



Life  
Course  
Centre

## WORKING PAPER SERIES

No. 2025-22

September 2025

# Aggregating epigenetic clocks to study human capital formation

Giorgia Menta

Pietro Biroli

Divya Mehta

Conchita D'Ambrosio

Deborah A. Cobb-Clark AO

The Australian Research Council Centre of Excellence  
for Children and Families over the Life Course

Phone +61 7 3346 7477 Email [lcc@uq.edu.au](mailto:lcc@uq.edu.au)

[lifecoursecentre.org.au](http://lifecoursecentre.org.au)



Australian Government

Australian Research Council



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA



THE UNIVERSITY OF  
SYDNEY



THE UNIVERSITY OF  
WESTERN  
AUSTRALIA



THE UNIVERSITY OF  
MELBOURNE

## Research Summary

### Why was the research done?

This research was conducted to explore how children's environments shape their biological development and future life chances. While genes are fixed, the way they are expressed can change depending on experiences such as family background, schooling, or adversity. Epigenetic "clocks" have become the main tool for measuring biological aging, but with multiple versions available, researchers lacked a clear, reliable measure. This paper develops a new combined measure – the Multi EpiGenetic Age (MEGA) clock – and applies it to study how early-life experiences affect life chances.

### What were the key findings?

The paper shows that: 1) the MEGA clock provides a more accurate and consistent measure of biological age than existing individual clocks; 2) teenagers with faster biological aging have lower educational attainment, worse mental health, and worse labour market outcomes in young adulthood; 3) exposure to child maltreatment before age 10 is linked to faster biological aging in adolescence, equivalent to about half a year; and 4) children who start school one year later show accelerated biological aging by age seven, especially those from disadvantaged families.

### What does this mean for policy and practice?

Policies aimed at improving children's environments can have lasting biological as well as social and economic benefits. Protecting children from maltreatment and ensuring early access to structured educational settings may help reduce inequalities in health and human capital.

## Citation

Menta, G., Biroli, P., Mehta, D., D'Ambrosio, C., & Cobb-Clark, D.A. (2025). 'Aggregating epigenetic clocks to study human capital formation', Life Course Centre Working Paper Series, 2025-22. Institute for Social Science Research, The University of Queensland.

## The authors

### **Giorgia Menta**

Luxembourg Institute of Socio-Economic Research

Email: [giorgia.menta@liser.lu](mailto:giorgia.menta@liser.lu)

<https://liser.elsevierpure.com/en/persons/giorgia-menta>

Twitter: @MentaGiorgia/@LISERinLux

### **Pietro Biroli**

University of Bologna

Email: [pietro.biroli@unibo.it](mailto:pietro.biroli@unibo.it)

<https://sites.google.com/site/pietrobiroli/>

### **Divya Mehta**

Queensland University of Technology

Email: [divya.mehta@qut.edu.au](mailto:divya.mehta@qut.edu.au)

### **Conchita D'Ambrosio**

University of Luxembourg

Email: [conchita.dambrosio@uni.lu](mailto:conchita.dambrosio@uni.lu)

### **Deborah A. Cobb-Clark AO**

School of Economics, University of Sydney

Email: [deborah.cobb-clark@sydney.edu.au](mailto:deborah.cobb-clark@sydney.edu.au)

## Acknowledgements/Funding Sources

We would like to thank Nancy Kong and Anita Sathyanarayanan for their input in establishing the conceptual underpinnings of the project. Tiger Mathieson and Sura Majeed provided excellent research assistance. We thank seminar and conference participants at the University of Bristol, the University of Essen, NTU Singapore, Chulalongkorn University, the 16th Annual IGSS Conference, the 38th ESPE Annual Conference, and the 2024 AASLE Conference for useful comments and suggestions. We are 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 U.K. Medical Research Council and Wellcome (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 here: <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), from the Fonds National de la Recherche Luxembourg (Grant C19/SC/13650569/ALAC), and from the European Union Research Council, under project #101073237 – ESSGN.

DISCLAIMER: The content of this Working Paper does not necessarily reflect the views and opinions of the Life Course Centre. Responsibility for any information and views expressed in this Working Paper lies entirely with the author(s).

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).



We acknowledge the Traditional Custodians of the lands on which we work and live across Australia.  
We pay our respects to Elders past and present and recognise their continued connections  
to land, sea and community.

## **Abstract**

Epigenetics is the study of how people's behavior and environments influence the way their genes are expressed, even though their DNA sequence is itself unchanged. By aggregating age-related epigenetic markers, epigenetic 'clocks' have become the leading tool for studying biological aging. We make an important contribution by developing a novel, integrated measure of epigenetic aging – the Multi EpiGenetic Age (MEGA) clock – which combines several existing epigenetic clocks to reduce measurement error and improve estimation efficiency. We use the MEGA clock in three empirical contexts to show that: i) accelerated epigenetic aging in adolescence is associated with worse educational, mental-health, and labor market outcomes in early adulthood; ii) exposure to child maltreatment before adolescence is associated with half a year higher epigenetic aging; and iii) that entering school one year later accelerates epigenetic aging by age seven, particularly among disadvantaged children. The MEGA clock is robust to alternative methods for constructing it, providing a flexible and interpretable approach for incorporating epigenetic data into a wide variety of settings.

**JEL Codes:** I12, I14, J24

**Keywords:** epigenetic clocks, DNA methylation, child abuse, human capital, ALSPAC data

## 1. Introduction

Nature and nurture form a two-way street: human biology and environmental conditions influence people's life chances through a dynamic interplay. Our genetic code affects the environments we grow up in and the trajectories we follow (Plomin and Bergeman, 1991; Cole 2009; Belsky and Harden, 2019). External shocks and policy interventions, in turn, can alter biological processes in ways that persist across the life course (Hertzman, 1999; Shonkoff *et al.*, 2009). In economics, early 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) has contributed to recent advances in behavioral genetics (e.g. Okbay *et al.*, 2022; Abdellaoui *et al.*, 2023; Benjamin *et al.*, 2007, 2024) and the integration of genes into economic models (e.g. Papageorge and Thom, 2020; Houmark *et al.*, 2024; Sanz-de-Galdeano and Terskaya, 2025). While genetic variants are fixed at conception, genetic expression is not. Epigenetics – the study of changes in gene expression that occur without alterations to the DNA sequence – has opened a new frontier for understanding the process through which people's environments affect their biological functioning. Epigenetic modifications can happen through exposure to adverse environments (Cole, 2009); leave a lasting biological footprint (Szyf, 2009); and shape economic choices (Almond and Currie, 2011). Social scientists are increasing their efforts to understand the way socio-economic conditions shape the epigenome; yet, with a few recent exceptions (Schmitz and Duque, 2022; Costi *et al.* 2025a, 2025b), the economics literature on epigenetics remains relatively limited. This is unfortunate because measuring, understanding, and causally estimating these epigenetic influences can provide insights into the pathways through which economic policies and social environments affect public health and productivity.

We contribute to this emerging field by introducing a novel measure of epigenetic aging and showing how it can be systematically incorporated into economic models to better understand the dynamics of health and human capital formation. Epigenetic aging refers to the pace at which the body ages biologically and is based on changes in gene expression rather than the passage of time. This process is typically captured using epigenetic 'clocks', which employ machine learning algorithms to aggregate DNA methylation data into empirically quantifiable measures of



biological age.<sup>1</sup> Biological age has been shown to be a more reliable predictor of physiological functioning, disease risk, health outcomes, and mortality than chronological age (Kotschy *et al.*, 2025). By summarizing complex epigenetic information, epigenetic clocks offer a useful and quantifiable measure of biological aging that is valuable in both biological and social-science research. The growing number of available clocks, however, can make interpretation challenging (Bell *et al.*, 2019; Teschendorff and Horvath, 2025). Intended to capture the same latent concept – epigenetic aging – the algorithms used to define different epigenetic clocks nevertheless diverge in terms of methodology and selection of relevant epigenetic sites, leading to conceptual differences across clocks as well as measurement error. Applied researchers are often left without clear guidance on which epigenetic measure to use. Moreover, while technological advances have made it less costly to include epigenetic markers in social-science data, sample sizes often remain small, limiting statistical power.

We make four key contributions, one methodological and three empirical. The first is to propose a methodological approach for combining existing epigenetic clocks into a single metric – the Multi EpiGenetic Age (MEGA) clock, which reduces measurement error and provides a more efficient and robust measure of latent epigenetic age. Our strategy integrates several well-established clocks (our empirical application uses four: Horvath, 2013; Hannum *et al.*, 2013; Levine *et al.*, 2018; Lu *et al.*, 2019), each trained to predict different aspects of aging and health, into a unified measure. We enhance the robustness of this approach by leveraging three alternative empirical methods for combining clocks: (i) exploratory factor analysis; (ii) weighted indexing; and (iii) structural equation modeling. Using epigenetic and behavioral data from a large birth cohort in the U.K., the Avon Longitudinal Study of Parents and Children (ALSPAC), we demonstrate that the MEGA clock reduces the measurement error inherent in individual epigenetic clocks, providing more precise estimates of the determinants and consequences of epigenetic aging. The MEGA clock also retains the intuitive age-scale interpretation of existing epigenetic clocks, making it directly comparable to chronological age. Given the growing availability of pre-computed epigenetic clocks in many large-scale data sets (e.g. U.K. Understanding Society, U.S.

---

<sup>1</sup> DNA methylation is a chemical process in which a small molecule, called a methyl group, is added to a DNA molecule. This addition can affect how genes are expressed, effectively turning them “on” or “off” which has implications for disease risk and long-term biological outcomes. DNA methylation is the most studied form of epigenetic modifications.

Health and Retirement Study, U.S. National Longitudinal Study of Adolescent to Adult Health, U.S. The Future of Families and Child Wellbeing Study), the MEGA clock is easy to implement without requiring in-depth biological expertise. Importantly, our methodological approach is designed to accommodate future developments – as new epigenetic clocks are developed, they can be readily incorporated, provided they capture the same latent construct of epigenetic aging.

Our three empirical contributions leverage the ALSPAC data to illustrate how the MEGA clock can be incorporated into both observational and causal analyses of human capital formation across the life cycle. In our first application, we assess the usefulness of the MEGA clock by investigating the association between age acceleration and human capital later in life. Faster epigenetic aging has been shown to be predictive of numerous outcomes, including mortality (e.g. Faul *et al.*, 2023; Lu *et al.*, 2019; Perna *et al.*, 2016) and disease incidence (e.g. Horvath, 2013; McCartney *et al.*, 2018). 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 socio-emotional skills, as suggested by studies looking at epigenetic aging and cognitive performance in children and young adults (e.g. Niccodemi *et al.*, 2022; Costi *et al.*, 2025a). We conceptualize the MEGA clock as a potential input in the human capital production function, hypothesizing that adolescents who experience accelerated epigenetic aging will exhibit worse cognitive and socio-emotional outcomes in early adulthood, even after controlling for family background, cognitive and socio-emotional skills, and health. Our results confirm that the MEGA clock is an efficient measure of biological aging and suggest that epigenetic age acceleration in adolescence predicts fewer years of education, worse mental health, and weaker attachment to the labor market, above and beyond traditional health biomarkers (e.g. BMI) and health behaviors (e.g. smoking and drinking).

We next turn to consideration of the shocks that might influence epigenetic aging. We begin by examining the association of exposure to child abuse on epigenetic age acceleration in late adolescence.<sup>2</sup> Child maltreatment is of particular interest because it is an external, child-specific experience with well-documented adverse consequences (Currie and Spatz Widom, 2010; Fletcher, 2009; Henkaus, 2022; Suglia *et al.*, 2014) and has been linked to widespread

---

<sup>2</sup> Age acceleration occurs when an individual’s epigenetic age (here, the MEGA clock) outpaces their chronological age, signaling a faster rate of biological aging relative to the passage of time.



modifications across the epigenome (Lawn *et al.*, 2018). We contribute to this literature by shifting the focus from general epigenetic changes to epigenetic aging, showing how experiences of child abuse are associated with accelerated epigenetic aging when we adopt a longitudinal approach that accounts for different abuse trajectories and critical developmental periods (Heckman, 2007; Cunha *et al.*, 2010). Our results show that exposure to child maltreatment before adolescence is associated with half a year of accelerated epigenetic aging, a result comparable in magnitude to the age acceleration observed for children born into socio-economic disadvantage (see Fiorito *et al.*, 2017).

In our final empirical study, we provide causal evidence of the biological consequences of early school exposure by examining how the timing of school entry influences epigenetic aging. Using a sharp regression discontinuity design (RDD) around the U.K. school-entry age cutoff, we identify the causal impact of entering school one year later on epigenetic age acceleration at age seven. This approach, widely used in economics to estimate the effects of school starting age on cognitive and labor market outcomes (e.g. Bedard and Dhuey, 2006; Black, Devereux, and Salvanes, 2011; Fredriksson and Öckert, 2014), allows us to be the first to isolate the biological consequences of delayed school entry. We find that children born just after the cutoff – who start school one year later – exhibit faster epigenetic aging (0.5 – 1.2 years) than those who start earlier, with especially pronounced effects among children from lower socio-economic backgrounds. These findings suggest that early educational environments may play a protective role in shaping biological development, potentially by promoting healthier routines and buffering stress exposure (Anderson *et al.*, 2011; Holford and Rabe, 2022). By embedding an RDD within an epigenetic framework, we show that institutional factors can causally affect epigenetic aging trajectories as early as mid-childhood.

Our research speaks to several literatures. First, we advance the emerging inter-disciplinary literature that uses epigenetic clocks by proposing a methodological innovation that addresses key challenges faced by researchers. Our approach provides a simple and scalable method to reduce reliance on arbitrary clock selection, improve the replicability of results across different clocks, and increase statistical power without expanding sample size. To our knowledge, this is the first study to build an integrated epigenetic clock based on other existing clocks. Recently, Martinez *et al.* (2025) have proposed a complementary, yet distinct, approach. They use a battery of

biomarkers, including epigenetic clocks, and confirmatory factor analysis, to identify three factors they interpret as epigenetic age, systemic biological age, and immune age.

Second, we are the first to estimate the causal effect of age at school-entry on health and biological outcomes, and among the first to estimate the causal effect of quasi-exogenous shocks or policies on epigenetic aging (see Schmitz and Duque, 2022 for an exception).<sup>3</sup> We also contribute to the rich observational literature linking abuse with epigenetic aging, incorporating measures taken across different developmental periods and perspectives (i.e., from mothers, fathers, and children), and adding a temporal dimension to the analysis (see Cecil *et al.*, 2020, and Rubens *et al.*, 2023, for systematic reviews). Using data from ALSPAC mothers, Lawn *et al.* (2018) find that sexual abuse experienced by the age of 17 is associated with an epigenetic age that is 3.4 year higher in adulthood (Horvath clock), after adjusting for childhood and adulthood socio-economic positioning. Marini *et al.* (2020) look at early childhood DNA methylation 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 seven. The effect is driven by girls and is only found when using the Hannum rather than the Horvath's clock. Lussier *et al.* (2023) also adopt a longitudinal approach, examining 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.

Third, our research supports the growing evidence that epigenetic age is a valid and policy-relevant measure of health that can serve as a feasible short-term endpoint for evaluating the impact of interventions aimed at improving well-being. Age acceleration has been shown to be predictive of all-cause mortality (Faul *et al.*, 2023; Lu *et al.*, 2019; Perna *et al.*, 2016), healthcare utilization (Davillas and Jones, 2024), and the occurrence of diseases such as cancer (Horvath, 2013; Dugué *et al.*, 2018; Perna *et al.*, 2016) and Alzheimer's disease (McCartney *et al.*, 2018). In children and young adults, age acceleration has been shown to be negatively associated with test-scores (Niccodemi *et al.*, 2022), cognitive functioning (Raffington *et al.*, 2023a; Costi *et al.*, 2025a), and

---

<sup>3</sup> A small clinical literature, typically based on samples of a few dozen participants, has examined the short-term biological-aging effects of randomized interventions aimed at improving diet and physical activity (e.g. Chen *et al.*, 2019; Fitzgerald *et al.*, 2021; Nwanaji *et al.*, 2021).

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 in the domains of health and labor market outcomes.

Finally, our life-course approach maps out experiences both in childhood and adolescence, providing a comprehensive view of the way that early-life conditions influence epigenetic 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.

## **2. The Conceptual Underpinnings of Epigenetic Clocks**

All humans age, but the rate at which we do so differs considerably from person to person. Chronological age measures how long someone has lived, while biological age captures the state of people's physiological and cellular systems based on various physiological and molecular markers. Biological aging is associated with progressive loss of function at the cellular, tissue, and organ levels, further accelerating the general decline in physical functioning and cognitive performance (López-Otín *et al.*, 2013). We are interested in a person's biological age because it is a more accurate measure of their functional capacity than is their chronological age.

But how can one measure biological aging? One well-studied process underlying biological aging is epigenetics, that is, the ensemble of reversible chemical and structural alterations to the genome that can lead to long-term changes in gene activity even though the underlying DNA sequence itself has not been altered (Klengel *et al.*, 2014). The most studied form of epigenetics is DNA methylation.<sup>4</sup> DNA methylation happens when a methyl group binds to a CpG site – a particular section of the DNA sequence where a cytosine nucleotide (C) is adjacent to a guanine nucleotide (G), with one phosphate (p) in between – playing a critical role in regulating gene expression by essentially turning genes “on” or “off”. DNA methylation patterns are relatively stable over time

---

<sup>4</sup> While other forms of epigenetic changes exist (e.g. histone modification), current biotechnologies have become increasingly cost-efficient in the stabilization and extraction of DNA methylation from human tissues. As a result, there is a wealth of data and well-characterized tools available for studying DNA methylation.

and can be measured reliably at scale in large cohort studies, allowing researchers to link biological regulation to health and human capital outcomes across the life course.

Research has shown that, as people age, certain parts of their DNA become systematically more or less methylated. This insight laid the foundation for the development of epigenetic biomarkers of aging – referred to as epigenetic clocks – that are designed to predict either chronological aging or physiological decline using DNA methylation levels. Epigenetic clocks are constructed using supervised machine-learning algorithms (e.g. penalized elastic net regression) to address the high-dimensionality of the data and prevent overfitting, since epigenetic datasets typically contain substantially fewer individuals than measured CpG sites. The algorithms usually consist of two phases: development and prediction. In the development phase, the parameters of the algorithm are adjusted to achieve the greatest prediction accuracy in large training samples. These estimated parameters are then used in prediction (estimation) samples to construct epigenetic clocks as weighted scores of DNA methylation at selected CpG sites, using the weights learned from the training sample used in model development. Adopting the same general approach, existing epigenetic clocks differ along four main dimensions: (i) which CpG sites are included in the training sample and the resulting weights identified by the algorithms; (ii) the type of tissue used to measure DNA methylation (e.g. blood, saliva); (iii) the outcome they are meant to predict (e.g., chronological age, physical fitness, mortality) and (iv) the sample that they are trained on.

## **2.1. First-Generation Clocks**

The first two widely used epigenetic clocks, Horvath's DNA methylation age clock and Hannum's clock, were developed in 2013 to predict chronological age using DNA methylation data. The Horvath clock captures age-related DNA methylation patterns, using the combined methylation status of 353 CpG sites. It is a multi-organ clock that predicts chronological age from embryo to old age using DNA methylation data on many tissues and organs (e.g. whole blood, cerebellum, colon, kidney, liver, lungs). The Hannum clock similarly uses elastic net regression to predict people's chronological age, identifying 71 CpG sites that can be used to accurately predict age (Hannum *et al.*, 2013). Epigenetic age as estimated by Hannum's clock has been shown to be a more accurate predictor of chronological age than Horvath's clock when adult blood samples are used, since this is the tissue on which the model was trained. Estimates of Hannum epigenetic age

have been shown to be biased, however, when applied in non-blood tissues (Simpkin *et al.*, 2016) and in children (Hannum *et al.*, 2013).

Because the algorithms underlying these clocks were trained on chronological age, critics have argued that these “first-generation” clocks may, in fact, exclude CpGs whose methylation patterns reflect biological age variation. As a result of being designed to detect age-dependent patterns, these algorithms can achieve remarkable chronological age prediction. Clocks that achieve almost perfect chronological age prediction, however, have been shown to be worse predictors of mortality (Zhang *et al.*, 2019).

## **2.2. Second-Generation Clocks**

In recent years, a second generation of DNA methylation-based biomarkers for aging has emerged in which CpGs associated with organ-system functioning are captured alongside chronological age. In 2018, Levine developed the PhenoAge clock, trained to predict not only chronological age but other aging-related indicators (e.g. blood glucose, liver and kidney markers of function).<sup>5</sup> This clock, while also a good predictor of chronological age, has been extensively validated; it can predict age-related health outcomes more effectively than the first-generation clocks and can differentiate morbidity and mortality risks among individuals of identical chronological age (Levine *et al.*, 2018). Subsequently, Horvath developed GrimAge which surpassed previous clocks in predicting both age-related disease as well as mortality. GrimAge is a linear combination of DNA methylation-based surrogate biomarkers for health-related plasma proteins, smoking pack-years, sex, and chronological age. It has been shown to be a stronger predictor of lifespan, age-related conditions, disease, and mortality risk compared to the widely used Horvath’s clock (Lu *et al.*, 2019).

The number of CpGs selected by the algorithms underlying epigenetic clocks vary substantially across different models. The Horvath clock incorporates 353 age-related CpGs, while the Hannum clock uses 71 CpGs, PhenoAge employs 513 CpGs, and GrimAge includes 1,030 CpGs. Beyond these numerical differences, the specific CpG sites selected for each model also show limited

---

<sup>5</sup> 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 DNA methylation PhenoAge has achieved greater applicability across chip platforms than other clocks.

overlap. The PhenoAge clock shares only 41 CpGs with the Horvath clock, while both PhenoAge and Horvath clocks have merely 5 CpGs in common with the Hannum clock (Levine *et al.*, 2018).<sup>6</sup>

### **2.3. From Epigenetic Age to Age Acceleration**

Epigenetic age, as estimated by DNA methylation-based clocks, provides a biological measure of aging that often differs from chronological age. The difference between a person’s predicted epigenetic age and their actual chronological age is commonly referred to as age acceleration, typically measured using the residuals from a regression of epigenetic age on chronological age. Understanding age acceleration is important because it provides insights into individual-specific variation in the aging process that cannot be gained by simply counting years lived. People of the same chronological age often differ substantially in their biological condition due to genetics, lifestyle, and environmental factors. Age acceleration highlights who is aging faster or slower than average, providing a personalized metric (biomarker) that better predicts health risks, disease progression, and life outcomes.

In all empirical applications, we control for chronological age, allowing estimates of the determinants and consequences of epigenetic age to be interpreted directly as age acceleration (see Section 5).

## **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. The substantial differences in training methodologies, CpG site selection, and construction approaches create challenges for researchers seeking to determine the most appropriate epigenetic aging measure for their specific analysis. Drawing from the social-sciences literature on the measurement of latent factors, we propose a novel way of harnessing the information coming from different epigenetic clocks into a unique measure of epigenetic aging which we name Multi EpiGenetic Age (MEGA) clock. Although our approach can be implemented

---

<sup>6</sup> Models that include a large number of CpG sites tend to be more robust and accurate than those with fewer (Liu *et al.*, 2020; Lu *et al.*, 2019). However, epigenetic clocks aim to capture broad features of the methylome, and Horvath and Raj (2018) show that including a moderate number of age-associated CpGs is sufficient to yield highly reliable models.

with several different clocks that capture the same underlying construct, our current focus is on four clocks: two first-generation clocks (the Horvath and the Hannum clocks), which have been trained to predict chronological age; and two second-generation clocks (the PhenoAge and GrimAge clocks), which have been trained to predict lifespan (functional stage). Our rationale for this choice is twofold. On the one hand, the robustness and replicability of these clocks have been widely shown across a variety of samples and tissues (Lu *et al.*, 2019; Maddock *et al.*, 2020; McCrory *et al.*, 2021). On the other hand, these clocks all rely on genome-wide DNA methylation data derived from the same methylation profiling technology (i.e. Illumina Infinium arrays)<sup>7</sup> and are constructed 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 ill health, it is reasonable to think of these clocks as separate indicators of the same latent concept. Combining the information contained in these four clocks would then be expected to reduce measurement error, leading to a more robust, holistic measure of DNA methylation age.

We construct three versions of the MEGA clock, each based on a different methodological approach with varying underlying identification assumptions. In the empirical applications we will rely on the four clocks defined above (Horvath, Hannum, PhenoAge, GrimAge), however, it is important to note that the procedures described below are generalizable to any set of epigenetic clocks. The first MEGA clock,  $MEGA_{WGT}$ , relies on a weighted index approach proposed by Anderson (2008). Here, clocks are combined using a weighted sum with the weights equal to the row-sums of the inverse variance-covariance matrix of the clocks. Specifically, let  $\{C_1, \dots, C_K\}$  be a set of epigenetic clocks, all expressed in the same unit (years of age), and  $M$  be their variance-covariance matrix. Further,  $I_K$  is a vector of ones with 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), using the  $MEGA_{WGT}$  clock constructed with this weighting procedure in a regression is analogous to the joint

---

<sup>7</sup> Illumina Infinium arrays are a profiling technology that measures DNA methylation. For example, Illumina Infinium 450k arrays measure DNA methylation in up to 450k CpG sites.



estimation of seemingly unrelated regression, one for each individual clock, constraining all clock-related coefficients to be equal, and corresponds to an efficient Generalized Least Squares (GLS) estimator.

The second MEGA clock,  $MEGA_{FA}$ , relies instead on exploratory factor analysis, which is commonly used to identify one or more latent constructs using a series of measures. 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 factor analysis. We interpret this factor as the  $MEGA_{FA}$ . 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:

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

where  $u_{k,1}$  is the  $k$ -th element of vector  $u = M^{-1}L_K$ . Both  $MEGA_{WGT}$  and  $MEGA_{FA}$  rely on the clocks' variance-covariance matrix  $M$ , but they do so in different ways.  $MEGA_{WGT}$  downweights redundant clocks and gives more weight to those adding independent information, whereas  $MEGA_{FA}$  additionally emphasizes the common variance across clocks, as captured by their factor loadings, thereby extracting the shared aging signal.

Lastly,  $MEGA_{SEM}$ , uses Structural Equation Modelling and relies on the strongest identification assumptions. The advantage of Structural Equation Modelling is that it allows us to simultaneously model both observed and latent variables and their direct relationships, as well as error terms. Informed by the factor analysis, the  $MEGA_{SEM}$  clock is estimated as the latent factor of a system of measurement equations based on the individual clocks and a behavioral (or structural) equation, simultaneously modelling its relationship with other shocks or outcomes – such as child abuse, age at school-entry, or labor market outcomes. The system of measurement equations can be illustrated as follows:

$$C_k = \lambda_k EA + \varepsilon_k, k = 1, \dots, K \quad (3)$$

where  $C_k$ , the  $k$ -th epigenetic clock, can be seen as an indicator of latent epigenetic age, denoted  $EA$ .  $\lambda_k$  is the factor loading parameter and  $\varepsilon_k$  is a 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:

$$EA = X'\gamma + \omega \quad (4)$$

where  $\gamma$  is a  $P \times 1$  vector of parameters and  $\omega$  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:

$$C_k = \lambda_k X'\gamma + \lambda_k \omega + \varepsilon_k, k = 1, \dots, K \quad (5)$$

Conditional on the behavioral equation being correctly specified, Structural Equation Modelling is generally more efficient because it uses all data to simultaneously estimate parameters. This simultaneous estimation minimizes potential information loss and takes into account the full covariance structure of the data.<sup>8</sup>

The estimation of the MEGA clock using 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 system of measurement equations.<sup>9</sup> 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 measurement 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.

---

<sup>8</sup> A similar approach can be followed when the latent Epigenetic Age  $EA$  variable is used as a control, not an outcome, in the behavioral equations. Consider the following vector of education, health, and labor-market outcomes  $Y = \{Y_1, \dots, Y_D\}$ ; the resulting behavioral equation would be  $Y = EA'\delta + H'\theta + u$ , where  $H'$  is a vector of controls and  $\delta, \theta$  are vectors of parameters to be estimated.

<sup>9</sup> Here, we use Stata v. 17, and commands “sem” for  $MEGA_{SEM}$ , “factor” for  $MEGA_{FA}$  and “egen weightave” for  $MEGA_{WGT}$ .

Combining information from different clocks has the advantage of reducing the measurement error associated with each individual clock, allowing us to capture the process of epigenetic aging more precisely. Minimizing measurement error and improving estimation precision is important given the limited sample sizes of most epigenetic datasets. Reducing measurement error also has the potential to mitigate the risk of attenuation bias in our estimates.

## **4. The Avon Longitudinal Study of Parents and Children**

### **4.1. Data Overview**

Our empirical analyses utilize data from the Avon Longitudinal Study of Parents and Children (ALSPAC), also known as “Children of the 90s”, an English birth cohort study that provides a unique opportunity 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, U.K., 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.<sup>10</sup> This comprehensive approach captures a wide array of information, including physical and mental health, cognitive abilities, socio-economic factors, and environmental exposures (Boyd *et al.*, 2013; Fraser *et al.*, 2013; Northstone *et al.*, 2019, 2023). The biological data collected includes genetic and epigenetic information, with blood samples taken at various time points. Longitudinal DNA methylation data for 1,022 children and their mothers have been collected using Illumina Infinium 450k arrays, the same profiling technology

---

<sup>10</sup> At age 18, study children were sent ‘fair processing’ materials describing ALSPAC’s intended use of their health and administrative records and were given clear means to consent or object via a written form. Data were not extracted for participants who objected, or who were not sent fair processing materials.

used in each of the clocks in the MEGA clock. For children, DNA methylation was extracted from cord blood samples at birth and from peripheral blood samples at child ages 7 and 15-19 years.<sup>11</sup>

Using these data, we then compute epigenetic age 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).<sup>12</sup> We first combine these clocks into a single measure of epigenetic age using three alternative methods (described in Section 3), and then use them separately in our three empirical applications to test the validity of our approach. Our analyses primarily rely on DNA methylation measured in late adolescence (median age 17.5), when children in our sample are biologically closer to the adult populations on which these clocks were originally developed.

## 4.2. Estimation Samples

After conditioning on the availability of DNA methylation data and the main variables of interest (see Section 5 for more details), our final estimation samples consist of 598 observations for Application 1 (human capital); 448 observations for Application 2 (child abuse); and 597 observations for Application 3 (school-entry age). Statistics describing the socio-demographic characteristics of each estimation sample, compared to the full ALSPAC sample, are reported in Table A2. Consistent with Relton *et al.* (2015), children in the epigenetic subsamples are over selected from families with higher socio-economic status. Standard *t*-tests of differences in means between our estimation samples and the general ALSPAC sample (reported in the last three columns of Table A2) reveal that children in our estimation samples are more likely to have slightly older and more educated mothers, and fathers with a higher social class relative to the average ALSPAC child. In addition, our estimation samples are made up of a larger share of first-born

---

<sup>11</sup> 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 (Research Electronic Data Capture) electronic data capture tools hosted at the University of Bristol (Harris *et al.*, 2009). REDCap 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/>). 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).

<sup>12</sup> As standard in the literature, we compute GrimAge using the PCGrimAge algorithm (Higgins-Chen *et al.*, 2022).

children and display a small gender imbalance in Applications 1 and 2 (respectively, 59 and 62 percent of girls as compared to boys).

## 5. Main Results

### 5.1. Features of the MEGA clocks

One key theoretical assumption of our approach is that a single dimension captures the common variance across existing epigenetic clocks. We can check the plausibility of this assumption leveraging the exploratory factor analysis. When performing factor analysis 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).<sup>13</sup> As shown in Table A3, all four clocks have factor loadings larger than 0.4 in all estimation samples (a threshold commonly used in the literature to retain items, see Stevens, 2002) and all have more than 50 percent unique variance (i.e. variance that is not explained by the remaining clocks).

The distributions of the three MEGA clocks, the four original clocks, and chronological age are displayed in Figure 1 for the estimation sample in our child abuse analysis (Application 2).<sup>14</sup> Descriptive statistics for the same variables can be found in Table 1, separately for the estimation samples of all empirical applications. In our analysis of child abuse, children range in age from 14.6 to 19.3 years old. The age distribution follows a bimodal distribution around ages 15 and 17 – the timing of the two clinical assessments collecting DNA methylation data in ALSPAC. The four traditional clocks which are trained on adult samples and therefore are not fine-tuned to predict adolescent age, display greater dispersion and less precise centering around the chronological age compared to the MEGA clocks. This pattern is the first rough indication that aggregation improves the signal-to-noise ratio in our epigenetic aging measures.

---

<sup>13</sup> The selection of only one factor is unambiguous, as the second-largest factor has eigenvalue equal to 0.10.

<sup>14</sup> Application 2, illustrated in section 5.3, is the first in which the MEGA clocks are used as outcomes. In Figure 1, the values of  $MEGA_{SEM}$  are computed as the linear prediction of the latent variable based on the coefficients from the structural model, described more in detail in section 5.3.1.

The predictive performance of individual clocks against chronological age, shown in Figure A2, reveals substantial heterogeneity across measures. The Hannum and Horvath clock have the highest accuracy, with linear fits that closely follow the 45-degree line representing perfect prediction. GrimAge has the largest intercept shift (GrimAge estimates of epigenetic age are on average 17.4 years higher than chronological age), but the smallest dispersion and second-closest slope to unity after Horvath.

When comparing 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 one. Consistently, *MEGA<sub>SEM</sub>* and *MEGA<sub>FA</sub>* display the highest correlation with age in our sample, closely followed by GrimAge and *MEGA<sub>WGT</sub>* (Figure A4). All in all, the MEGA clocks appear to be better predictors of chronological age in our adolescent sample than the traditional clocks.<sup>15</sup> This enhanced performance reflects the variance reduction achieved through aggregation, which attenuates measurement error while preserving the signal common to all clocks.

## **5.2. Application 1: Epigenetic Aging and Human Capital**

Our first empirical application speaks to the relevance of epigenetic aging in economics research, by documenting its association with measures of human capital. We show that accelerated epigenetic aging in adolescence predicts fewer years of education, worse mental health, and weaker attachment to the labor market in early adulthood – even after accounting for family background, skills, and traditional health markers. This highlights epigenetic age as a new, policy-relevant indicator of human capital formation, extending recent evidence on its predictive power for health and cognition (e.g. Faul et al., 2023; Raffington et al., 2023a).

### **5.2.1. Estimating Equation**

We begin by estimating the association between accelerated epigenetic aging in late adolescence and early-adulthood cognitive and mental-health outcomes. We do so estimating the following linear regression model:

---

<sup>15</sup> These descriptive results presented for the estimation sample of Application 2 hold in the larger samples used in Applications 1 and 3.

$$Y_{t+1} = \delta_0 + \delta_1 MEGA_{i,t} + \delta_2 Health_{i,t} + \theta X_i + \varepsilon_{i,t} \quad (6)$$

where  $Y_{t+1}$  is one of the four outcomes measured in early adulthood (period  $t + 1$ ). All outcomes are coded as binary indicators for negative events, measured in the participants' mid-twenties. We consider two key human capital outcomes for young adults: (i) not having attained a university degree by age 26; and (ii) being neither in employment, education, nor training (NEET) at age 25. Our third and fourth outcomes relate to socio-emotional and mental health: (iii) an indicator for being above the diagnostic threshold of 12 for the self-assessed Short Moods and Feelings Questionnaire (SMFQ) at age 25 (Angold *et al.*, 1995), a widely used psychometric scale of mental health employed as a screening tool for depressive symptoms, with values ranging from 0 to 26; and (iv) having been diagnosed with depression by age 22. Further,  $MEGA_{i,t}$  is one of the MEGA clocks for child  $i$  measured in period  $t$ , namely late adolescence.  $X_i$  is a vector of the following individual controls: mother's age at birth of the study child; binary indicators for mother's education, father's social class, and the child's gender, birth year, and birth order. Importantly,  $X_i$  includes the child's age at the time of the DNA methylation assessment used to build the MEGA clocks, meaning that we can interpret the outcome as a measure of age acceleration. Lastly,  $Health_{i,t}$  is a vector of individual health and behavioral outcomes measured at age 15: BMI, being a smoker and drinking.<sup>16</sup>

A causal interpretation of  $\delta_1$  requires strong assumptions: conditional on the observed covariates included in the regression, the MEGA clock is unrelated to any unobserved confounders. Although a wide set of individual and family characteristics  $X_i$  are included as linear controls, other unobserved factors likely remain. Thus, the estimate of  $\delta_1$  should be interpreted as a conditional association, rather than a causal effect.

In an augmented version of equation (6), we also control for two latent factors interpretable as measures of cognitive and socio-emotional skills. These are obtained by extending our Structural Equation Model with two additional sets of measurement equations, using the same procedure applied to the MEGA clock. Both factors draw on age-16 assessments. The cognitive skills factor

---

<sup>16</sup> Being a smoker is defined as a binary indicator for smoking at least one cigarette per week, while 'Drinking' is defined as a binary indicator for any strictly positive answer to the question "Number of times young person has had 5+ full drinks in 24 hours, in the last 2 years". Both measures are reported by adolescents themselves.



is based on Mathematics, English, Science and aggregate test-scores, taken from linked administrative records in the National Pupil Database on General Certificate of Secondary Education (GCSE) exams, or an equivalent certification. The socio-emotional 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).<sup>17</sup>

### 5.2.2. Results

Estimates from equation (6) are reported in Tables 2 and 3, and illustrated in Figure 2. We find that one extra year of epigenetic age acceleration is associated with a 2 percentage points (pp) increase in the probability of not having a university degree by age 26; a 2 to 3 pp increase in the probability of being NEET at 25; a 2 pp increase in the probability of being above the diagnostic SMFQ threshold at age 25; and between a 1 to 2 pp increase in the probability of being diagnosed with depression by age 22. The observed effect sizes range from 4 to 27 percent of the mean outcome values, indicating that people with faster age acceleration in late adolescence experience a small to moderate deterioration in economic and mental-health outcomes by early adulthood. Our results on educational attainment are consistent with Mareckova *et al.* (2023), who find evidence of a modest negative association between Horvath age acceleration and IQ in a sample of young adult women.

Age acceleration is predictive of these outcomes even after controlling for negative health status and risky behaviors in adolescence (see Tables 2 and 3). The tables show estimates for coefficients  $\delta_1$  and  $\delta_2$  from Equation (6), 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). Our results reveal that the coefficients attached to the MEGA clocks reduce only slightly after the introduction of the *Health* control vector, suggesting that epigenetic age acceleration is capturing something more than the usual health and behavioral scales that can be measured in adolescence, and provide additional predictive power in explaining early adulthood outcomes.

---

<sup>17</sup> 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.

To benchmark the effect sizes of the associations between the MEGA clocks and early-adulthood outcomes, we include two additional sets of measurement equations to our Structural Equation Model, which are interpretable as latent factors for cognitive and socio-emotional skills, as described in section 5.2.1. Results are shown in Figure A5. We find that the MEGA clock weakly, but consistently predicts worse early-adulthood outcomes, consistent with Figure 2.<sup>18</sup> In contrast, the cognitive and socio-emotional factors show stronger predictive power, but mostly within their respective domains: the cognitive factor predicts better education and labor market outcomes, the socio-emotional factor better mental health. In particular, cognitive skills are strongly associated with the likelihood of completing a university degree, with a one standard-deviation increase in the cognitive skills factor linked to an 11 pp lower probability of not having a university degree by age 26. Similarly, socio-emotional skills show stronger associations with mental health indicators, with up to an 8 pp reduction in the probability of being diagnosed with depression by age 22. By comparison, the MEGA clock effects are smaller, but more consistent across the four outcomes, predicting human capital more broadly. Importantly, the associations with the MEGA clock are of comparable magnitude to those of the socio-emotional skill factor, just slightly smaller and in the opposite direction. This suggests that the MEGA clock associations, while modest, are still economically meaningful when benchmarked against established predictors of human capital development.

### **5.3. Application 2: The Consequences of Child Abuse for Epigenetic Aging**

Having established that epigenetic age predicts key early-adulthood outcomes, we next turn to its determinants, focusing on childhood shocks. Child abuse is a particularly salient event, with well-documented long-term consequences and broad epigenetic effects (Currie and Spatz Widom, 2010; Lawn *et al.*, 2018). We extend the existing literature by (i) specifically looking at epigenetic age rather than broader changes in DNA methylation, (ii) incorporating measures across different developmental periods and multiple reporters, and (iii) adding a longitudinal dimension to the analysis of child abuse. Our results show that abuse experienced before age 10 is associated with

---

<sup>18</sup> The MEGA clocks and the cognitive and socio-emotional factors are standardized in the figure, to enhance comparability.

half a year of accelerated epigenetic aging in late adolescence – an effect comparable to that of growing up in socio-economic disadvantage (Fiorito *et al.*, 2017).

### 5.3.1. Patterns of Child Abuse

We measure child abuse in ALSPAC using the responses to prospectively collected questions on child cruelty from mothers ( $M$ ) and their partners ( $P$ ).<sup>19</sup> Caregivers provided prospective reports of physical and emotional cruelty towards the study child, coming from themselves or their partners, several times throughout the data collection period.<sup>20</sup> 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.

Once cohort children reached adulthood (22+ years), they retrospectively reported any instances of physical or sexual abuse. These retrospective questions asked for the frequency with which 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’, assess how frequently adults in the family were violent toward the study child. We dichotomized these responses using specific cutoff values (Table A1) and created a composite measure of self-reported cruelty for each developmental period. This binary indicator equals one if participants reported experiencing any form of physical or emotional cruelty at least once during that period, 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). As before, we build one binary indicator of self-reported sex abuse for each developmental period,

---

<sup>19</sup> 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.

<sup>20</sup> 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.

equal to one if the study child reported having experienced any sexual abuse at least once and zero otherwise.

Figure 3 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.<sup>21</sup> This is consistent with the literature on parental reporting versus child self-reporting of adverse childhood experiences, which highlights substantial discrepancies in reporting 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 prospective parental reports to capture the full spectrum of childhood maltreatment, accounting for potential underreporting by parents and minimizing measurement error.

Table 4 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's 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

---

<sup>21</sup> The correlation table looks similar when restricting the sample to children with available DNA methylation information, which is how we define the estimation sample used to produce the main results of this paper (see Figure A1).

cells but those in column (C) are empty.<sup>22</sup> 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 increases 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.<sup>23</sup> 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 percent. There is a relatively large persistence in abuse over time: 15.6 percent of children in the sample have experienced some form of abuse in both developmental periods, suggesting that almost half of those experiencing abuse in early childhood (0-10 years old) are also victim of abuse in adolescence.

### 5.3.2. Estimating Equation

We study the associations between measures of child abuse over time defined above and the MEGA clocks by estimating the following linear regression model:

$$MEGA_{i,t} = \beta_0 + \beta_1 Abuse_{i,t} + \beta_2 Abuse_{i,t-1} + \gamma X_i + \varepsilon_{i,t} \quad (7)$$

where  $MEGA_{i,t}$  is again one of the MEGA clocks for child  $i$  measured in period  $t$ , here corresponding to late adolescence. Moreover,  $Abuse_{i,t}$  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. Lastly,  $X_i$  is the same vector of controls defined in section 5.2.1.

The previous literature on the epigenetic effects of childhood adversity leads us to expect both coefficients  $\beta_1$  and  $\beta_2$  to be positive. That is, we expect experiences of child cruelty or sex abuse to be associated with increased age acceleration. There is limited evidence on how abuse unfolds

---

<sup>22</sup> '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.

<sup>23</sup> The relatively high incidence of child abuse is potentially 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.

longitudinally across childhood, however. It is thus unclear *a priori* whether more recent abuse has a stronger association with epigenetic age acceleration in late adolescence ( $\beta_1 > \beta_2$ ) or whether early childhood adversity correlates with a stronger, long-lasting scar ( $\beta_1 < \beta_2$ ). Given the same logic as in our human capital analysis, here too, we interpret  $\beta_1$  and  $\beta_2$  as conditional associations rather than as the causal effect of child abuse on epigenetic age acceleration.

### 5.3.3. Results

Our OLS estimates of the coefficients in Equation (7) are reported in Table 5. Different versions of the MEGA clock appear across the columns:  $MEGA_{SEM}$  in column 1,  $MEGA_{FA}$  in column 2 and  $MEGA_{WGT}$  in column 3. The results are very similar across the different versions of the MEGA clock – experiencing any form of child abuse (sexual or cruelty) between ages 0 and 10 is associated with over half a year of accelerated epigenetic aging. Abuse experienced after age 11 does not appear to be significantly associated with age acceleration – the coefficient is generally negative and small in magnitude. This negative relationship is driven by the strong intertemporal correlation of our measure of abuse. When we control for child abuse occurring in one developmental period and not the other, estimated coefficients are always positive (albeit only significant for the 0-10 age range).

The half-a-year acceleration in epigenetic aging (an increase ranging from 21 to 28 percent of a standard deviation) is within the estimated range of other studies that have looked at the association between early life adversity and age acceleration: it is comparable, for example, to the age acceleration observed for children born in low socio-economic positions (Fiorito *et al.*, 2017). 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).

When we disaggregate child abuse into child cruelty and sex abuse, we find that only child cruelty between ages 0 and 10 is strongly and precisely associated with epigenetic age acceleration (see Table 6). The positive coefficient amounts to about half a year greater age acceleration for children exposed to child cruelty by age 11, similar in magnitude to the effects reported in Table 5. Estimated coefficients for sex abuse are also positive for both periods, albeit about half the size

and imprecisely estimated (statistically insignificant at conventional thresholds). This suggests that the epigenetic trace left by the combined abuse measure may be driven by child cruelty more than sex abuse. This interpretation requires caution, however, given the composition of our abuse measure. As child cruelty represents the largest component of overall abuse (as detailed in Table 4) and only 3.6 percent of the sample reports having experienced sexual abuse between ages 0 and 10, our analyses may lack sufficient statistical power to detect significant effects of sex abuse in isolation.

#### **5.4. Application 3: The Effect of School-Entry Age on Epigenetic Aging**

We explore the determinants of epigenetic age acceleration in our third, and final, empirical application. Specifically, we exploit the U.K. school-entry age cutoff in a regression discontinuity design to isolate the causal impact of delayed school entry on epigenetic age acceleration at age seven. This approach, standard in economics for studying the effects of school starting age on cognitive and labor market outcomes (e.g., Bedard and Dhuey, 2006; Black, Devereux, and Salvanes, 2011; Fredriksson and Öckert, 2014), allows us to provide the first causal estimates of the impact of the timing of school entry on biological development. We are also one of the first to examine the causal effects of quasi-exogenous shocks or policy changes on epigenetic aging (see Schmitz and Duque, 2022, for a related exception). We find that delayed school entry leads to accelerated epigenetic aging for children from low-SES backgrounds and small anti-aging effect for children from high-SES families.

##### **5.4.1. Estimating Equation**

We use a sharp regression discontinuity design (RDD) based on the school-entry cutoff in the U.K. to estimate the causal effect of delayed school entry on epigenetic age acceleration. Children born on or after September 1<sup>st</sup> are required to enter school a year later than those born just before this cutoff. This rule induces quasi-random variation in school starting age among children born within a narrow window around the cutoff, which we exploit for identification. Our estimating equation is the following:

$$MEGA_{i,t} = \theta_0 + \theta_1 Treat_i + \theta_2 MoB_i + \theta_3 Treat_i \times MoB_i + \mu W_i + v_{i,t} \quad (8)$$



where  $Treat_i$  is a binary indicator for being born on or after September 1<sup>st</sup> of a given year and  $MoB_i$  is the running variable, i.e. the child's month of birth normalized to 0 in September (thus taking values -1 for August, 1 for October, etc.). As previously,  $MEGA_{i,t}$  is child  $i$ 's MEGA clock measured at time  $t$ , which in this case we take as age 7 – the earliest DNA methylation measurement point since school entry. Lastly,  $W_i$  is a vector of exogenous child controls, namely the child's gender, age at the time of DNA methylation measurement and their year of birth. In our main specification, we restrict the analysis to children born within four months around the cutoff, that is those born from May to December. To test for the robustness of our results, we additionally estimate equation (8) using children born within three and within two months from the cutoff.

The parameter  $\theta_1$  captures the effect of being born after the cutoff (and thus entering school a year later) on epigenetic aging, measured by the MEGA clock at age 7. While we may observe a non-zero average treatment effect, it is also plausible that the impact of delayed school entry varies according to children's home or childcare environments. Specifically, children from high socioeconomic status (SES) backgrounds may benefit from delayed entry, whereas those from lower SES backgrounds might experience more favorable developmental conditions within the school setting than at home (Anderson *et al.*, 2011; Holford and Rabe, 2022). To account for potential treatment effect heterogeneity by SES, we also run an augmented model specification controlling for an occupation-based measure of paternal social class at the time of the child's birth and interacting it with  $Treat_i$ ,  $MoB_i$  and their interaction.

#### 5.4.2. Results

We first look at the distribution of age acceleration by month of birth in the estimation sample. We compute age acceleration as the residuals from regressions of the MEGA clock on chronological age, run separately for each of the three time points at which DNA methylation is observed (child's birth, age 7 and age 15-19). For simplicity, Figure A6 only shows results for the MEGA factor clock. As expected, there is no discontinuity in age acceleration by birth month at the time of birth, a result that primarily serves as a placebo test.<sup>24</sup> In contrast, we observe clear discontinuities at the school-entry cutoff when age acceleration is measured at ages 7 and 15, with higher values for

---

<sup>24</sup> Because the epigenetic clocks used to build the MEGA clock (Hannum, Horvath, PhenoAge, and GrimAge) were primarily developed and validated in adult populations, their application in children – particularly in newborns – should be interpreted with caution.

children born in September compared to those born in August. These patterns provide a first piece of evidence consistent with an effect of school-entry age on epigenetic aging.

Moving to a regression framework, Panel A of Table 7 shows results from estimating equation (8). Consistent with the evidence from Figure A6, our RDD estimates indicate that delaying school entry by one year leads to a statistically significant acceleration in epigenetic age of 0.3 to 0.6 years, measured at age 7.

We disaggregate the main results by paternal social class at birth (results using the MEGA factor clock are shown in the blue estimates in Figure 5) and find that the average detrimental effect of delaying school entry on epigenetic aging is driven entirely by children from low-SES backgrounds. No significant impact is observed among children from higher-SES families. These findings suggest that earlier access to structured school environments may offer protective biological effects, especially for socio-economically disadvantaged children.

This study provides the first causal evidence linking formal schooling to biological aging processes during early childhood. These findings highlight the importance of early institutional exposure not only for cognitive and social development, as previously documented in the economics literature (e.g., Bedard and Dhuey, 2006; Black, Devereux, and Salvanes, 2011; Fredriksson and Öckert, 2014), but also for objectively measurable biological outcomes.

## 5.5. Sensitivity Tests

We first compare results from our three empirical applications with those that would be obtained by using traditional epigenetic clocks individually, rather than the MEGA clock. Figure A7 presents regression coefficients for the four clocks on early-adulthood human capital outcomes (Application 1). The results indicate that the patterns shown in Figure 2 are driven by the GrimAge clock, which is the only one to yield point estimates significantly larger than zero at least at the 10 percent level. In comparison to models based on individual clocks, regressions using the MEGA clock yield more precisely estimated and larger coefficients, consistent with a reduction in classical measurement error.

We turn now to consider our analysis of child abuse (Application 2). Estimates of coefficients  $\beta_1$  and  $\beta_2$  from Equation (7) for each of the outcomes indicated in the legend are reported in Figure 4. First, we present results for each of the single clocks used to build the MEGA clock (blue round

markers) and, second, we replicate the estimated coefficients for our MEGA clocks (see Table 5) for convenience (red diamond markers). Estimates of child abuse before age 11 using traditional clocks result in larger average 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 GrimAge (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 A8, we replicate the same exercise but disaggregate child abuse into its two components: cruelty and sex abuse (similar to Table 6). Results are again consistent with those from Figure 4. Estimates based on the MEGA clocks have smaller confidence intervals on average than the traditional clocks, while point estimates are aligned with those from the GrimAge clock.

Lastly, we repeat this exercise for our analysis of school-entry age (Application 3), replicating results from Panel A of Table 7 (see Figure A9). The estimated treatment effects of being born after August 31<sup>st</sup> are positive for all individual clocks except Hannum's, with point estimates from the MEGA clocks falling within the range of the individual clocks' ones. Here again we confirm that using the MEGA clocks yields smaller standard errors than each of the four traditional clocks individually, thereby improving precision.

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 in all estimation samples. Results for our analysis of human capital are displayed in Figure A10. The point estimates from even-numbered columns of Tables 2 and 3 remain quite robust when excluding either one of the Hannum, Horvath or PhenoAge clocks. Interestingly, our 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. Results for our analysis of child abuse are also robust to the exclusion of either one clock (see Figure A11). 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 percent level. We also apply this leave-one-out computational strategy to our analysis of the effects

of school-entry age (see Figure A12). Results are presented only for the factor analysis and weighted index approach, due to a lack in convergence for the Structural Equation Models. Here too results appear to be remarkably robust to the exclusion of either one clock, remaining close to the treatment effects displayed in Panel A of Table 7. The only exception appears to be Horvath's clock, which is also the one displaying the largest point estimates in Figure A9.

We then turn to consider the sensitivity of our results to our measurement of exposure to 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 A4 shows that our results in Table 5 are not driven by this measurement choice: the estimated 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 A4) results in the lowest standard errors across all methods, suggesting that harnessing these different sources of information can help reduce the measurement error linked to the under-reporting of sensitive constructs.

Next, we examine the sensitivity of our school-entry age results to bandwidth selection. Theoretically, smaller bandwidths imply lower bias in the causal estimates but bigger confidence intervals, due to the smaller sample size. Restricting the bandwidth from four to three months around the cutoff (Panel B of Table 7) yields similar point estimates as the main results in Panel A, although they are no longer statistically different from zero due to the loss in precision. Further narrowing the sample to children born between July and October (Panel C of Table 7) results in a doubling of the standard errors but also of the point estimates; focusing only on those born just before or just after the cutoff – the cleanest approach from a theoretical perspective – suggests that delaying school entry increases epigenetic aging by 0.9 to 1.2 years.

Finally, we assess whether the aging penalty associated with delayed school entry persists beyond age 7, into adolescence. Table A5 indicates that it does not: while point estimates remain positive on average, they are much smaller (one third to one fifth of the results at age 7) and noisily estimated, such that none of them is statistically different from zero at conventional levels. The epigenetic aging differences induced by the school-entry cutoff fade over time, mirroring a common finding in the literature that the effects of school starting age on various outcomes tend to weaken as children age (e.g. Bedard and Dhuey, 2006).

## **6. Potential Mechanisms**

We now discuss how epigenetic aging may serve as a biological pathway connecting early-life exposures, including child abuse and age at school-entry, to downstream outcomes in cognition and mental health. Due to the limited range of biomarkers in our data, we are constrained in our ability to directly investigate biological channels beyond DNA methylation. Consequently, we draw on insights from the molecular epigenetics literature and observed behavioral patterns to provide a coherent narrative of how environmental exposures may translate into accelerated epigenetic aging and later into worse adult outcomes. The goal is not to identify definitive causal pathways, but to offer biologically and behaviorally grounded interpretations of our findings.

### **6.1. Human Capital**

Accelerated epigenetic aging is likely to influence later cognitive and mental-health outcomes through multiple interconnected biological pathways. The machine learning approaches used to develop epigenetic clocks are fundamentally agnostic to the underlying biology. This means that the CpG sites they rely on capture a wide range of processes rather than a single mechanism. These sites regulate cell growth and survival, inflammation, antiviral responses, and DNA repair (Horvath, 2013; Hannum *et al.*, 2013; Levine *et al.*, 2018), and are over-represented in gene sets involved in the immune system, lipid function, and adipocytes communication (Lu *et al.*, 2019). Although the adults in our sample are still too young to show clinical aging traits, the molecular processes captured by epigenetic clocks could already affect brain function and correlate with early cognitive deterioration (Felt *et al.*, 2023). Some of the most central gene regions in epigenetic clocks are involved in neuronal pathways, including neurogenesis, neuron differentiation, and neuron death (Han *et al.*, 2018), providing a potential biological link to learning, memory, and other human-capital outcomes.

### **6.2. Child Abuse**

Child abuse may accelerate epigenetic aging through both biological and behavioral stress responses. One plausible biological mechanism involves changes in the proportion of immune

cells, particularly leukocytes, which in turn affects methylation patterns (Lima *et al.*, 2022).<sup>25</sup> Child abuse, in particular, has been consistently associated with changes in immune cell proportions, driven by heightened inflammatory activity (D’Elia *et al.*, 2018; Renna *et al.*, 2021). To test whether blood cell counts mediate the relationship between child abuse and epigenetic aging in our sample, we compute standard cell counts for peripheral blood samples from the age 15-19 DNA methylation data in ALSPAC,<sup>26</sup> 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 A13). When controlling for blood cell counts in our main regressions, the magnitude of the estimated coefficients decreases by approximately one-third, suggesting a mediating role (Table A6).

Beyond cellular changes, stress-related pathways involving the hypothalamic-pituitary-adrenal axis and cortisol production are also likely involved (Dammering *et al.*, 2021; Suarez *et al.*, 2018). Individual differences in resilience – through emotion regulation and self-control – have also been shown to moderate stress induced epigenetic aging (Harvanek *et al.*, 2021, 2023).<sup>27</sup> Risky behaviors and lifestyle changes can also play a role in how stress affects epigenetic aging (Schmitz *et al.*, 2022; Jung *et al.*, 2023). Both biological and behavioral stress responses can additionally affect other hallmarks of aging, including circadian rhythms, immune system functioning, and nutrition, which subsequently influence the epigenome (Harvanek *et al.*, 2023).

### 6.3. School-Entry Age

We show that structured educational environments appear to influence epigenetic aging, particularly among socioeconomically disadvantaged children. While much of the existing research has focused on non-cognitive skills, academic performance, and behavioral outcomes (e.g., Bedard and Dhuey, 2006; Fredriksson and Öckert, 2014), systematic evidence on the potential protective health effects of early exposure to structured educational environments

---

<sup>25</sup> Since distinct blood cell types exhibit unique methylation patterns, cell counts are typically controlled for in epigenetic analyses. Yet, these counts may function as ‘bad controls’, as they mediate the relationship between environmental exposures and the epigenome.

<sup>26</sup> The cell types considered here are B-lymphocytes (Bcell), CD4+ T-lymphocytes (CD4T), CD8+ T-lymphocytes (CD8T), granulocytes (Gran), monocytes (Mono), and natural killers (NK).

<sup>27</sup> Research on posttraumatic stress disorder (PTSD) suggests that specific symptom clusters, such as emotional withdrawal, sleep problems, and cognitive dysfunction, correlate more strongly with changes in epigenetic aging than the overall PTSD severity (Katrinli *et al.*, 2020; Na *et al.*, 2022).

remains limited. Some studies, such as Anderson *et al.* (2011), suggest that the transition from a relatively unstructured home environment to the routine and regulation of school life may reduce snacking opportunities and promote physical activity – plausibly contributing to improved physical health. Holford and Rabe (2022) similarly find that children assessed later in the school year (and thus exposed to more schooling) display lower weight-for-height ratios, suggesting cumulative effects of structured schooling on health-related behaviors.

Building on these intuitions, we further examine treatment effect heterogeneity by SES in our data. To test whether school systematically offers healthier environments for low-SES children, we analyze a range of health and nutrition outcomes in the full ALSPAC sample in Figure 5. Children from manual occupation households who enter school later because of their birthdate display not only accelerated epigenetic aging, but also worse general health (as reported by their mothers), and higher consumption of processed foods and sugary and fatty items at age 7.<sup>28</sup> Conversely, children from higher social classes benefit from delayed school entry, exhibiting healthier dietary patterns. While all effects point in the same direction, the magnitude of the impact on epigenetic aging is substantially larger, suggesting that it may capture broader or more cumulative biological consequences than these behavioral outcomes – highlighting the value of measuring it directly. These results suggest that structured school environments may serve as a compensatory setting protecting the health and epigenetic aging trajectories of disadvantaged children, supporting the hypothesis that early-life social environments can mitigate the biological embedding of inequality.

## 7. Conclusion

As economists work to develop richer models of human behavior, there is growing interest in measuring and analyzing the epigenetic processes through which people’s environments affect their biological functioning. A focus on biological, rather than chronological, age opens the door to a deeper understanding of people’s health and disease risk, physical functioning, and cognitive performance as they age. Epigenetic clocks have emerged as the leading tool for summarizing age-

---

<sup>28</sup> In Figure 5, ‘high-sugar, high-fat diet’ refers to a standardized dietary pattern score derived by Ambrosini *et al.* (2016) using reduced rank regression, associated with more energy density, higher percent energy from free sugars and total fat, and lower density of fiber. Similarly, ‘Processed diet’ is a dietary pattern score indicating high consumption of processed foods, chips and soft drinks, as derived by Smith *et al.* (2013).



related epigenetic markers – now available in many standard large-scale, population-representative data sets – into a single measure of biological age.

We make an important methodological contribution by developing a new metric – the MEGA clock – which allows researchers to combine established epigenetic clocks to increase the estimation precision in models of the determinants and consequences of epigenetic aging. Importantly, the methodological approach we propose is flexible and designed to accommodate future epigenetic clocks as they are developed. The results of our empirical applications not only establish the validity and robustness of the MEGA clock, but also the usefulness of epigenetic clocks in understanding the biological mechanisms through which socio-economic factors influence health and human capital.

The results of our empirical applications add to the growing evidence that epigenetic age is a policy-relevant measure of health and well-being. We are led to two key conclusions. First, there is a potential for epigenetics to be one mechanism linking environmental conditions in one lifecycle stage to social and economic well-being in the next. For example, we find that early-life exposure to child abuse is associated with accelerated adolescent epigenetic aging, highlighting the biological impact of early-life adversity. Accelerated epigenetic aging in adolescence, in turn, predicts worse cognitive and mental-health outcomes in early adulthood with lifetime implications for health and productivity.

Second, although many of the empirical relations we examine are correlational, we are among the first to provide some causal evidence of the effect of childhood events on the pace of epigenetic aging. Our research breaks new ground by using a regression discontinuity design to identify the causal impact of the timing of school entry on epigenetic aging. Children starting school one year later experience faster epigenetic aging, particularly if they are from low-SES backgrounds. Early access to structured educational environments may be a protective factor against the biological embedding of stressful or less structured home settings.

Our exploration of potential mechanisms suggests that accelerated epigenetic aging likely reflects both biological and behavioral responses to early-life experiences. For instance, exposure to child abuse appears to alter immune cell profiles and stress-response systems, while early school entry may provide structured environments that promote healthier routines, particularly for disadvantaged children.

More broadly, our findings underscore the potential for epigenetic measures to reshape how economists study health, well-being, and human capital development. By tracing how early-life environments become biologically embedded, epigenetic clocks offer an avenue to evaluate not only long-run consequences of adversity but also the protective effects of early interventions. The MEGA clock offers an easily interpretable, reliable tool for studying the link between environmental factors and biological aging processes, without requiring specialized training in molecular biology, enabling more targeted and effective policy interventions. Future research should continue refining these measures and exploring their applications, investigating persistence across the life cycle, testing the generalizability of results across populations, and linking specific policies – such as preschool or nutrition programs – to biological aging trajectories. By bridging biology and economics, our work contributes to laying the groundwork for a more comprehensive understanding of human development and the complex interplay between nature and nurture.

## References

- Abdellaoui, A., Yengo, L., Verweij, K. J., and Visscher, P. M. (2023). 15 years of GWAS discovery: realizing the promise. *The American Journal of Human Genetics*, 110, 179–194.
- Almond, D., and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, 25, 153–172.
- Ambrosini, G. L., Johns, D. J., Northstone, K., Emmett, P. M., and Jebb, S. A. (2016). Free sugars and total fat are important characteristics of a dietary pattern associated with adiposity across childhood and adolescence. *Journal of Nutrition*, 146, 778–784.
- 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, 1481–95.
- Anderson, P. M., Butcher, K. F., Cascio, E. U., and Schanzenbach, D. W. (2011). Is being in school better? The impact of school on children's BMI when starting age is endogenous. *Journal of Health Economics*, 30, 977–986.
- Angold, A., Costello, E. J., Messer, S. C., and 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.
- Baldwin, J. R., Reuben, A., Newbury, J. B., and Danese, A. (2019). Agreement between prospective and retrospective measures of childhood maltreatment: a systematic review and meta-analysis. *JAMA psychiatry*, 76, 584–593.
- Beauchamp, J. P., Cesarini, D., and Johannesson, M. (2017). The psychometric and empirical properties of measures of risk preferences. *Journal of Risk and Uncertainty*, 54, 203–237.
- Bedard, K., and Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *Quarterly Journal of Economics*, 121, 1437–1472.
- Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., ... and Rakyan, V. K. (2019). DNA methylation aging clocks: challenges and recommendations. *Genome Biology*, 20, 249.
- Belsky, D. W., and Harden, K. P. (2019). Phenotypic Annotation: Using Polygenic Scores to Translate Discoveries From Genome-Wide Association Studies From the Top Down. *Current Directions in Psychological Science*, 28, 82–90.
- Benjamin, D. J., Chabris, C. F., Glaeser, E. L., Gudnason, V., Harris, T. B., Laibson, D. I., Launer, L., and Purcell, S. (2007). Genoeconomics. In Weinstein, M., Vaupel, J. W., and Wachter, K. W. (eds.), *Biosocial Surveys*. Committee on Population, Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press.

- Benjamin, D. J., Cesarini, D., Turley, P., and Young, A. S. (2024). *Social-Science Genomics: Progress, Challenges, and Future Directions*. National Bureau of Economic Research Working Paper No. 32404.
- Björklund, A., Lindahl, M., and Plug, E. (2006). The origins of intergenerational associations: Lessons from Swedish adoption data. *Quarterly Journal of Economics*, 121, 999–1028.
- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *Review of Economics and Statistics*, 93, 455–467.
- Bowles, S., and Gintis, H. (2002). The inheritance of inequality. *Journal of Economic Perspectives* 16, 3–30.
- 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., and Nolte, T. (2020). Childhood maltreatment and DNA methylation: A systematic review. *Neuroscience and Biobehavioral Reviews*, 112, 392–409.
- Cesarini, D., Johannesson, M., Lichtenstein, P., Sandewall, Ö., and Wallace, B. (2010). Genetic variation in financial decision-making. *Journal of Finance*, 65, 1725–1754.
- Chen, L., Dong, Y., Bhagatwala, J., Raed, A., Huang, Y., and Zhu, H. (2019). Effects of vitamin D3 supplementation on epigenetic aging in overweight and obese African Americans with suboptimal vitamin D status: a randomized clinical trial. *The Journals of Gerontology: Series A*, 74, 91–98.
- Cole, S. W. (2009). Social regulation of human gene expression. *Current Directions in Psychological Science*, 18, 132–37.
- Costi, C., Menta, G., Fiorito, G., and D’Ambrosio, C. (2025a). The intergenerational association of epigenetic modifications between mothers and offspring, from birth to adolescence. *Economics and Human Biology*, 58, 101509.
- Costi, C., Menta, G., Caro, J. C., Clark, A. E., and D’Ambrosio, C. (2025b). *Poverty gets under the skin: UK evidence from biological aging*. Mimeo.
- Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78, 883–931.
- Currie, J., and 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., and 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., ... and 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., and 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., ... and 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, 1611–1619.
- Faul, J. D., Kim, J. K., Levine, M. E., Thyagarajan, B., Weir, D. R., and 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, e2215840120.
- Felt, J. M., Yusupov, N., Harrington, K. D., Fietz, J., Sliwinski, M. J., Ram, N., ... and 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.
- Fiorito, G., Polidoro, S., Dugué, P. A., Kivimaki, M., Ponzi, E., Matullo, G., ... and Vineis, P. (2017). Social adversity and epigenetic aging: a multi-cohort study on socioeconomic differences in peripheral blood DNA methylation. *Scientific Reports*, 7, 16266.
- Fisher, H.L., Bunn, A., Jacobs, C., Moran, P., Bifulco, A. (2011). Concordance between mother and offspring retrospective reports of childhood adversity. *Child Abuse and Neglect*, 35, 117–122.
- Fletcher, J. M. (2009). Childhood mistreatment and adolescent and young adult depression. *Social Science and Medicine*, 68, 799–806.
- 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.
- Fredriksson, P., and Öckert, B. (2014). Life-cycle effects of age at school start. *Economic Journal*, 124, 977–1004.
- Goldberger, A. S. (1976). *Twin methods: A skeptical view*. Social Systems Research Institute, University of Wisconsin-Madison.
- Goodman, R., Ford, T., Simmons, H., Gatward, R., and 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, 534–539.
- Hannum, G., Guinney, J., Zhao, L., Zhang, L. I., Hughes, G., Sadda, S., ... and Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 49, 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., and Sinha, R. (2021). Psychological and biological resilience modulates the effects of stress on epigenetic aging. *Translational Psychiatry*, 11, 601.
- Harvanek, Z. M., Boks, M. P., Vinkers, C. H., and Higgins-Chen, A. T. (2023). The cutting edge of epigenetic clocks: In search of mechanisms linking aging and mental health. *Biological Psychiatry*, 94, 694–705.
- Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences*, 104, 13250–13255.
- Henkhaus, L. E. (2022). The lasting consequences of childhood sexual abuse on human capital and economic well-being. *Health Economics*, 31, 1954–1972.
- Hertzman, C. (1999). The biological embedding of early experience and its effects on health in adulthood. *Annals of the New York Academy of Sciences*, 896, 85–95.
- Higgins-Chen, A. T., Thrush, K. L., Wang, Y., Minter, C. J., Kuo, P. L., Wang, M., ... and Levine, M. E. (2022). A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. *Nature Aging*, 2, 644–661.
- Hoffmann, A., and Spengler, D. (2012). DNA memories of early social life. *Neuroscience*, 4, 64–75.
- Holford, A., and Rabe, B. (2022). Going universal. The impact of free school lunches on child body weight outcomes. *Journal of Public Economics Plus*, 3, 100016.
- Houmark, M. A., Ronda, V., and 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., ... and Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, 13, 86.
- Houtepen, L. C., Heron, J., Suderman, M. J., Tilling, K., and 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, 3156.
- Horvath, S., and Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, 19, 371–384.

- Jung, J., McCartney, D. L., Wagner, J., Yoo, J., Bell, A. S., Mavromatis, L. A., ... and Lohoff, F. W. (2023). Additive effects of stress and alcohol exposure on accelerated epigenetic aging in alcohol use disorder. *Biological Psychiatry*, 93, 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., ... and Smith, A. K. (2020). Evaluating the impact of trauma and PTSD on epigenetic prediction of lifespan and neural integrity. *Neuropsychopharmacology*, 45, 1609–1616.
- Klengel, T., Pape, J., Binder, E. B., and 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., and 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., and Scott, A. J. (2025). On the limits of chronological age. *Annual Review of Economics*, 17, 151-173.
- Lawn, R. B., Anderson, E. L., Suderman, M., Simpkin, A. J., Gaunt, T. R., Teschendorff, A. E., ... and 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, A. (1974). Home Investments in Children. *Journal of Political Economy*, 82, S111–S131.
- Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., ... and Horvath, S. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*, 10, 573.
- Lima, C. N., Suchting, R., Scaini, G., Cuellar, V. A., Del Favero-Campbell, A., Walss-Bass, C., ... and Fries, G. R. (2022). Epigenetic GrimAge acceleration and cognitive impairment in bipolar disorder. *European Neuropsychopharmacology*, 62, 10–21.
- Liu, Z., Leung, D., Thrush, K., Zhao, W., Ratliff, S., Tanaka, T., ... and Levine, M. E. (2020). Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging Cell*, 19, e13229.
- López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell*, 153, 1194–1217.
- Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., ... and Horvath, S. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*, 11, 303.
- Lussier, A. A., Zhu, Y., Smith, B. J., Cerutti, J., Fisher, J., Melton, P., ... and 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 and Adolescent Health*, 7, 532.
- Maddock, J., Castillo-Fernandez, J., Wong, A., Cooper, R., Richards, M., Ong, K. K., ... and Hardy, R. (2020). DNA methylation age and physical and cognitive aging. *The Journals of Gerontology: Series A*, 75, 504–511.
- Mareckova, K., Pacinkova, A., Marecek, R., Sebejova, L., Izakovicova Holla, L., Klanova, J., ... and 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., ... and 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., and Aiello, A.E. (2025). *Addressing conceptual and measurement disagreement of biological age: A confirmatory factor analysis*. Mimeo.
- Maselli del Giudice, A., La Mantia, I., Barbara, F., Ciccarone, S., Ragno, M. S., de Robertis, V., ... and Di Stadio, A. (2022). Use of nutraceuticals in elderly to fight inflammation and immunosenescence: a randomized case-control study. *Nutrients*, 14, 3476.
- McCartney, D. L., Stevenson, A. J., Walker, R. M., Gibson, J., Morris, S. W., Campbell, A., ... and Marioni, R. E. (2018). Investigating the relationship between DNA methylation age acceleration and risk factors for Alzheimer's disease. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, 10, 429–437.
- McCrory, C., Fiorito, G., Hernandez, B., Polidoro, S., O'Halloran, A. M., Hever, A., ... and 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, 741–749.
- Na, P. J., Montalvo-Ortiz, J. L., Nagamatsu, S. T., Southwick, S. M., Krystal, J. H., Gelernter, J., and 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, 41671.
- Newbury, J. B., Arseneault, L., Moffitt T. E., Caspi, A., Danese, A., Baldwin, J. R., and 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.
- Niccodemi, G., Menta, G., Turner, J., and D'Ambrosio, C. (2022). Pace of aging, family environment and cognitive skills in children and adolescents. *SSM-Population Health*, 20, 101280.



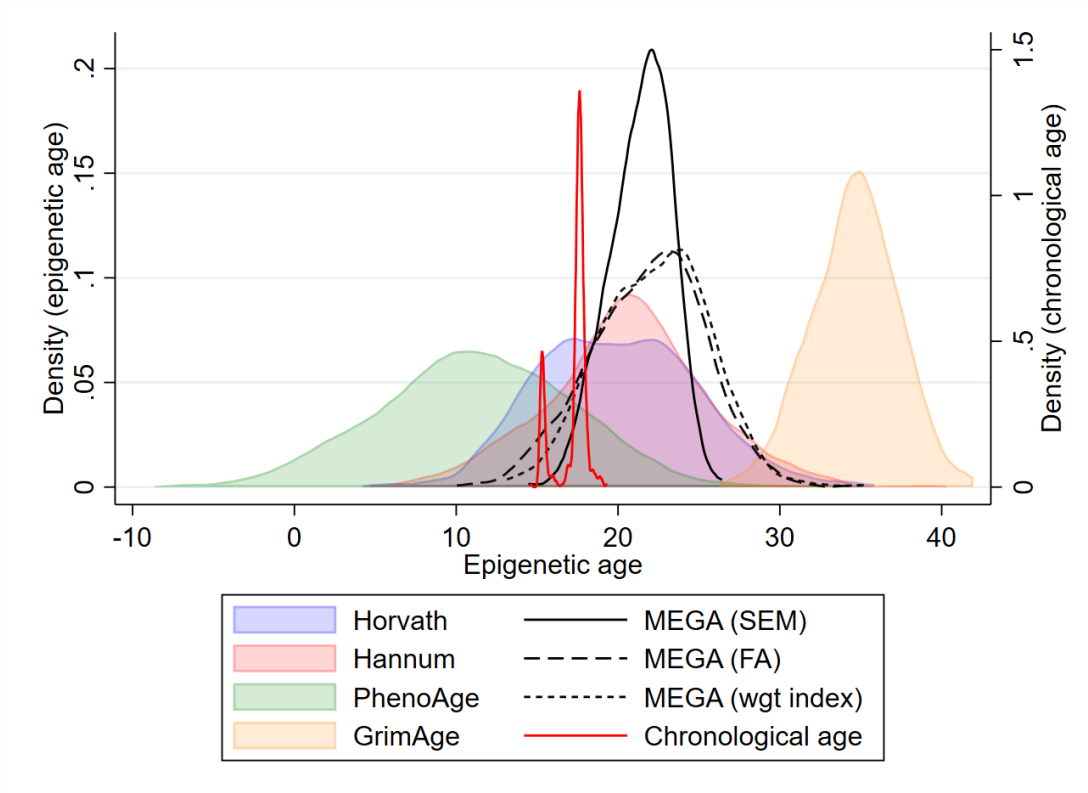
- Northstone, K., Lewcock, M., Groom, A., Boyd, A., Macleod, J., Timpson, N., and Wells, N. (2019). The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Research*, 4, 15132.1.
- Northstone, K., Shlomo, Y. B., Teyhan, A., Hill, A., Groom, A., Mumme, M., ... and Golding, J. (2023). The Avon Longitudinal Study of Parents and Children ALSPAC G0 Partners: A cohort profile. *Wellcome Open Research*, 8, 18782.2.
- Nwanaji-Enwerem, J. C., Chung, F. F. L., Van der Laan, L., Novoloaca, A., Cuenin, C., Johansson, H., ... and Herceg, Z. (2021). An epigenetic aging analysis of randomized metformin and weight loss interventions in overweight postmenopausal breast cancer survivors. *Clinical Epigenetics*, 13, 224.
- Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., ... and 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., and 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., Holleczeck, B., Saum, K. U., and Brenner, H. (2016). Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clinical Epigenetics*, 8, 64.
- Petrovic, D., Carmeli, C., Sandoval, J. L., Bodinier, B., Chadeau-Hyam, M., Schrepft, S., ... and Stringhini, S. (2023). Life-course socioeconomic factors are associated with markers of epigenetic aging in a population-based study. *Psychoneuroendocrinology*, 147, 105976.
- Plomin, R., and Bergeman, C. S. (1991). The nature of nurture: Genetic influence on “environmental” measures. *Behavioral and Brain Sciences* 14, 373–386.
- Raffington, L., Tanksley, P. T., Sabhlok, A., Vinnik, L., Mallard, T., King, L. S., ... and 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., ... and Harden, K.P. (2023b). Associations of DNA-Methylation Measures of Biological Aging With Social Disparities in Child and Adolescent Mental Health. *Clinical Psychological Science*, 12, 551–562.
- Renna, M. E., Peng, J., Shrout, M. R., Madison, A. A., Andridge, R., Alfano, C. M., ... and 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., ... and 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., ... and Mehta, D. (2023). Childhood maltreatment and DNA methylation: a systematic review. *Neuroscience and Biobehavioral Reviews*, 147, 105079.
- Sanz-de-Galdeano, A., and Terskaya, A. (2025). Sibling Differences in Genetic Propensity for Education: How do Parents React? *Review of Economics and Statistics*, forthcoming.
- Schmitz, L. L., Zhao, W., Ratliff, S. M., Goodwin, J., Miao, J., Lu, Q., ... and 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, 589–611.
- Shonkoff, J. P., Boyce, W. T., and McEwen, B. S. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities. *JAMA*, 301, 2252–2259.
- Simpkin, A. J., Hemani, G., Suderman, M., Gaunt, T. R., Lyttleton, O., Mcardle, W. L., ... and Smith, G. D. (2016). Prenatal and early life influences on epigenetic age in children: a study of mother–offspring pairs from two cohort studies. *Human Molecular Genetics*, 25, 191–201.
- Smith, A., Emmett, P., Newby, P., and Northstone, K. (2013). Dietary patterns obtained through principal components analysis: The effect of input variable quantification. *British Journal of Nutrition*, 109, 1881–1891.
- Soares, A.G., Zimmerman, A., Zammit, S., Karl, A., Halligan, S.L., and 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., ... and Raikonen, K. (2018). The epigenetic clock and pubertal, neuroendocrine, psychiatric, and cognitive outcomes in adolescents. *Clinical Epigenetics*, 10, 96.
- Suglia, S. F., Clark, C. J., Boynton-Jarrett, R., Kressin, N. R., and Koenen, K. C. (2014). Child maltreatment and hypertension in young adulthood. *BMC Public Health*, 14, 1149.
- Szyf, M. (2009). The early life environment and the epigenome. *Biochimica et Biophysica Acta*, 1790, 878–85.
- Taubman, P. (1976). Earnings, education, genetics, and environment. *Journal of Human Resources*, 11, 447–461.
- Teschendorff, A. E., and Horvath, S. (2025). Epigenetic ageing clocks: statistical methods and emerging computational challenges. *Nature Reviews Genetics*, 26, 350–368.
- Warren, A.-S., Goldsmith, K. A., and 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, 888–896.

- Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., ... and Lee, J. Y. (2022). A saturated map of common genetic variants associated with human height. *Nature*, 610, 704–712.
- Zhang, Q., Vallerger, C. L., Walker, R. M., Lin, T., Henders, A. K., Montgomery, G. W., ... and Visscher, P. M. (2019). Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Medicine*, 11, 54.

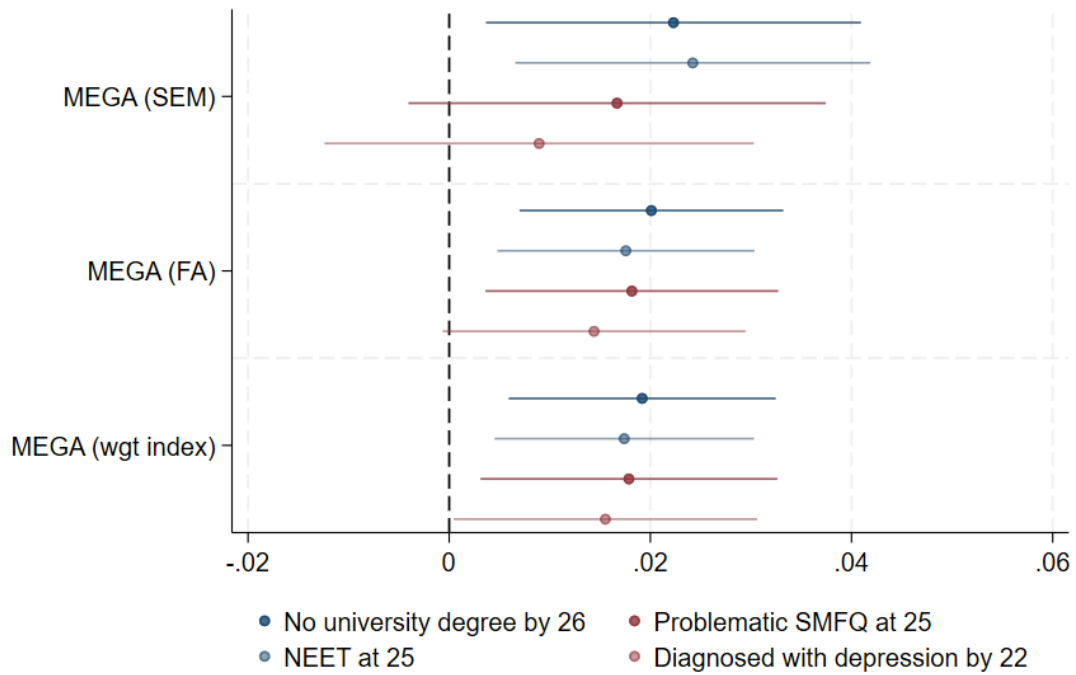
## Figures and Tables

Figure 1: Distribution of epigenetic age by clock and chronological age



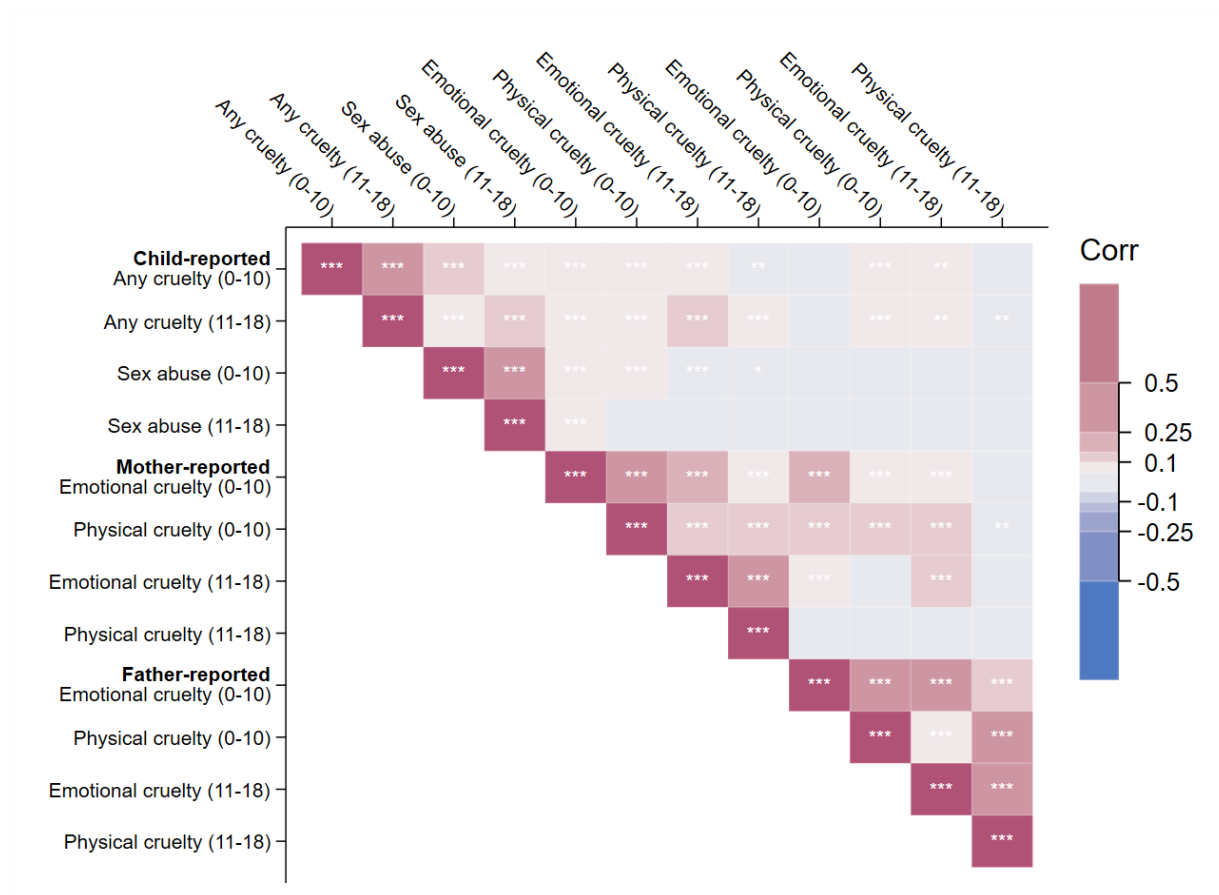
*Notes:* The figure displays the distribution of epigenetic age as measured by four individual clocks (Horvath, Hannum, PhenoAge, and GrimAge), as well as the three MEGA composite measures: MEGA (SEM), MEGA (FA), and MEGA (weighted index) constructed for the estimation sample in Application 2 (N=448). The red line represents the distribution of chronological age in the sample (measured in two waves around age 15–19).

Figure 2: Age acceleration and early-adulthood outcomes



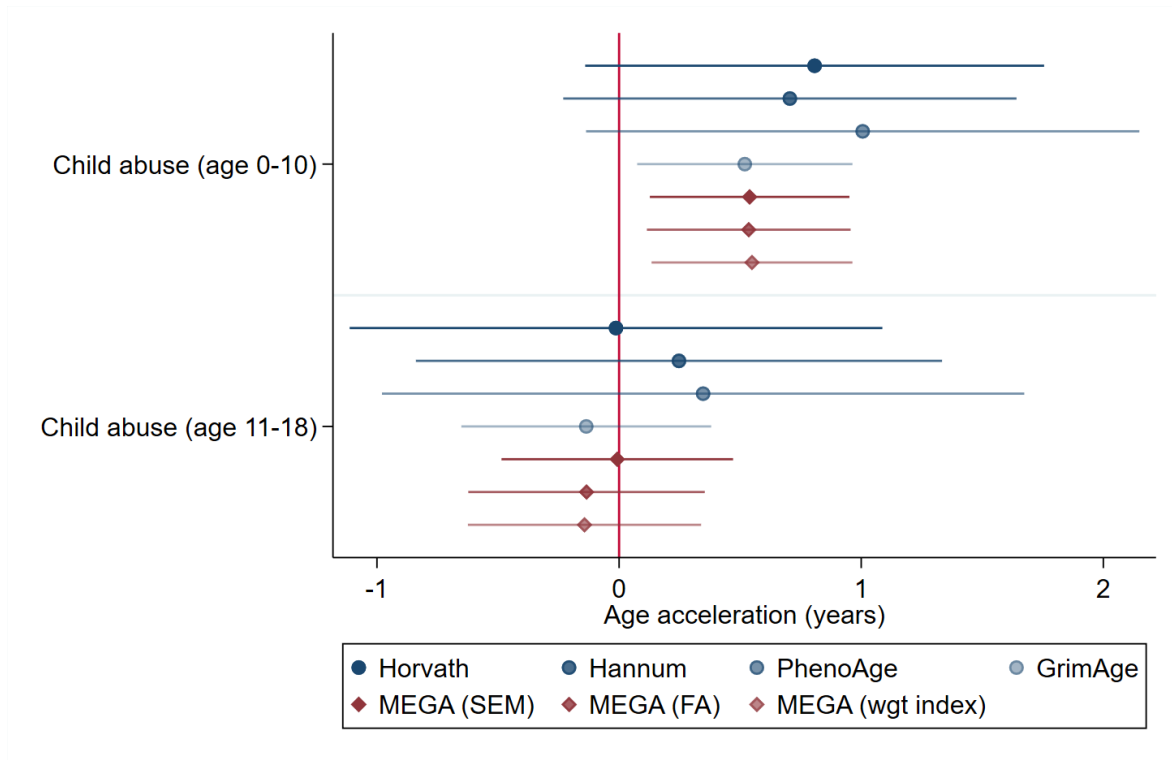
*Notes:* The figure displays point estimates for the MEGA clocks from columns (2), (4), and (6) of Tables 2 and 3. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, age, birth year, and birth order. Health outcomes (BMI, smoking and drinking) are additionally controlled for. Horizontal spikes are for 90 percent confidence intervals.

Figure 3: Cross-rater, intertemporal correlations of child cruelty and sex abuse



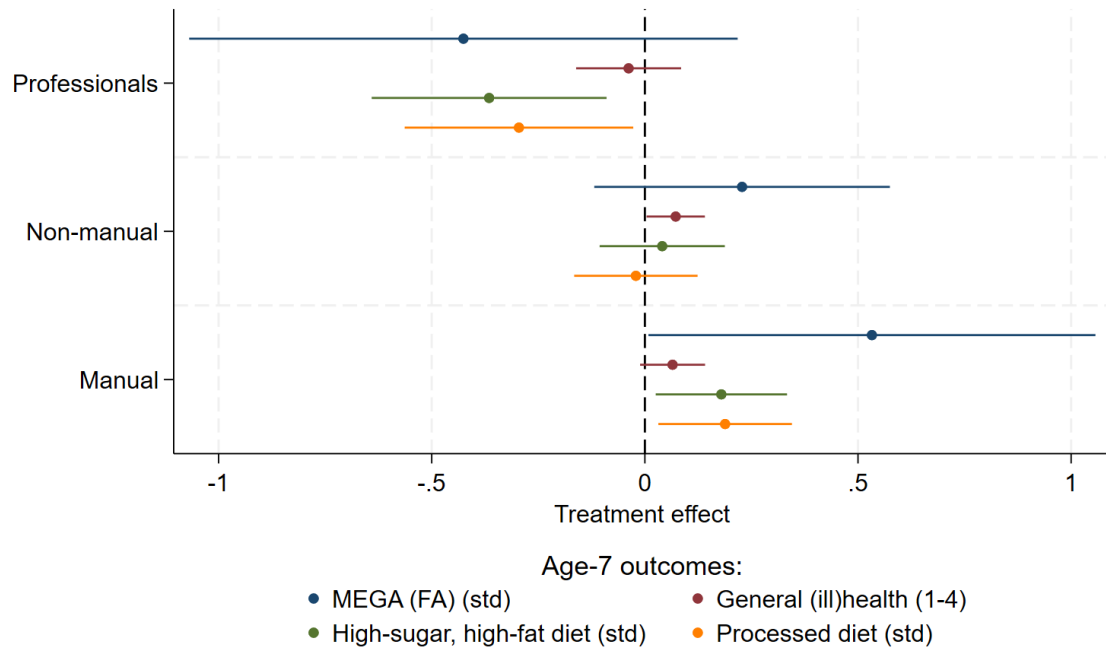
Notes: The correlation matrix is based on the sample of 3937 children in ALSPAC for whom all measures above are available.

Figure 4: Child abuse and age acceleration: individual clocks



*Notes:* The blue dots in the figure replicate point estimates from Table 5 using the individual clocks as dependent variables, while the red dots report results for the MEGA clocks. 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, age, birth year, and birth order. Horizontal spikes are for 90 percent confidence intervals.

Figure 5: The effect of delayed school entry on age-7 outcomes, by paternal social class



*Notes:* The figure displays estimated treatment effects by paternal social class, for all outcomes displayed in the legend. Estimates come from a model specification of equation (8) in which paternal social class is controlled for and where the treatment status, the month of birth and their interaction is additionally interacted for paternal social class dummies. All regressions control for the child's age and dummies for the child's gender and birth year. Horizontal spikes are for 90 percent confidence intervals.



Table 1: Descriptive statistics of chronological age and epigenetic age

	<i>N</i>	Mean	Std. Dev.	Min	Max
<b>Panel A: Application 1</b>					
Chronological age	598	17.18	1.00	14.6	19.3
Horvath	598	19.9	4.92	4.2	35.8
Hannum	598	20.47	5.00	4.6	40.3
PhenoAge	598	11.05	6.23	-11.4	34.7
GrimAge	598	34.53	2.73	26.1	43.5
MEGA clocks					
$MEGA_{SEM}$ (GrimAge)	598	0	1.85	-6.5	5.5
$MEGA_{SEM}$ (Horvath)	598	0	1.51	-5.3	4.5
$MEGA_{SEM}$ (Hannum)	598	0	2.97	-10.4	8.8
$MEGA_{SEM}$ (PhenoAge)	598	0	3.21	-11.3	9.5
$MEGA_{FA}$	598	33.09	2.56	25.6	41.2
$MEGA_{WGT}$	598	32.92	2.52	26	40.7
<b>Panel B: Application 2</b>					
Chronological age	448	17.23	0.95	14.6	19.3
Horvath	448	19.97	4.95	4.2	35.8
Hannum	448	20.62	5.03	4.6	40.3
PhenoAge	448	11.29	6.15	-8.7	34.7
GrimAge	448	34.64	2.67	26.3	42.0
MEGA clocks					
$MEGA_{SEM}$ (GrimAge)	448	21.40	1.90	14.5	26.4
$MEGA_{SEM}$ (Horvath)	448	16.45	1.46	11.1	20.3
$MEGA_{SEM}$ (Hannum)	448	33.27	2.95	22.5	41.1
$MEGA_{SEM}$ (PhenoAge)	448	34.54	3.07	23.3	42.7
$MEGA_{FA}$	448	33.78	2.55	26.3	40.8
$MEGA_{WGT}$	448	33.19	2.50	26.6	40.6
<b>Panel C: Application 3</b>					
Chronological age	597	7.46	0.14	7.1	8.8
Horvath	597	11.03	3.42	2.6	30.8
Hannum	597	9.49	4.54	-3.6	32.8
PhenoAge	597	-1.34	6.74	-23	20.8
GrimAge	597	26.1	2.48	19.6	33.2
MEGA clocks					
$MEGA_{SEM}$ (GrimAge)	597	1.86	1.07	-1.5	5.6
$MEGA_{SEM}$ (Horvath)	597	9.4	1.59	4.3	14.6
$MEGA_{SEM}$ (Hannum)	597	12.92	2.18	5.9	20
$MEGA_{SEM}$ (PhenoAge)	597	19.41	3.28	8.8	30.1
$MEGA_{FA}$	597	21.96	2.05	15.7	30.1
$MEGA_{WGT}$	597	22.9	2.02	16.3	31.6

Notes: Descriptive statistics refer to the largest estimation sample for each of the three applications. For the second and third applications, the mean values of the  $MEGA_{SEM}$  are computed as the linear prediction of the latent variable based on the coefficients from the structural model. In the first application, the  $MEGA_{SEM}$  factors are instead centered on zero, as they appear on the right-hand side of the structural equation and no linear prediction can be computed for them.

Table 2: Age acceleration and early-adulthood cognitive outcomes

	SEM		FA		Wgt Index	
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A. No University degree by 26</b>						
MEGA	0.023** (0.011)	0.022** (0.011)	0.023*** (0.008)	0.020** (0.008)	0.022*** (0.008)	0.019** (0.008)
BMI at age 15		-0.001 (0.006)		-0.000 (0.006)		-0.000 (0.006)
Smoking at age 15		0.240*** (0.073)		0.233*** (0.074)		0.231*** (0.074)
Drinking at age 15		-0.007 (0.037)		-0.006 (0.038)		-0.006 (0.038)
Observations	525	525	525	525	525	525
Adjusted R-squared			0.148	0.166	0.147	0.165
<b>Panel B. NEET at 25</b>						
MEGA	0.028** (0.011)	0.024** (0.011)	0.019** (0.008)	0.018** (0.008)	0.018** (0.008)	0.017** (0.008)
BMI at age 15		0.006 (0.005)		0.007 (0.005)		0.007 (0.005)
Smoking at age 15		0.003 (0.075)		0.002 (0.077)		-0.001 (0.077)
Drinking at age 15		-0.047 (0.036)		-0.046 (0.037)		-0.046 (0.037)
Observations	352	352	352	352	352	352
Adjusted R-squared			0.012	0.035	0.011	0.035

Notes: Standard errors in parentheses. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. 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, age, birth year, and birth order. Mean outcome values in the estimation samples are 0.24 for not having a university degree and 0.11 for being NEET. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 3: Age acceleration and early-adulthood mental health outcomes

	SEM		FA		Wgt Index	
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A. Problematic SMFQ at 25</b>						
MEGA	0.015 (0.013)	0.017 (0.013)	0.017* (0.009)	0.018** (0.009)	0.016* (0.009)	0.018** (0.009)
BMI at age 15		-0.002 (0.006)		-0.002 (0.006)		-0.002 (0.006)
Smoking at age 15		0.075 (0.082)		0.065 (0.084)		0.062 (0.084)
Drinking at age 15		-0.063 (0.041)		-0.063 (0.041)		-0.063 (0.041)
Observations	434	434	434	434	434	434
Adjusted R-squared			0.021	0.027	0.021	0.026
<b>Panel B. Diagnosed with depression by 22</b>						
MEGA	0.010 (0.013)	0.009 (0.013)	0.015* (0.009)	0.014 (0.009)	0.016* (0.009)	0.016* (0.009)
BMI at age 15		0.006 (0.007)		0.005 (0.007)		0.005 (0.007)
Smoking at age 15		0.032 (0.083)		0.022 (0.085)		0.020 (0.085)
Drinking at age 15		0.043 (0.041)		0.043 (0.042)		0.043 (0.042)
Observations	451	451	451	451	451	451
Adjusted R-squared			0.000	0.007	0.000	0.008

*Notes:* Standard errors in parentheses. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. 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, age, birth year, and birth order. Mean outcome values in the estimation samples are 0.18 for SMFQ and 0.20 for depression. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 4: Prevalence of abuse in the Application 2 estimation sample

	M (1)	P (2)	C (3)	MP (4)	CM (5)	CP (6)	CMP (7)
<b>Age 0-10</b>							
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%
<b>Age 11-18</b>							
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 for Application 2. 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 5: Child abuse and age acceleration from the MEGA clock

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

*Notes:* Standard errors in parentheses. 'SEM' stands for Structural Equation Modelling, 'FA' for Factor Analysis, and 'wgt index' for weighted index. 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, age, birth year, and birth order. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 6: Child abuse and age acceleration from the MEGA clock: disaggregation

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

*Notes:* Standard errors in parentheses. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. 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, age, birth year, and birth order. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

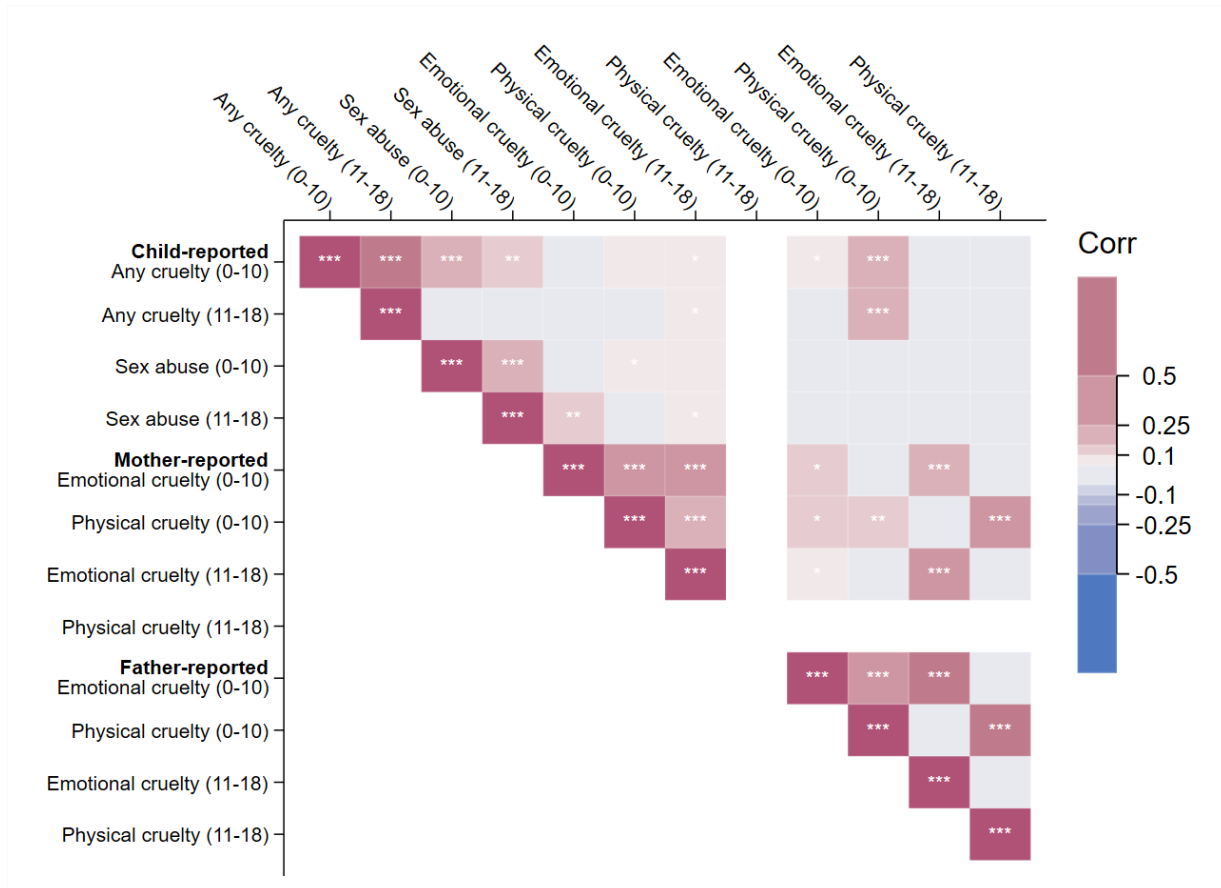
Table 7: The effect of delayed school entry on age acceleration at age 7

	(1) SEM	(2) FA	(3) Wgt Index
<b>Panel A: May - December</b>			
Treat	0.317 (0.235)	0.542* (0.321)	0.579* (0.313)
MoB	-0.142* (0.074)	-0.192* (0.104)	-0.198* (0.101)
Treat * MoB	0.056 (0.117)	-0.017 (0.142)	-0.002 (0.139)
Observations	597	597	597
Adjusted R-squared		0.116	0.130
<b>Panel B: June - November</b>			
Treat	0.328 (0.252)	0.440 (0.398)	0.448 (0.392)
MoB	-0.091 (0.109)	-0.105 (0.167)	-0.107 (0.164)
Treat * MoB	-0.134 (0.163)	-0.174 (0.226)	-0.131 (0.223)
Observations	446	446	446
Adjusted R-squared		0.083	0.094
<b>Panel C: July - October</b>			
Treat	0.891 (0.587)	1.158** (0.541)	1.114** (0.531)
MoB	-0.457 (0.365)	-0.647** (0.310)	-0.625** (0.304)
Treat * MoB	0.054 (0.341)	0.376 (0.444)	0.448 (0.436)
Observations	307	307	307
Adjusted R-squared		0.069	0.076

Notes: Robust standard errors in parentheses. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. All regressions control for the child’s age and dummies for the child’s gender and birth year. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

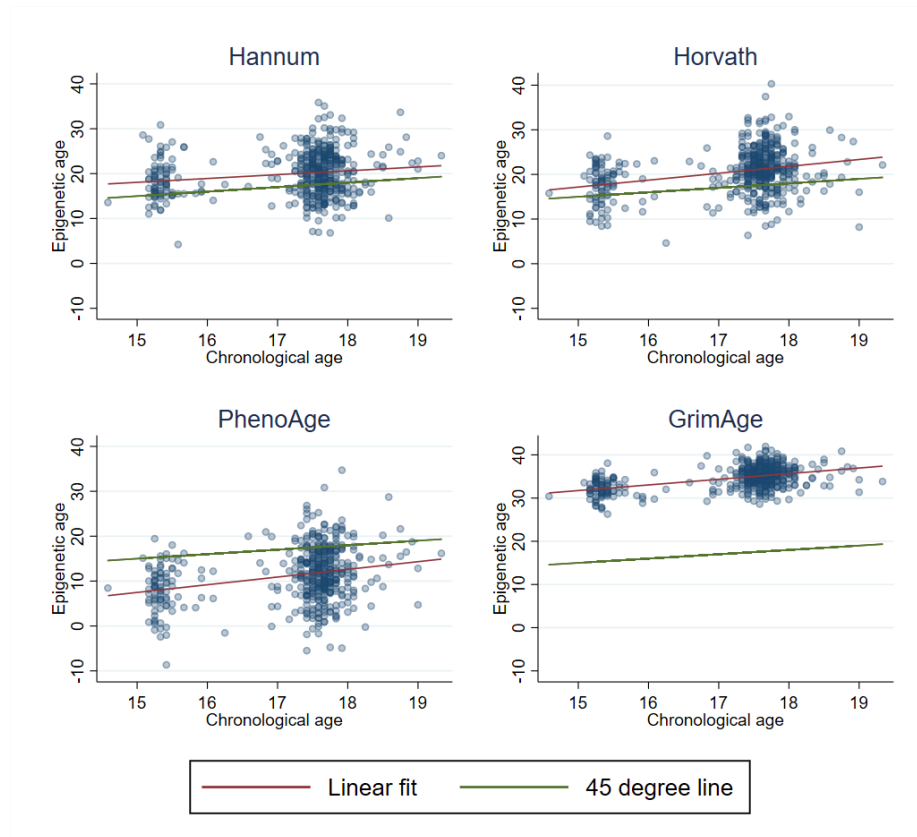
## Appendix A: Supplementary Figures and Tables

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



Notes: The graph replicates Figure 3 in the estimation sample of 448 participants from Application 2. The row and column corresponding to mother-reported physical cruelty between ages 11 and 18 are blank as there are no cases of mother-reported physical cruelty between child ages 11 and 18 in the estimation sample.

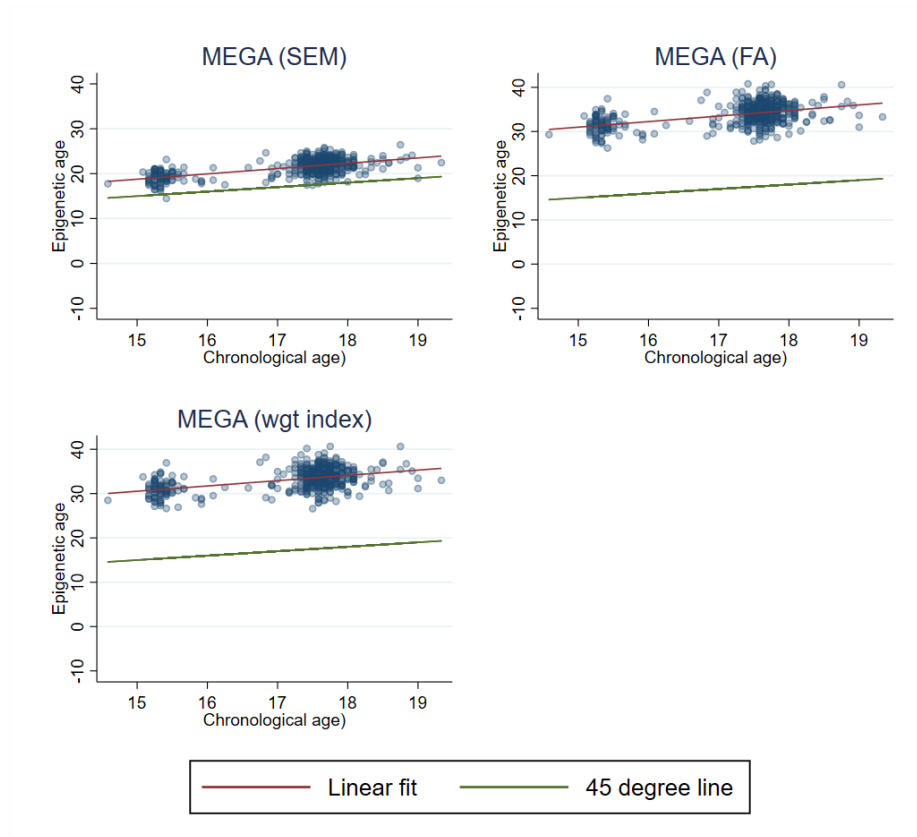
Figure A2: Chronological age and epigenetic age in four epigenetic clocks



*Notes:* The graphs are scatterplots of epigenetic age against chronological age in the estimation sample of Application 2 (448 observations). Linear fits of the bivariate association are in red, while the 45-degree line is in green.

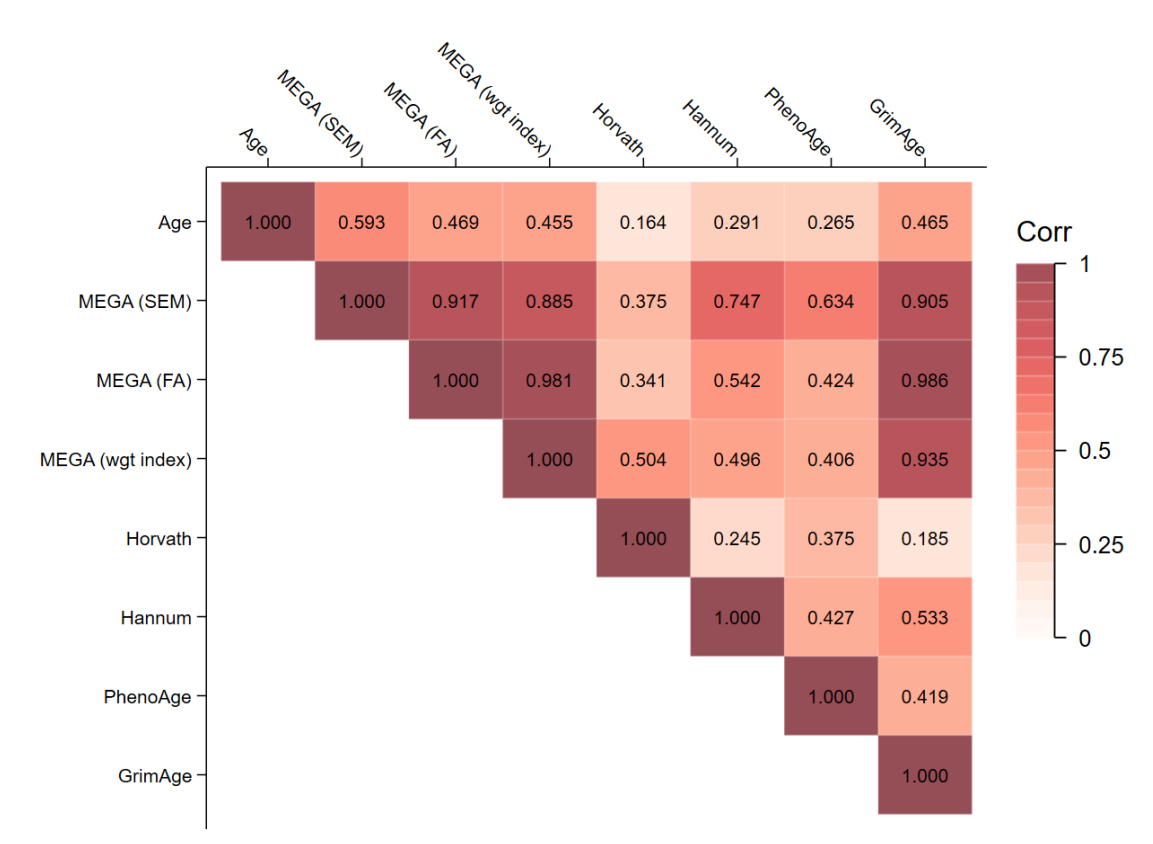


Figure A3: Chronological age and epigenetic age in the MEGA clocks



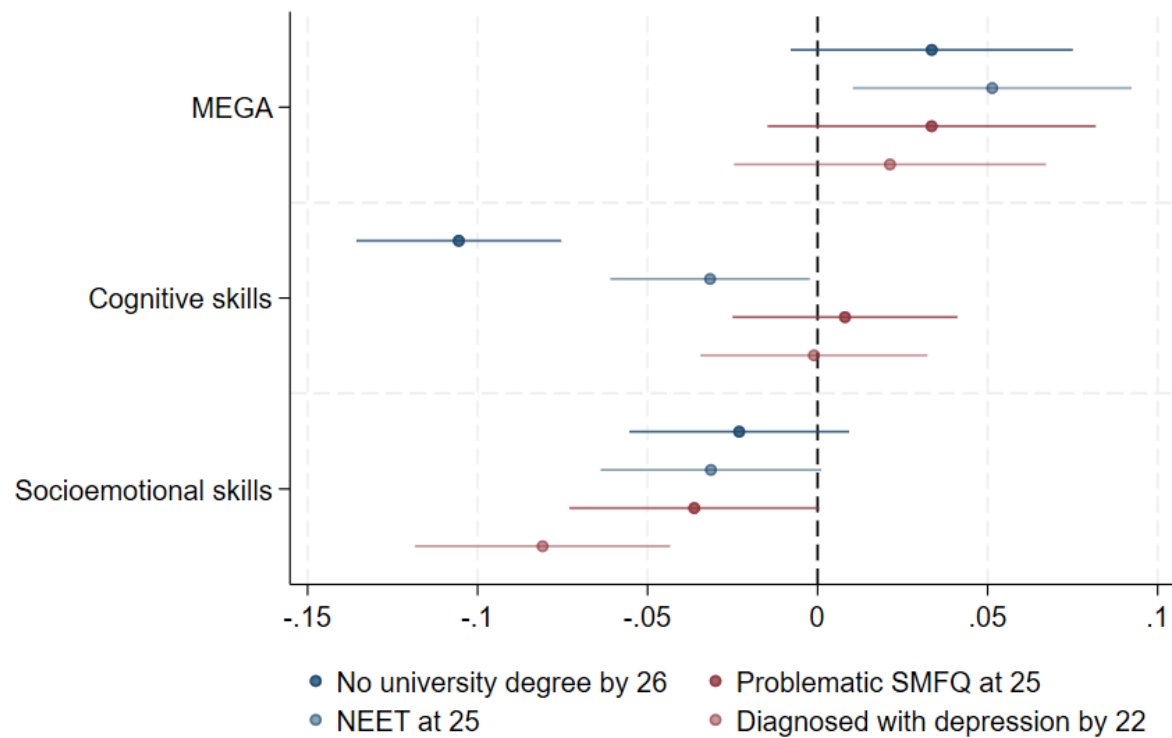
*Notes:* The graphs are scatterplots of epigenetic age against chronological age in the estimation sample of Application 2 (448 observations). ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. Linear fits of the bivariate association are in red, while the 45-degree line is in green.

Figure A4: Correlation coefficients across clocks



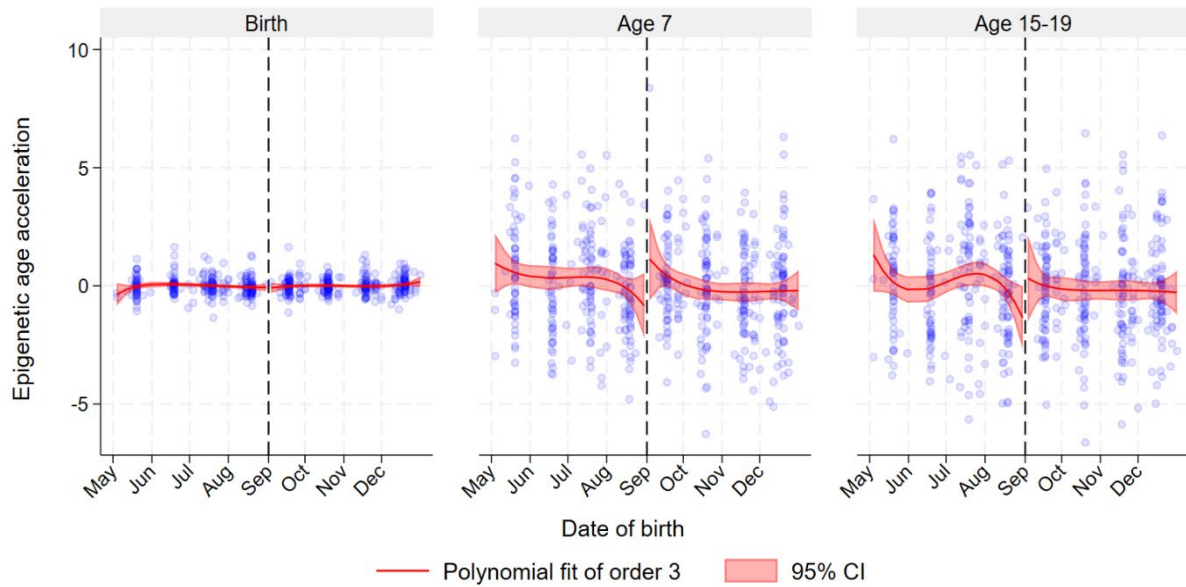
*Notes:* The correlation matrix refers to the estimation sample from Application 2 (448 observations). ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index.

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



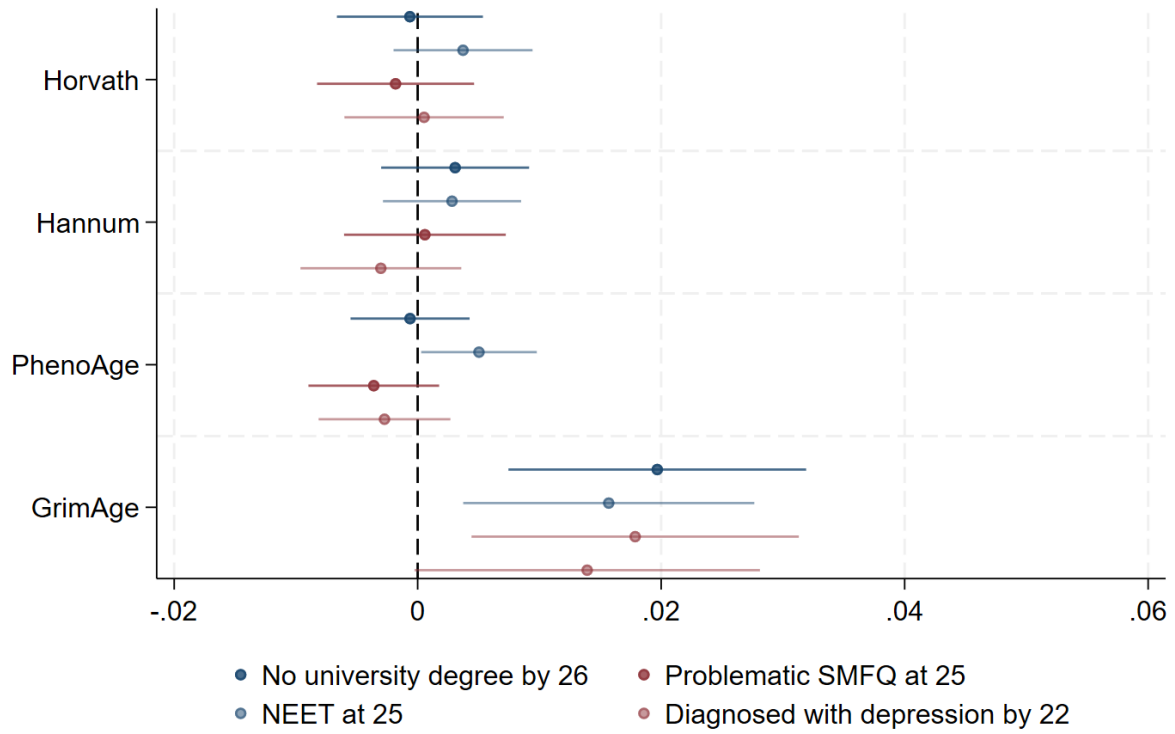
*Notes:* The figure replicates results in column (1) of Tables 2 and 3, additionally controlling for a cognitive and a socio-emotional skills factor. The MEGA clock and the cognitive and socio-emotional factors all come from a Structural Equation Model and are standardized in the figure, to enhance comparability. 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. Horizontal spikes are for 90 percent confidence intervals.

Figure A6: Age acceleration by date of birth ( $MEGA_{FA}$ )



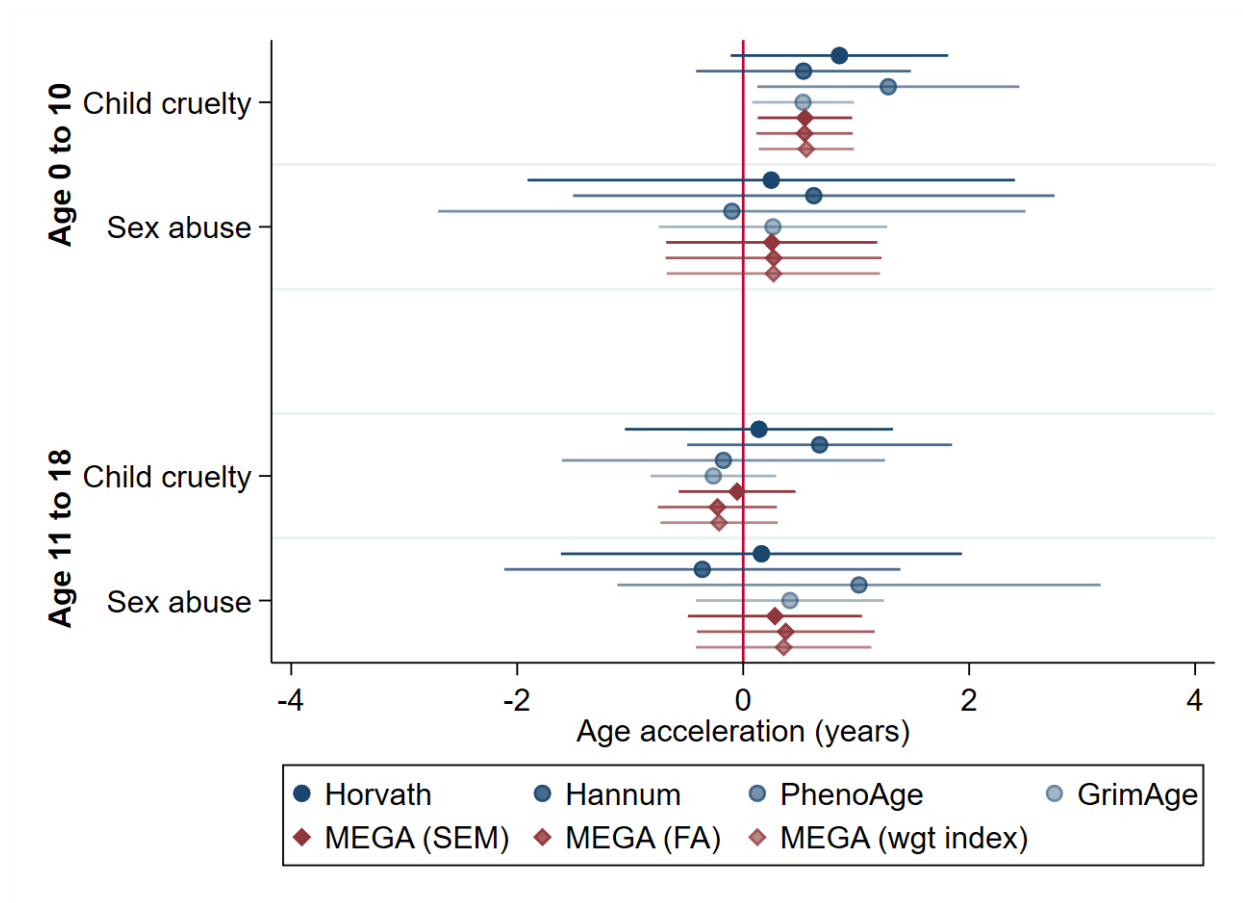
*Notes:* Scatter plot of epigenetic age acceleration, observed at the child's birth, age 7 and age 15-19, against date of birth. Epigenetic age acceleration is here computed as residuals from regressions of the MEGA (factor analysis) clock on chronological age, run for each of the three time points separately. To achieve greater granularity than we would by simply using month of birth, date of birth is here derived as the mid-point of a plausible date-of-birth interval, given by cross-referencing the child's birth month and birth year with the month and year of completion of the first child-based questionnaire, and the child's age in weeks at the time of completion. The vertical dashed line marks the 1<sup>st</sup> of September, the birthdate cut-off from which children's school entry is delayed by one academic year.

Figure A7: Age acceleration and early-adulthood cognitive outcomes: individual clocks



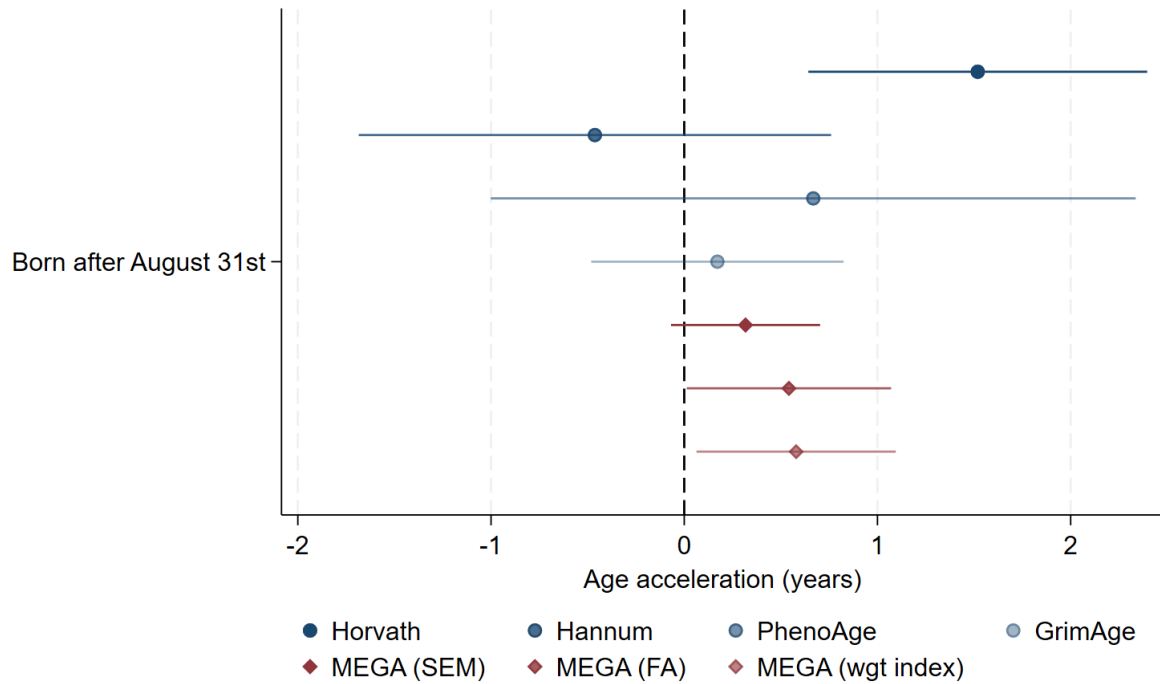
*Notes:* The figure replicates results from the even-numbered columns of Tables 2 and 3, using individual clocks as dependent variables instead of the MEGA clock. All regressions control for mother's age at birth of the study child and binary indicators for mother's education, father's social class, and the child's gender, birth year, birth order, and age 15 health outcomes (BMI, smoking, and drinking). Horizontal spikes are for 90 percent confidence intervals.

Figure A8: Child abuse (disaggregated) and age acceleration: individual clocks



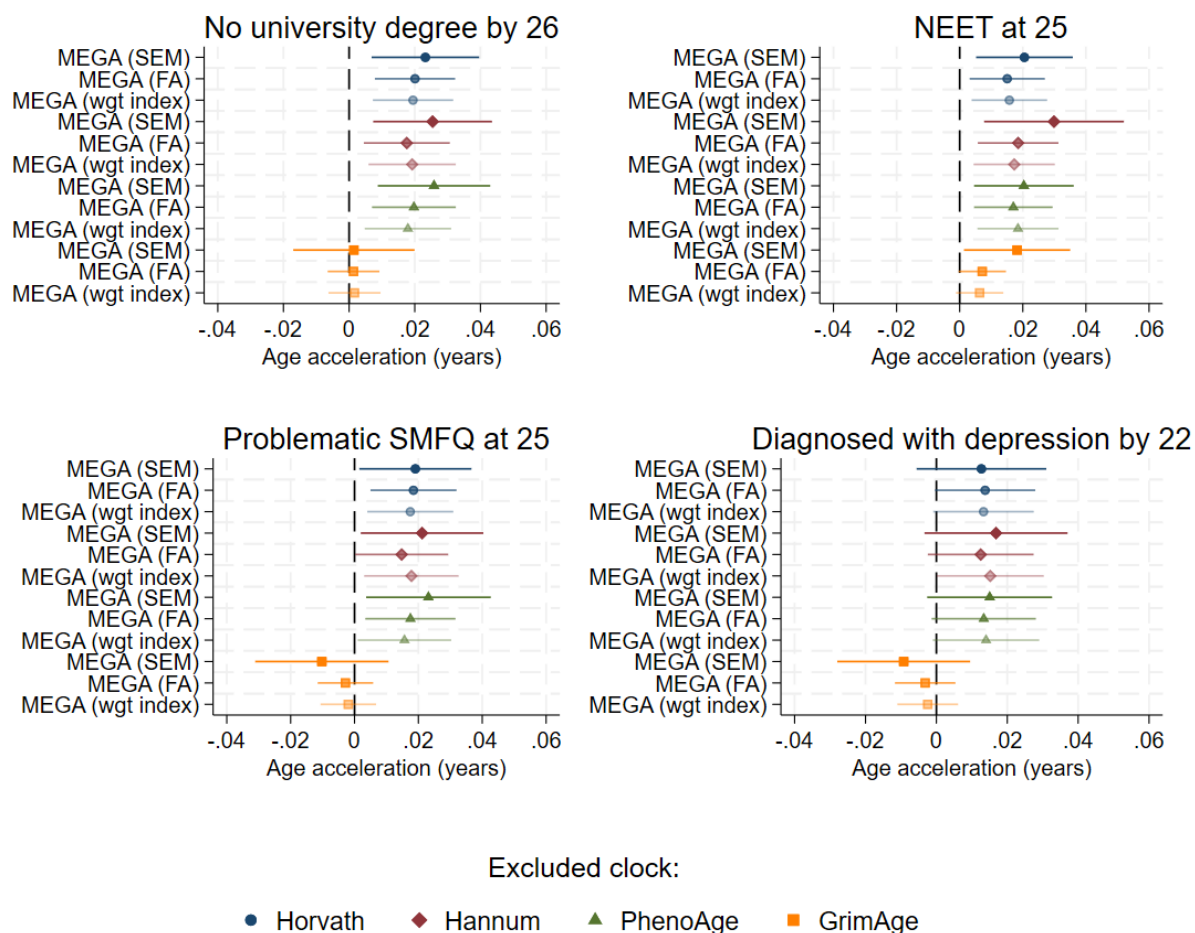
*Notes:* The blue dots in the figure replicate point estimates from Table 6 using the individual clocks as dependent variables, while the red dots report results for the MEGA clocks. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. All regressions 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. Horizontal spikes are for 90 percent confidence intervals.

Figure A9: The effect of delayed school entry on age acceleration at age 7: individual clocks



*Notes:* The blue dots in the figure replicate point estimates of the treatment dummy from Panel A of Table 7 using the individual clocks as dependent variables, while the red dots report results for the MEGA clocks. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. All regressions control for the child’s gender, age and birth year. Horizontal spikes are for 90 percent confidence intervals.

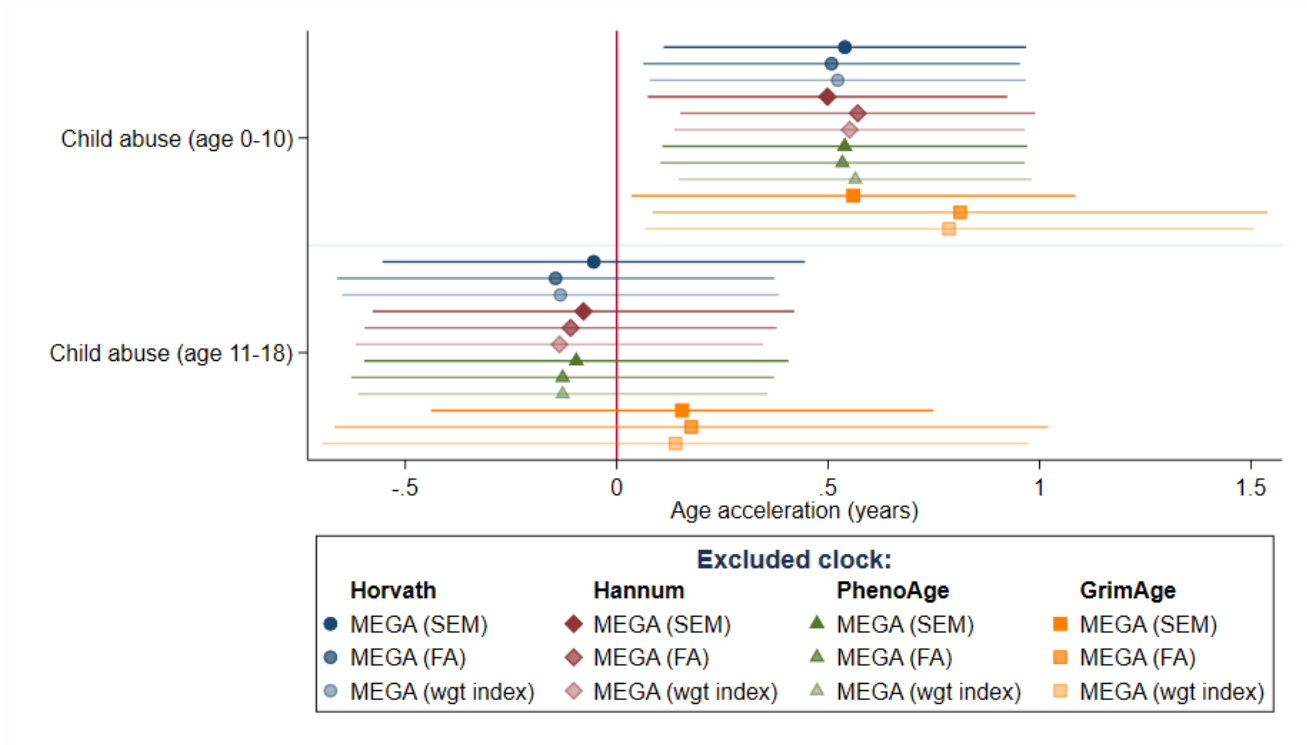
Figure A10: Age acceleration and early-adulthood outcomes: leave-one-out MEGA



*Notes:* The figure replicates results from even-numbered columns in Tables 2 and 3 using leave-one-out versions of the MEGA clock. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. 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. Horizontal spikes are for 90 percent confidence intervals.

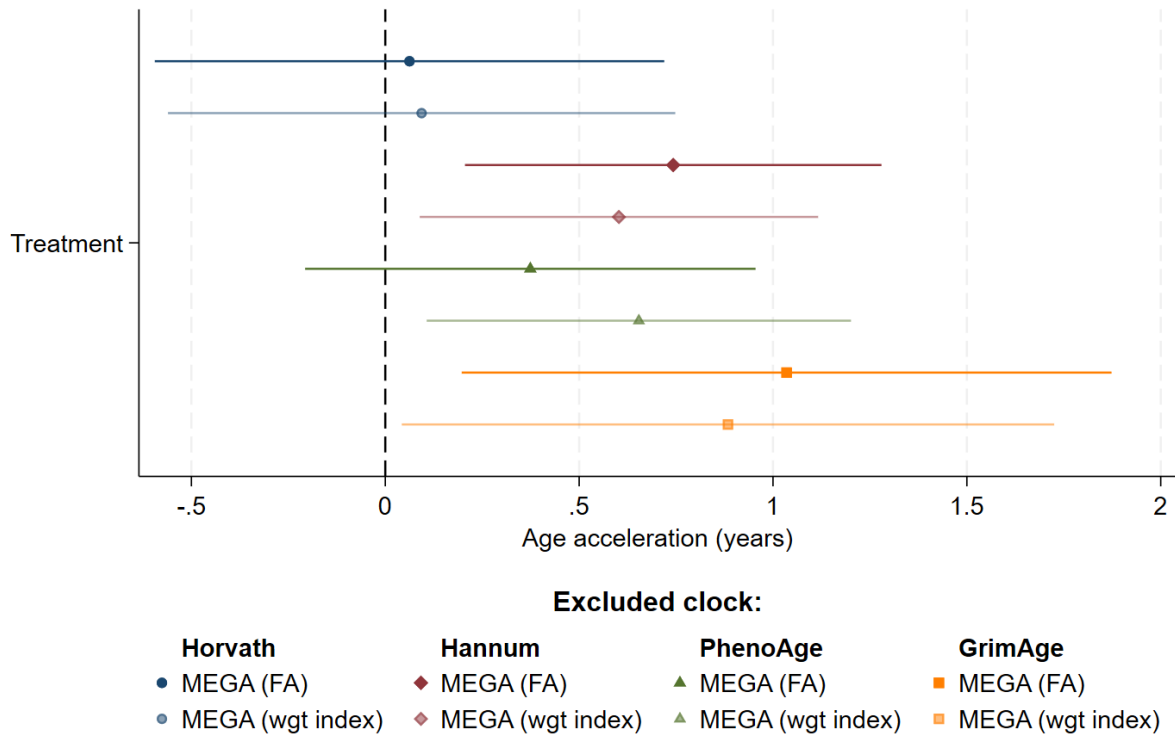


Figure A11: Child abuse and age acceleration: leave-one-out MEGA



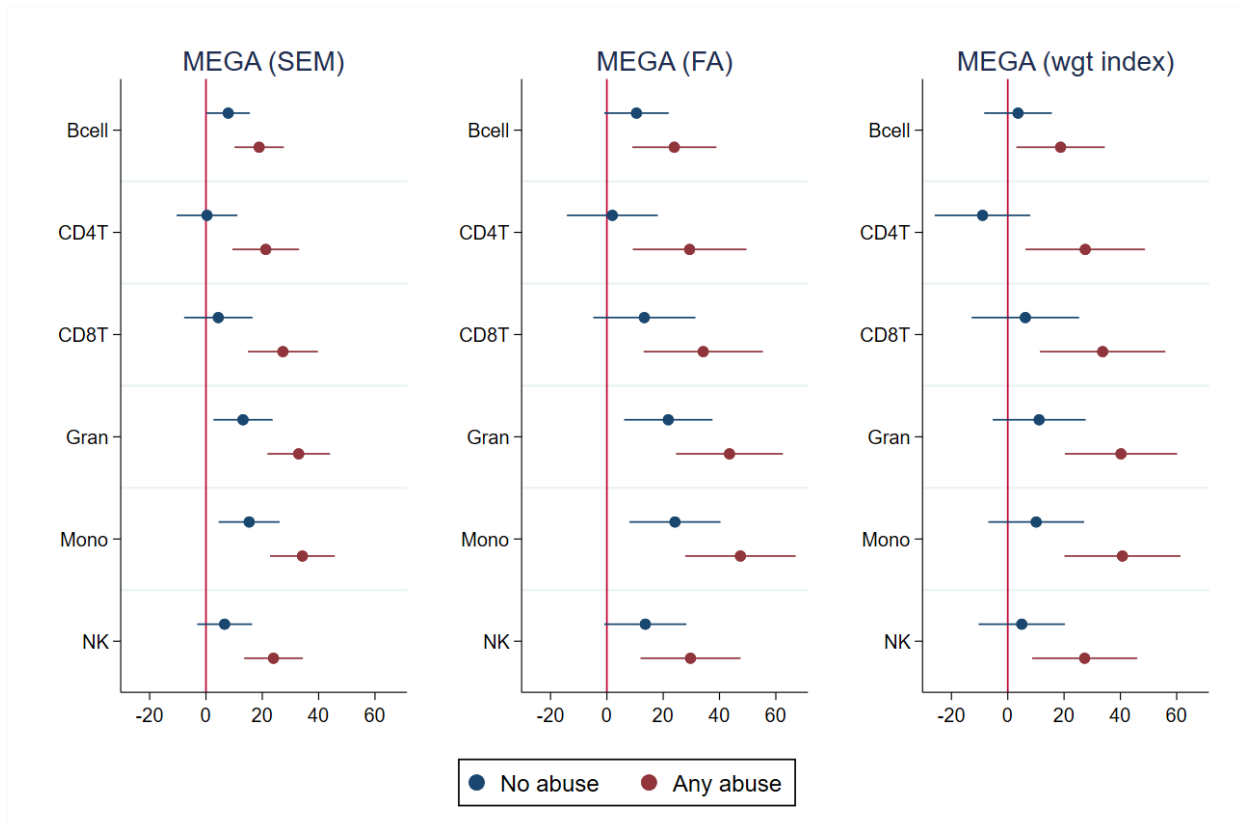
*Notes:* The figure replicates results from Table 5 using leave-one-out versions of the MEGA clock. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. All regressions control for mother’s age at birth of the study child and binary indicators for mother’s education, father’s social class, and the child’s gender and birth order. The child’s birth year is also included as a control, except for the SEM model excluding the Hannum clock, where it is excluded in order to achieve model convergence. Horizontal spikes are for 90 percent confidence intervals.

Figure A12: Delayed school entry and age acceleration at age 7: leave-one-out MEGA



*Notes:* The figure replicates results from Panel A of Table 7 using leave-one-out versions of the MEGA clock. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. Dots are for the estimated coefficients of the treatment status variable. All regressions control for the child’s age and binary indicators for the child’s gender and birth year. Estimates relying on MEGA (SEM) are not illustrated here due to a lack of model convergence. Horizontal spikes are for 90 percent confidence intervals.

Figure A13: 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. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. Spikes are for 90 percent 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 samples

	Sample from Application:				Differences		
	No. 1 (1)	No. 2 (2)	No. 3 (3)	Full (4)	(4)-(1)	(4)-(2)	(4)-(3)
Female	0.594 [0.492] <i>598</i>	0.618 [0.486] <i>448</i>	0.499 [0.500] <i>597</i>	0.489 [0.500] <i>14997</i>	-0.105*** (0.021)	-0.129*** (0.024)	-0.011 (0.021)
Age	17.177 [0.997] <i>598</i>	17.228 [0.951] <i>448</i>	7.457 [0.138] <i>597</i>	17.123 [1.042] <i>925</i>	-0.054 (0.054)	-0.105 (0.058)	-0.003 (0.008)
Born in 1992	0.674 [0.469] <i>598</i>	0.685 [0.465] <i>448</i>	0.595 [0.491] <i>597</i>	0.563 [0.496] <i>15468</i>	-0.111*** (0.021)	-0.122*** (0.024)	-0.030 (0.021)
First-born	0.490 [0.500] <i>598</i>	0.500 [0.501] <i>448</i>	0.482 [0.500] <i>593</i>	0.440 [0.496] <i>13320</i>	-0.050* (0.021)	-0.060* (0.024)	-0.043* (0.021)
Mother's age at birth	29.843 [4.319] <i>598</i>	29.946 [4.307] <i>448</i>	29.668 [4.462] <i>597</i>	27.989 [4.969] <i>14023</i>	-1.854*** (0.206)	-1.957*** (0.238)	-1.677*** (0.207)
Mother's education (ref: Lower-secondary)							
<i>Upper-secondary</i>	0.296 [0.457] <i>598</i>	0.292 [0.455] <i>448</i>	0.295 [0.456] <i>597</i>	0.182 [0.386] <i>15612</i>	-0.114*** (0.016)	-0.111*** (0.019)	-0.113*** (0.016)
<i>Post-secondary</i>	0.258 [0.438] <i>598</i>	0.277 [0.448] <i>448</i>	0.238 [0.426] <i>597</i>	0.104 [0.305] <i>15612</i>	-0.154*** (0.013)	-0.173*** (0.015)	-0.134*** (0.013)
Father's social class (ref.: Professionals)							
<i>Non-manual</i>	0.467 [0.499] <i>598</i>	0.482 [0.500] <i>448</i>	0.447 [0.498] <i>597</i>	0.316 [0.465] <i>15584</i>	-0.151*** (0.019)	-0.166*** (0.022)	-0.131*** (0.019)
<i>Manual</i>	0.289 [0.454] <i>598</i>	0.259 [0.439] <i>448</i>	0.317 [0.466] <i>597</i>	0.311 [0.463] <i>15584</i>	0.022 (0.019)	0.052* (0.022)	-0.006 (0.019)

Notes: The table plots means of covariates and their differences across the estimation samples of the three empirical applications 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<sub>FA</sub>*

	Application 1 (age 15-19, N=598)		Application 2 (age 15-19, N=448)		Application 3 (age 7, N=597)	
	Factor loadings	Uniqueness	Factor loadings	Uniqueness	Factor loadings	Uniqueness
Horvath	0.422	0.822	0.410	0.832	0.500	0.750
Hannum	0.659	0.565	0.662	0.562	0.603	0.636
PhenoAge	0.652	0.575	0.633	0.599	0.701	0.509
GrimAge	0.623	0.612	0.639	0.592	0.527	0.722

*Notes:* Values are loadings obtained from factor analysis of the four epigenetic clocks, run on the largest estimation sample from each application. All applications support a unifactoral model, with only one factor displaying eigenvalue greater than one. For Application 1, this is the sample of 598 observations for which all controls, all clocks and at least one of the four adult outcomes are available. For Application 2, this is the estimation sample of 448 observations. Last, for Application 3, we rely on the May-December estimation sample of 597 observations at child age 7.

Table A4: Child abuse and age acceleration from the MEGA clock: sensitivity to the rater of abuse

	M (1)	P (2)	C (3)	MP (4)	CM (5)	CP (6)	CMP (7)
<b>A. SEM</b>							
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>B. FA</b>							
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
<b>C. Wgt index</b>							
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 Structural Equation Modelling in panel A, with factor analysis 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 A5: The effect of delayed school entry on age acceleration at age 15-19

	(1) SEM	(2) FA	(3) Wgt Index
<b>Panel A: May - December</b>			
Treat	0.132 (0.343)	0.108 (0.356)	0.174 (0.353)
MoB	0.009 (0.114)	-0.016 (0.118)	-0.047 (0.117)
Treat * MoB	-0.088 (0.153)	-0.087 (0.159)	-0.044 (0.157)
Observations	600	600	600
Adjusted R-squared		0.291	0.280
<b>Panel B: June - November</b>			
Treat	.	0.146 (0.431)	0.268 (0.427)
MoB	.	-0.003 (0.188)	-0.039 (0.187)
Treat * MoB	.	-0.124 (0.254)	-0.108 (0.252)
Observations		454	454
Adjusted R-squared		0.302	0.286
<b>Panel C: July - October</b>			
Treat	.	0.234 (0.597)	0.356 (0.592)
MoB	.	-0.745** (0.344)	-0.820** (0.341)
Treat * MoB	.	0.739 (0.490)	0.803* (0.486)
Observations		317	317
Adjusted R-squared		0.314	0.299

Notes: Standard errors in parentheses. 'SEM' stands for Structural Equation Modelling, 'FA' for Factor Analysis, and 'wgt index' for weighted index. SEM estimates are not provided in Panel B and C since the estimating algorithm did not autonomously converge, likely due to the small sample size. Omitting some controls or changing the starting point of the algorithm would allow for convergence but at the expense of clarity and comparability. All regressions control for the child's age and dummies for the child's gender and birth year. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .



Table A6: Child abuse and age acceleration from the MEGA clock (controlling for cell type counts)

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

*Notes:* Standard errors in parentheses. ‘SEM’ stands for Structural Equation Modelling, ‘FA’ for Factor Analysis, and ‘wgt index’ for weighted index. SEM estimates are not provided since the estimating algorithm did not reliably converge, likely due to the small sample size. 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. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .