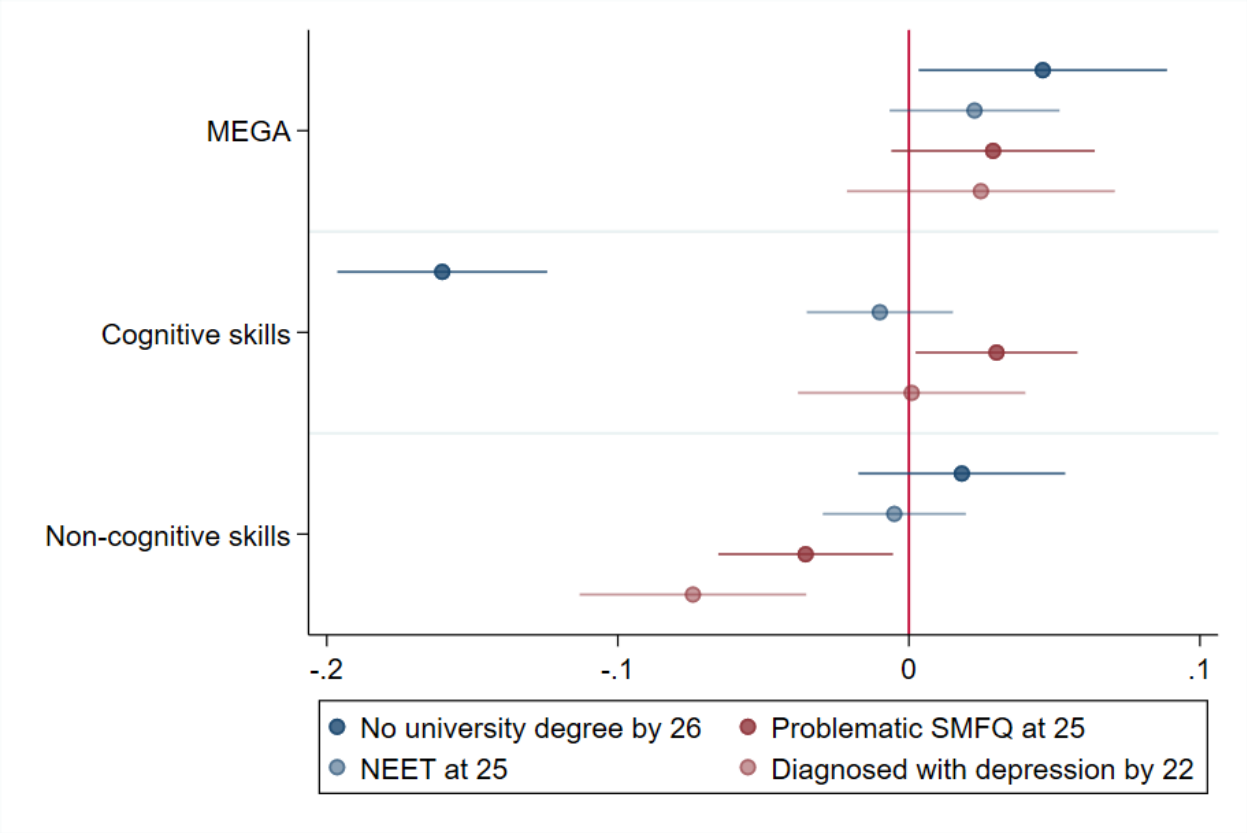
Notes: N=448.

Figure A4: Correlation coefficients across clocks



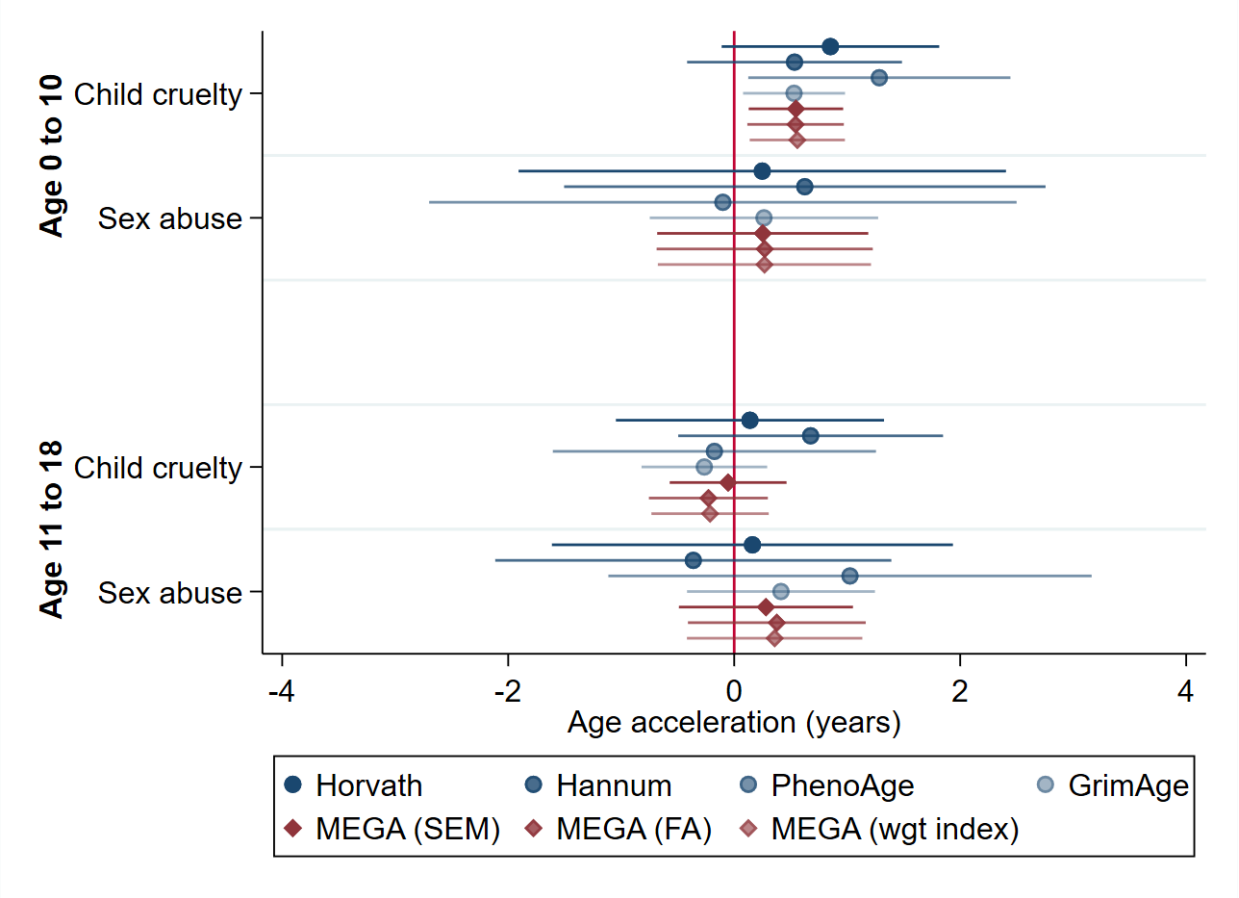
Notes: N=448.

Figure A5: Age acceleration, cognitive and non-cognitive skills, and early-adulthood outcomes



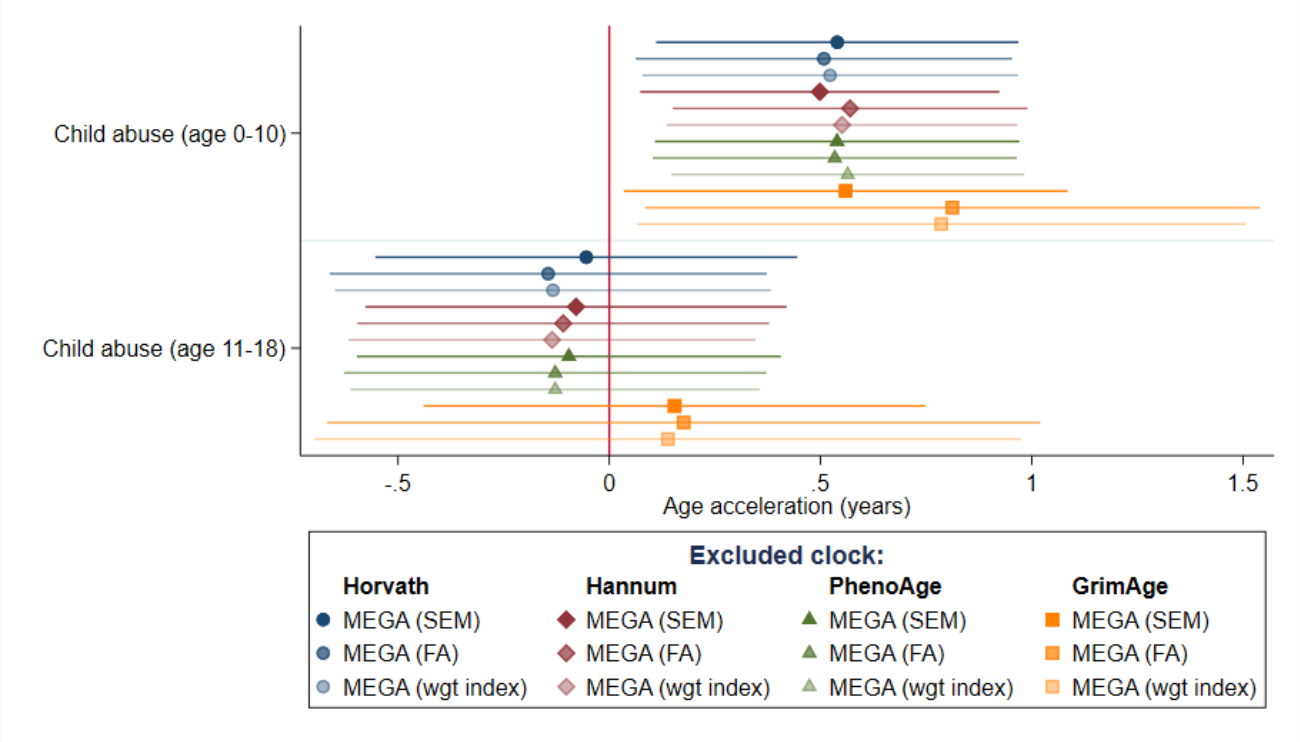
Notes: The figure replicates results the odd-numbered columns from Tables 5 and 6, additionally controlling for a cognitive and a non-cognitive skills factor. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. Spikes are for 90% confidence intervals.

Figure A6: Child abuse and age acceleration from four different clocks: disaggregation



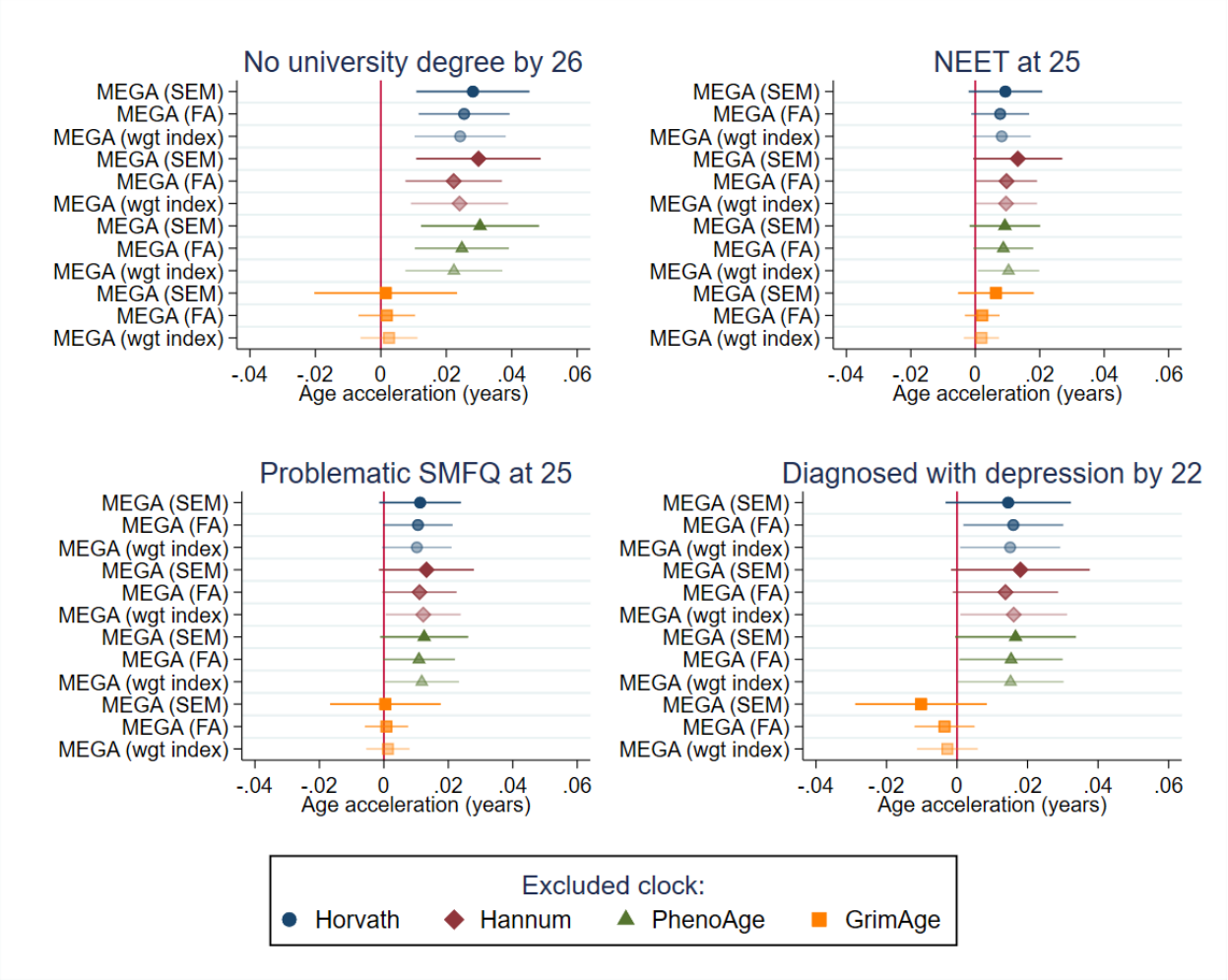
Notes: The figure replicates results from Table 4 using the single clocks as dependent variables instead of the MEGA clock. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. Spikes are for 90% confidence intervals.

Figure A7: Child abuse and age acceleration: leave-one-out MEGA



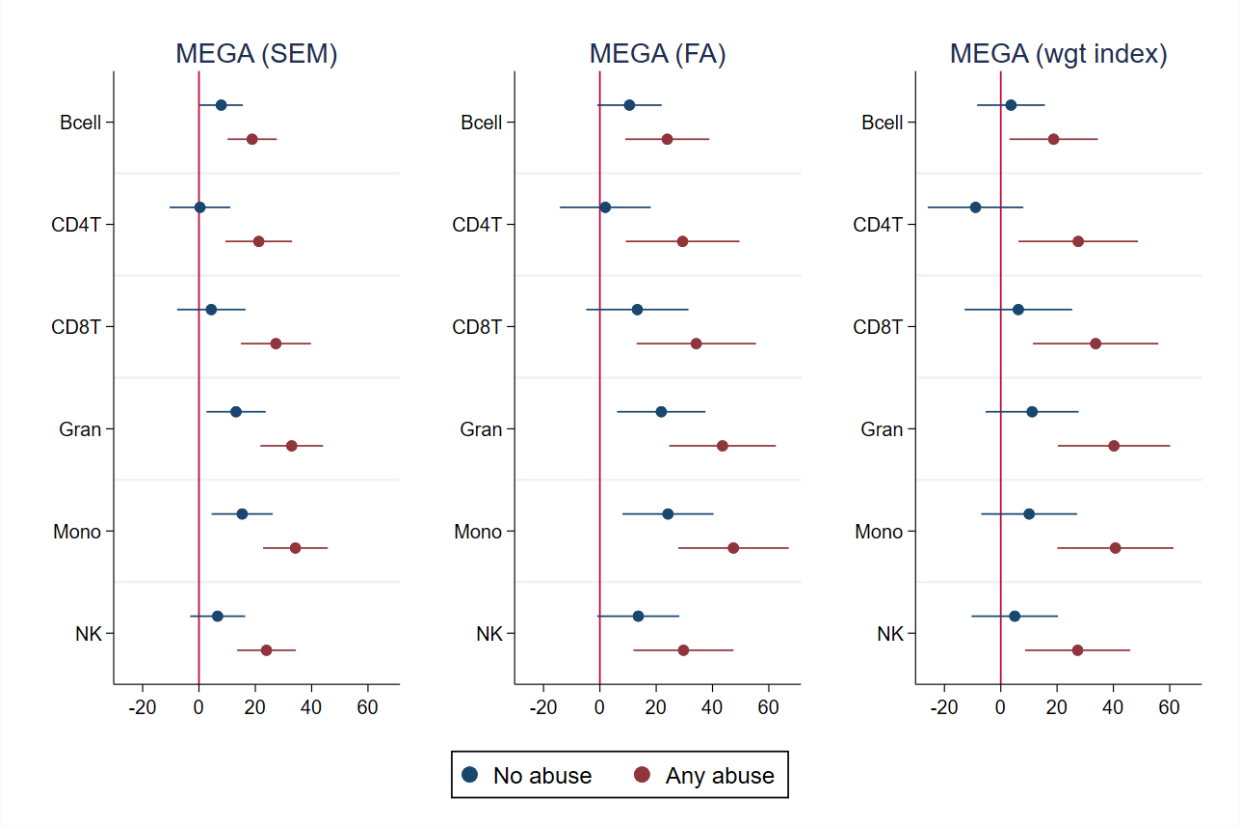
Notes: The figure replicates results from Table 3 using the single clocks as dependent variables instead of the MEGA clock. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. Spikes are for 90% confidence intervals.

Figure A8: Age acceleration and early-adulthood outcomes: leave-one-out MEGA



Notes: The figure replicates results from even-numbered columns in Tables 5 and 6 using leave-one-out versions of the MEGA clock. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. Spikes are for 90% confidence intervals.

Figure A9: Age acceleration and blood cell counts by exposure



Notes: The figure plots associations between blood cell counts and the MEGA clocks in the estimation sample, from a linear regression model where the only control is age. Spikes are for 90% confidence intervals.

Table A1: Self-reported measures of child abuse in ALSPAC (age 22+)

Variable label	Dichotomization
Frequency adult in family pushed, grabbed or shoved respondent	Happened at least 'sometimes'
Frequency adult in family smacked respondent for discipline	Happened at least 'sometimes'
Frequency adult in family punished respondent in a way that seemed cruel	Happened at least 'sometimes'
Frequency adult in family threatened to kick, punch, hit respondent with something that could hurt respondent or physically attack respondent in another way	Happened at least 'sometimes'
Frequency adult in family actually kicked, punched, hit respondent with something that could hurt respondent or physically attacked respondent in another way	Happened at least 'rarely'
Frequency adult in family hit respondent so hard it left bruises or marks	Happened at least 'rarely'
Respondent was touched in a sexual way by adult or older child, or was forced to touch adult or older child in a sexual way	Happened at least once
Adult or older child forced, or attempted to force, respondent into any sexual activity by threatening or holding respondent down or hurting respondent in some way	Happened at least once

Table A2: Selection on observables of the estimation sample

	Full sample	Estimation sample	Difference (1)-(2)
	(1)	(2)	(3)
Female	0.489	0.618	-0.129***
	[0.500]	[0.486]	(0.024)
	14997	*448*	
Age	17.123	17.228	-0.105
	[1.042]	[0.951]	(0.058)
	925	*448*	
Born in 1992	0.563	0.685	-0.122***
	[0.496]	[0.465]	(0.024)
	15468	*448*	
First-born	0.440	0.500	-0.060*
	[0.496]	[0.501]	(0.024)
	13320	*448*	
Mother's age at birth	27.989	29.946	-1.957***
	[4.969]	[4.307]	(0.238)
	14023	*448*	
Mother's education (ref: Lower-secondary)			
Upper-secondary	0.182	0.292	-0.111***
	[0.386]	[0.455]	(0.019)
	15612	*448*	
Post-secondary	0.104	0.277	-0.173***
	[0.305]	[0.448]	(0.015)
	15612	*448*	
Father's social class (ref.: Professionals)			
Non-manual	0.316	0.482	-0.166***
	[0.465]	[0.500]	(0.022)
	15584	*448*	
Manual	0.311	0.259	0.052*
	[0.463]	[0.439]	(0.022)
	15584	*448*	

Notes: The table plots means of covariates and their differences across the estimation sample and the largest ALSPAC sample in which each covariate is available. Standard deviations in brackets and standard errors in parentheses. Sample sizes are indicated in italics below standard deviations. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Factor analysis results for $MEGA_{FA}$

	Factor loadings	Uniqueness
Horvath	0.410	0.832
Hannum	0.662	0.562
PhenoAge	0.633	0.599
GrimAge	0.639	0.592

Table A4: Child abuse and age acceleration from the MEGA clock (standardized results)

	SEM (1)	FA (2)	Weighted index (3)
Any child abuse (0-10)	0.252**	0.221**	0.232**
	(0.124)	(0.106)	(0.107)
Any child abuse (11-18)	-0.003	0.021	0.010
	(0.136)	(0.123)	(0.124)
Observations	448	448	448
Adjusted R-squared	.	0.182	0.171

Notes: Standard errors in parentheses. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Child abuse and age acceleration from the MEGA clock: disaggregation (standardized results)

	SEM (1)	FA (2)	Weighted index (3)
Any child cruelty (0-10)	0.255**	0.228**	0.240**
	(0.126)	(0.108)	(0.109)
Any sex abuse (0-10)	0.118	0.090	0.092
	(0.265)	(0.242)	(0.243)
Any child cruelty (11-18)	-0.025	0.012	0.010
	(0.146)	(0.133)	(0.134)
Any sex abuse (11-18)	0.131	0.097	0.094
	(0.219)	(0.199)	(0.200)
Observations	448	448	448
Adjusted R-squared	.	0.181	0.170

Notes: Standard errors in parentheses. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Child abuse and age acceleration from the MEGA clock: sensitivity to the rater of abuse

	M	P	C	MP	CM	CP	CMP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. SEM							
Any abuse (0-10)	.	0.528	0.488*	0.294	0.489*	0.558**	0.539**
	.	(0.458)	(0.277)	(0.283)	(0.251)	(0.273)	(0.251)
Any abuse (11-18)	.	0.447	-0.010	0.321	0.024	-0.007	-0.007
	.	(0.450)	(0.315)	(0.362)	(0.292)	(0.311)	(0.291)
Observations	448	448	448	448	448	448	448
B. FA							
Any abuse (0-10)	0.408	0.400	0.308	0.427	0.470*	0.382	0.536**
	(0.295)	(0.466)	(0.284)	(0.286)	(0.256)	(0.280)	(0.255)
Any abuse (11-18)	0.367	0.482	-0.035	0.327	-0.084	-0.058	-0.134
	(0.372)	(0.457)	(0.324)	(0.367)	(0.297)	(0.320)	(0.296)
Observations	448	448	448	448	448	448	448
Adjusted R-squared	0.174	0.177	0.181	0.176	0.181	0.186	0.184
C. Weighted index							
Any abuse (0-10)	0.468	0.392	0.315	0.489*	0.480*	0.386	0.549**
	(0.290)	(0.460)	(0.280)	(0.282)	(0.252)	(0.277)	(0.252)
Any abuse (11-18)	0.348	0.443	-0.033	0.299	-0.086	-0.065	-0.143
	(0.367)	(0.451)	(0.319)	(0.362)	(0.293)	(0.316)	(0.292)
Observations	448	448	448	448	448	448	448
Adjusted R-squared	0.263	0.259	0.258	0.264	0.262	0.259	0.264

Notes: Standard errors in parentheses. The dependent variable is the MEGA clock age acceleration, computed with SEM in panel A, with FA in panel B and with the weighted index in panel C. Letters in the column headers indicate the person who reported the measure of child cruelty used in the definition of 'Any abuse': 'M' is for mothers, 'P' is for the mother's partner, and 'C' is for the child. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Child abuse and age acceleration from the MEGA clock (controlling for cell type counts)

	SEM	FA	Weighted index
	(1)	(2)	(3)
Any child abuse (0-10)	.	0.305	0.340*
	.	(0.192)	(0.203)
Any child abuse (11-18)	.	-0.360	-0.355
	.	(0.222)	(0.235)
Observations	.	448	448
Adjusted R-squared	.	0.595	0.525

Notes: Standard errors in parentheses. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.