Aggregating epigenetic clocks to study human capital formation

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

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

19th Canazei Winter School

January 7th, 2026

Why epigenetics?

- Many policy-relevant outcomes are shaped early, but we rarely observe the **biological embedding** of early-life conditions
- **Epigenetics**, the study of changes in gene expression without alterations to the DNA sequence, is one of the ways in which the environment (**nurture**) can affect human biology (**nature**)
    - Environmental exposures, e.g. SES, stress, leave long lasting biological traces that influence behaviors and economic outcomes (Neu, 2023)

Why epigenetics?

- Many policy-relevant outcomes are shaped early, but we rarely observe the **biological embedding** of early-life conditions
- **Epigenetics**, the study of changes in gene expression without alterations to the DNA sequence, is one of the ways in which the environment (**nurture**) can affect human biology (**nature**)
  - Environmental exposures, e.g. SES, stress, leave long lasting biological traces that influence behaviors and economic outcomes (Neu, 2023)

Why epigenetics?

- Potential of epigenetics in economics:
  - the biological footprint of life experiences ('latent health')
  - useful to study mechanisms linking early shocks to adult outcomes
  - measurable *before* education, earnings, disease, etc.
- Dominant summary measure: **epigenetic clocks** (measured in years of biological aging)
- This presentation:
  - A paper ("Aggregating epigenetic clocks to study human capital formation")
  - A digression on epigenetic clocks

Why epigenetics?

- Potential of epigenetics in economics:
    - the biological footprint of life experiences ('latent health')
    - useful to study mechanisms linking early shocks to adult outcomes
    - measurable *before* education, earnings, disease, etc.
- Dominant summary measure: **epigenetic clocks** (measured in years of biological aging)
- This presentation:
    - A paper ("Aggregating epigenetic clocks to study human capital formation")
    - A digression on epigenetic clocks

Why epigenetics?

- Potential of epigenetics in economics:
  - the biological footprint of life experiences ('latent health')
  - useful to study mechanisms linking early shocks to adult outcomes
  - measurable *before* education, earnings, disease, etc.
- Dominant summary measure: **epigenetic clocks** (measured in years of biological aging)
- This presentation:
  - A paper ("Aggregating epigenetic clocks to study human capital formation")
  - A digression on epigenetic clocks

The paper

Methodological contribution:

- Three aggregation strategies to construct the **Multi EpiGenetic Age (MEGA)** clock, from 4 widely-used clocks
- Reduces measurement error and improves reliability of epigenetic aging measures.

Empirical applications: three, based on ALSPAC data

- **Consequences**: In adolescence, faster epigenetic aging predicts worse early-adulthood education, employment, and mental health *even controlling for adolescent health behaviors*
- **Determinants**:
  - *Child abuse*: exposure to abuse before adolescence correlates with a 0.5-years increase in biological aging.
  - *School entry age*: delaying school entry increases epigenetic aging for low-SES children.

The paper

Methodological contribution:

- Three aggregation strategies to construct the **Multi EpiGenetic Age (MEGA)** clock, from 4 widely-used clocks
- Reduces measurement error and improves reliability of epigenetic aging measures.

Empirical applications: three, based on ALSPAC data

- **Consequences**: In adolescence, faster epigenetic aging predicts worse early-adulthood education, employment, and mental health *even controlling for adolescent health behaviors*
- **Determinants**:
  - *Child abuse*: exposure to abuse before adolescence correlates with a 0.5-years increase in biological aging.
  - *School entry age*: delaying school entry increases epigenetic aging for low-SES children.

The paper

Methodological contribution:

- Three aggregation strategies to construct the **Multi EpiGenetic Age (MEGA)** clock, from 4 widely-used clocks
- Reduces measurement error and improves reliability of epigenetic aging measures.

Empirical applications: three, based on ALSPAC data

- **Consequences**: In adolescence, faster epigenetic aging predicts worse early-adulthood education, employment, and mental health *even controlling for adolescent health behaviors*
- **Determinants**:
  - *Child abuse*: exposure to abuse before adolescence correlates with a 0.5-years increase in biological aging.
  - *School entry age*: delaying school entry increases epigenetic aging for low-SES children.

The paper

Methodological contribution:

- Three aggregation strategies to construct the **Multi EpiGenetic Age (MEGA)** clock, from 4 widely-used clocks
- Reduces measurement error and improves reliability of epigenetic aging measures.

Empirical applications: three, based on ALSPAC data

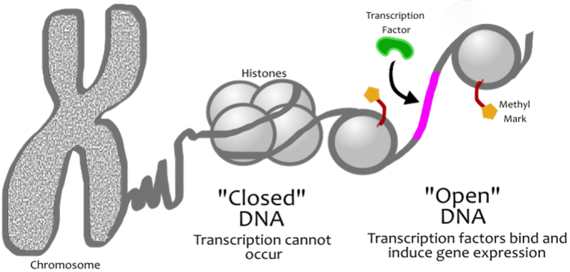
- **Consequences**: In adolescence, faster epigenetic aging predicts worse early-adulthood education, employment, and mental health *even controlling for adolescent health behaviors*
- **Determinants**:
  - *Child abuse*: exposure to abuse before adolescence correlates with a 0.5-years increase in biological aging.
  - *School entry age*: delaying school entry increases epigenetic aging for low-SES children.

Contribution to the literature

- Measurement of biological aging (e.g. Horvath, 2013; Hannum et al., 2013; Levine et al., 2018; Lu et al., 2019; Martinez et al., 2024)
- Early life adversity and epigenetic aging (Lawn et al., 2018; Cecil et al., 2020; Marini et al., 2020; Lussier et al., 2023; Rubens et al., 2023)
- Consequences of accelerated epigenetic aging (e.g. Horvath, 2013; Perna et al., 2016; Dugué et al., 2018; McCartney et al., 2018; Lu et al., 2019; Faul et al., 2023; Davillas and Jones, 2024; Niccodemi et al., 2022; Raffington et al., 2023a; Raffington et al., 2023b)
- Adverse health consequences of child abuse (Currie and Spatz Widom, 2010; Fletcher, 2009; Henkaus, 2022; Suglia et al., 2014)

The digression: A primer on Epigenetics

The mechanisms of gene activity regulation **independent of DNA sequence** that determine **'turn-on' and 'turn-off' of the genes**

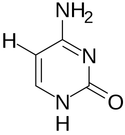


Epigenetic modifications

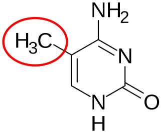
- **Epigenetic reprogramming**: (almost) clean ‘epigenetic slate’ at birth
- Epigenetic modifications can occur as a result of **environmental exposures**
    - E.g. stress, disease, smoking, diet
- Most common kinds of epigenetic modifications:
    - **DNA Methylation**
    - Histone modification

DNA Methylation

DNA methylation (DNAm): process by which a methyl group is added to a CpG dineuclotide



cytosine



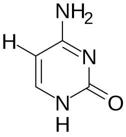
methylated cytosine

At each CpG site, DNAm is expressed as the **proportion of methylated cells** out of the observed total:

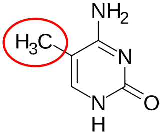
$$\beta = \frac{M}{M+U}, \qquad \beta \in [0,1]$$

DNA Methylation

DNA methylation (DNAm): process by which a methyl group is added to a CpG dineuclotide



cytosine



methylated cytosine

At each CpG site, DNAm is expressed as the **proportion of methylated cells** out of the observed total:

$$\beta = \frac{M}{M+U}, \qquad \beta \in [0,1]$$

DNAm: interpretation problems

Some caveats:

- Epigenetic data are tissue-specific (and so are clocks!)
- Cell type heterogeneity
- Underpinning biological mechanisms are still largely undiscovered
- Issues of reverse causality

Epigenetic clocks

- Aging isn't naturally a linear process, though we typically model it as such
  - Often focus on **chronological age**, ignoring **biological aging**
- DNAm transforms with age: global shifts over time
- **Epigenetic clocks**: 'scores' of DNAm, based on predictiveness of chronological age and physical decline
  - Mostly based on supervised ML (penalized lasso or elastic net regressions)
  - Epigenetic clocks are **accurate predictors of chronological age**
  - Age Acceleration (AA): residuals of '`reg DNAmAge Age`'
  - AA is a predictor of **age-related diseases and mortality**

Epigenetic clocks

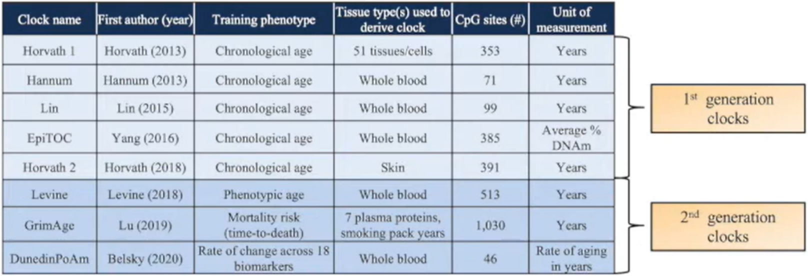
- Aging isn't naturally a linear process, though we typically model it as such
  - Often focus on **chronological age**, ignoring **biological aging**
- DNAm transforms with age: global shifts over time
- **Epigenetic clocks**: 'scores' of DNAm, based on predictiveness of chronological age and physical decline
  - Mostly based on supervised ML (penalized lasso or elastic net regressions)
  - Epigenetic clocks are **accurate predictors of chronological age**
  - Age Acceleration (AA): residuals of '`reg DNAmAge Age`'
  - AA is a predictor of **age-related diseases and mortality**

Epigenetic clocks

- Aging isn't naturally a linear process, though we typically model it as such
    - Often focus on **chronological age**, ignoring **biological aging**
- DNAm transforms with age: global shifts over time
- **Epigenetic clocks**: 'scores' of DNAm, based on predictiveness of chronological age and physical decline
    - Mostly based on supervised ML (penalized lasso or elastic net regressions)
    - Epigenetic clocks are **accurate predictors of chronological age**
    - Age Acceleration (AA): residuals of '`reg DNAmAge Age`'
    - AA is a predictor of **age-related diseases and mortality**

Epigenetic clocks: generations

Different ‘generations’ of clocks:



Clock name	First author (year)	Training phenotype	Tissue type(s) used to derive clock	CpG sites (#)	Unit of measurement
Horvath 1	Horvath (2013)	Chronological age	51 tissues/cells	353	Years
Hannum	Hannum (2013)	Chronological age	Whole blood	71	Years
Lin	Lin (2015)	Chronological age	Whole blood	99	Years
EpiTOC	Yang (2016)	Chronological age	Whole blood	385	Average % DNAm
Horvath 2	Horvath (2018)	Chronological age	Skin	391	Years
Levine	Levine (2018)	Phenotypic age	Whole blood	513	Years
GrimAge	Lu (2019)	Mortality risk (time-to-death)	7 plasma proteins, smoking pack years	1,030	Years
DunedinPoAm	Belsky (2020)	Rate of change across 18 biomarkers	Whole blood	46	Rate of aging in years

1^{st} generation clocks

2^{nd} generation clocks

Source: Lauren Schmitz

Little overlap in the CpGs identified by the different clocks

DNAm: datasets

Epigenetic data are becoming increasingly available to social scientists

Some studies including DNAm data:

- **US**: HRS, AddHealth, Framingham Heart Study, Midlife in the United States
- **UK**: ALSPAC, Understanding Society, Generation Scotland, NICOLA, E-Risk Twin Study
- **Netherlands**: Generation R Study, Netherlands Twin Registry
- **Other countries**: TILDA (Ireland), Dunedin Study (New Zealand), PROGRESS (Mexico), etc.

The MEGA clock

We develop the **Multi EpiGenetic Age (MEGA)** clock

- Combine 4 clocks (here: Horvath, Hannum, PhenoAge, GrimAge).
- Compare methods: weighted index, Factor, SEM

Why aggregating clocks?

- Each clock is a noisy proxy for latent biological aging
- Classical measurement error attenuates coefficients and inflates standard errors
- Aggregation extracts the shared aging signal and **downweights redundant noise**

Practical advantage: no need to be a biologist! Clocks are increasingly available in survey data

The MEGA clock

We develop the **Multi EpiGenetic Age (MEGA)** clock

- Combine 4 clocks (here: Horvath, Hannum, PhenoAge, GrimAge).
- Compare methods: weighted index, Factor, SEM

Why aggregating clocks?

- Each clock is a noisy proxy for latent biological aging
- Classical measurement error attenuates coefficients and inflates standard errors
- Aggregation extracts the shared aging signal and **downweights redundant noise**

Practical advantage: no need to be a biologist! Clocks are increasingly available in survey data

The MEGA clock

We develop the **Multi EpiGenetic Age (MEGA)** clock

- Combine 4 clocks (here: Horvath, Hannum, PhenoAge, GrimAge).
- Compare methods: weighted index, Factor, SEM

Why aggregating clocks?

- Each clock is a noisy proxy for latent biological aging
- Classical measurement error attenuates coefficients and inflates standard errors
- Aggregation extracts the shared aging signal and **downweights redundant noise**

Practical advantage: no need to be a biologist! Clocks are increasingly available in survey data

The MEGA clock

Let $C_1, \ldots, C_K$ be K epigenetic clocks and M their var-cov matrix.

(1) Weighted index (Anderson-style)

$$MEGA_{WGT} = \frac{\sum_{k=1}^{K} w_{k,1} C_k}{\sum_{k=1}^{K} w_{k,1}}, \quad w = M^{-1} I_K.$$

(2) Factor analysis FA

One common factor explains clock co-movement; wgt use factor loadings:

$$\mathrm{MEGA}_{\mathrm{FA}} = \frac{\sum_{k=1}^{K} u_{k,1} C_k}{\sum_{k=1}^{K} u_{k,1}}, \quad u = M^{-1} L.$$

(3) Structural Equation Model (SEM)

Measurement equations $C_k = \lambda_k EA^* + \varepsilon_k$ and behavioral equation $EA^* = X'\gamma + \omega$ estimated jointly.

Data

The Avon Longitudinal Study of Parents and Children (ALSPAC): around 14,000 pregnant mothers were recruited in the early 1990s in the county of Avon, UK.

- Still ongoing, yearly interviews with flexible modules
- Data on children, family, clinical observations, schools, etc.
- DNA methylation observed at: **birth**, **age 7**, and **age 15–19**.
- **Outcomes**, by early adulthood (ages 22–26): university completion, NEET, SMFQ mental health, depression diagnosis.
- **Key shocks**: child abuse (0–10 vs 11–18), and school-entry age cutoff (Sept. 1).

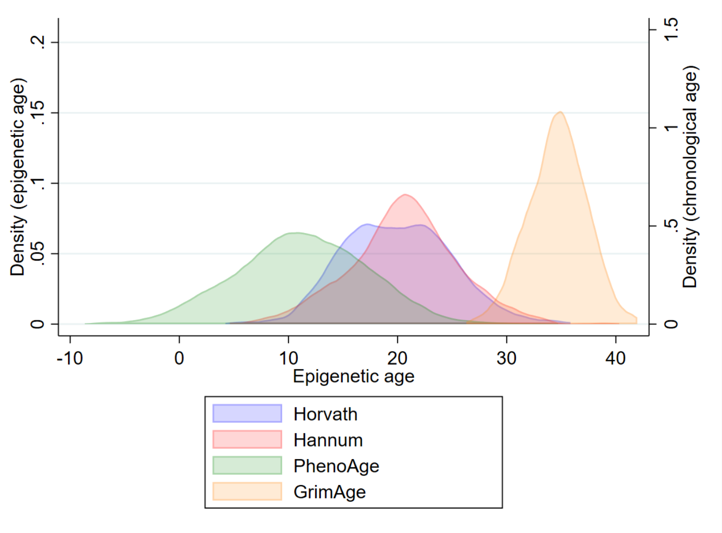
Data

The Avon Longitudinal Study of Parents and Children (ALSPAC): around 14,000 pregnant mothers were recruited in the early 1990s in the county of Avon, UK.

- Still ongoing, yearly interviews with flexible modules
- Data on children, family, clinical observations, schools, etc.
- DNA methylation observed at: **birth**, **age 7**, and **age 15–19**.
- **Outcomes**, by early adulthood (ages 22–26): university completion, NEET, SMFQ mental health, depression diagnosis.
- **Key shocks**: child abuse (0–10 vs 11–18), and school-entry age cutoff (Sept. 1).

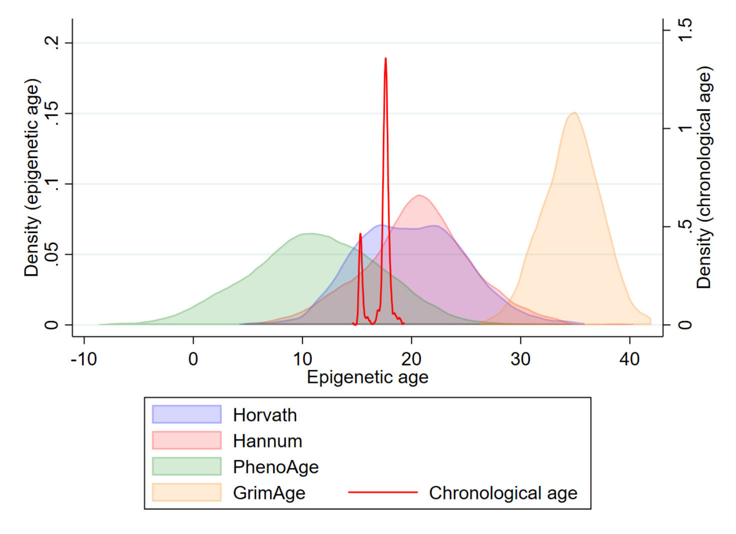
Distribution of epigenetic age by clock

◂ Correlations



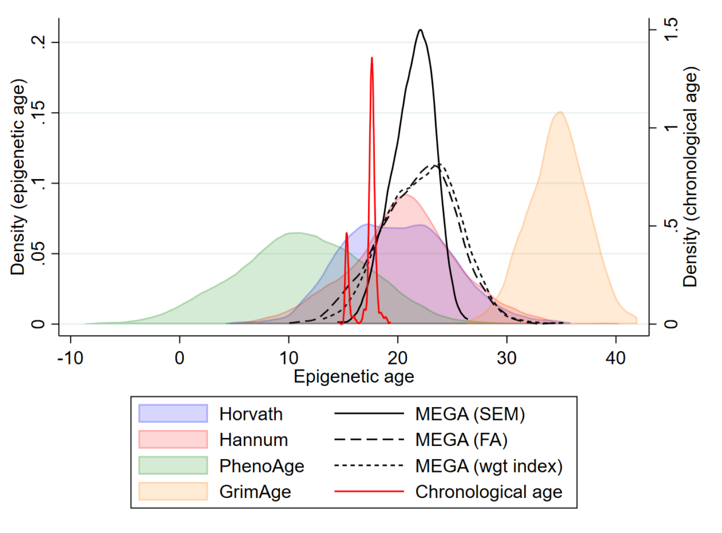
Distribution of epigenetic age by clock

◂ Correlations



Distribution of epigenetic age by clock

◂ Correlations



Empirical Roadmap

- **Application 1 (predictive)**: does adolescent epigenetic age acceleration predict early-adulthood outcomes?
- **Application 2 (determinants)**: is childhood abuse associated with accelerated aging?
- **Application 3 (causal)**: does delaying school entry causally affect biological aging? (sharp RDD)

Application 1: Empirical Strategy

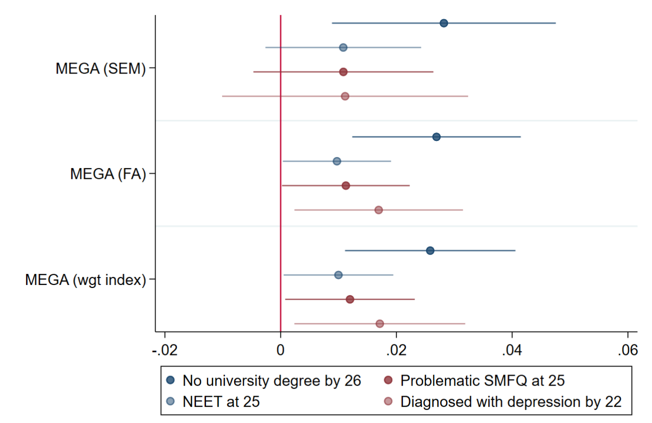
Application 1: Biological Aging and Human Capital

$$Y_{i,t+1} = \delta_0 + \delta_1 MEGA_{i,t} + \delta_2 Health_{i,t} + \theta X_i + u_{i,t}$$

- $Y_{i,t+1}$: binary negative early-adulthood outcomes (no uni degree, NEET, poor mental health, depression diagnosis)
- $Health_{i,t}$: BMI, smoking, drinking (age 15)
- X_i: child's gender, birth year, birth order, and **age at t**; mother's age at birth, mother's education, father's social class.

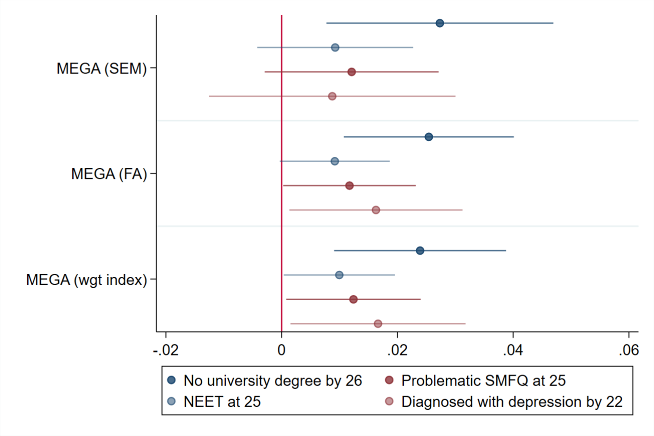
- H1: $\delta_1 > 0, \delta_2 > 0$

Application 1: Main Results



MEGA predicts moderately **worse** early-adulthood outcomes

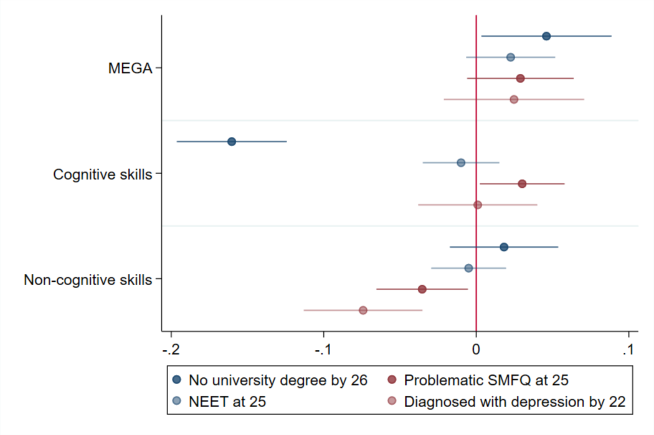
Application 1: Main Results (+ health)



MEGA predicts moderately **worse** early-adulthood outcomes

- Above and beyond health & risky behaviors (BMI, smoking, drinking)

Application 1: MEGA (SEM) vs cognitive and non-cognitive skills



Magnitude comparable to that of non-cognitive skills factor

Application 2: Empirical Strategy

Application 2: Consequences of Child Abuse for Biological Aging

$$MEGA_{i,t} = \beta_0 + \beta_1 Abuse_{i,t} + \beta_2 Abuse_{i,t-1} + \gamma X_i + \epsilon_{i,t}$$

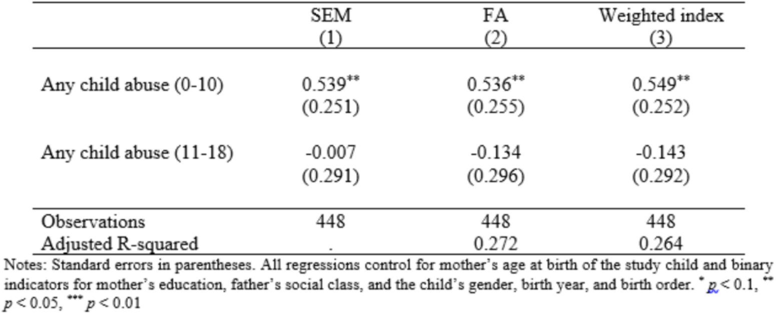
- $Abuse_{i,t}$: dummy for child cruelty or sexual abuse from *t-1* to *t*, reported by either M, P, or C (prospective vs retrospective)
- X_i: child's gender, birth year, birth order, and **age at t**; mother's age at birth, mother's education, father's social class.

- H1: $\beta_1 > 0, \beta_2 > 0$
- Unclear, *a priori*, whether $\beta_1 > \beta_2$ or vice versa

Application 2: Main Results

Abuse

Child abuse and age acceleration from the MEGA clock



	SEM (1)	FA (2)	Weighted 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. 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. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Application 3: Empirical Strategy

Application 3: RDD design to estimate the effect of school-entry age on epigenetic aging.

Sharp school-entry age cutoff: children who turn 4 by August 31st are allowed to start school on the same year.

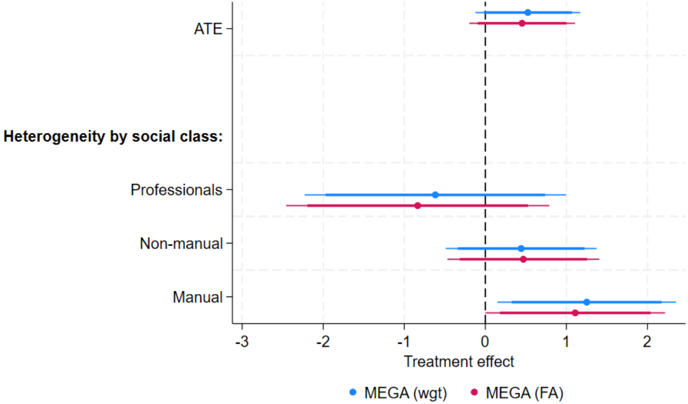
$$MEGA_i^7 = \gamma_0 + \gamma_1 Treat_i + \gamma_2 MoB_i + \gamma_3 Treat_i \times MoB_i + \lambda X_i + \omega_{i,t}$$

- $MEGA_i^7$: MEGA clock measured at age 7
- $Treat_i$: dummy for being born after September 1st
- X_i: child's gender, birth year, and **exact age at collection of DNAm (around age 7)**.
- Bandwidth: ± 3 months from the cutoff
- γ_1: estimated effect of delaying one year of schooling on age acceleration

Application 3: Main Results

Non-parametric jump

Delaying school entry increases age acceleration by 0.5 years by age 7



Social-class gradient in the aging effects of delayed school entry

Mechanisms

❶ **Why does faster biological aging predict worse early-adult outcomes?**

- Onset of symptoms of biological aging
- Scarce evidence on the link between epigenetic clocks and cognitive and non-cognitive skills maintenance and production
  - With the exception of Han et al. (2018): 3/10 most involved gene regions regulate neuronal pathways

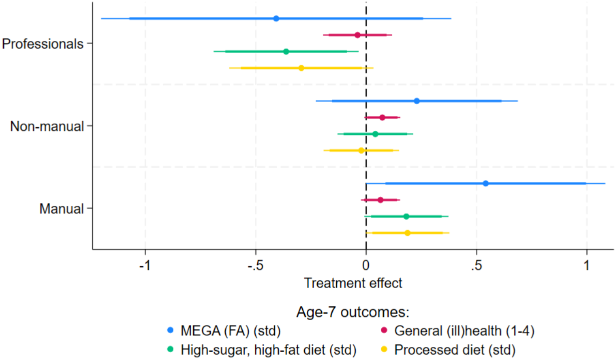
❷ **Why do children who experience abuse in childhood age faster?**

- Biological channels: blood cell-type counts ◂ Figure, cortisol production (Dammering et al., 2021; Suarez et al., 2018)
- Behavioral channels: psychological resilience (Harvanek et al., 2021, 2023), PTSD symptoms (Na et al., 2022), risky behaviors and lifestyle (Schmitz et al., 2022; Harvanek et al., 2023)

Mechanisms

❸ **Why does delaying school entry have adverse biological aging effects?**

- Structured school environment may reduce snacking and increase physical activity (Anderson et al., 2011), with improvements in weight-for-height (Holford and Rabe, 2022)



Conclusion

- **We develop a tool**, the MEGA clock, to measure epigenetic age with increased accuracy
  - Easy to implement for social scientists, does not require direct access to epigenetic data, just to popular clocks
- We provide evidence on the **usefulness of MEGA** through empirical applications
  - Faster aging in adolescence correlates with worse cognitive and non-cognitive outcomes in adulthood
  - Exposure to abuse by age 10 correlates with 0.5-years faster age acceleration
  - Entering school one year later increases age acceleration for low-SES children
- **Policy relevance**: early-life environments (abuse; schooling timing) leave measurable biological traces, with stronger effects for disadvantaged children

Conclusion

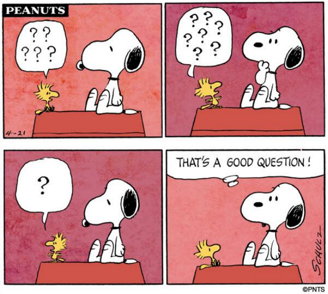
- **We develop a tool**, the MEGA clock, to measure epigenetic age with increased accuracy
  - Easy to implement for social scientists, does not require direct access to epigenetic data, just to popular clocks
- We provide evidence on the **usefulness of MEGA** through empirical applications
  - Faster aging in adolescence correlates with worse cognitive and non-cognitive outcomes in adulthood
  - Exposure to abuse by age 10 correlates with 0.5-years faster age acceleration
  - Entering school one year later increases age acceleration for low-SES children
- **Policy relevance**: early-life environments (abuse; schooling timing) leave measurable biological traces, with stronger effects for disadvantaged children

Conclusion

- **We develop a tool**, the MEGA clock, to measure epigenetic age with increased accuracy
  - Easy to implement for social scientists, does not require direct access to epigenetic data, just to popular clocks
- We provide evidence on the **usefulness of MEGA** through empirical applications
  - Faster aging in adolescence correlates with worse cognitive and non-cognitive outcomes in adulthood
  - Exposure to abuse by age 10 correlates with 0.5-years faster age acceleration
  - Entering school one year later increases age acceleration for low-SES children
- **Policy relevance**: early-life environments (abuse; schooling timing) leave measurable biological traces, with stronger effects for disadvantaged children

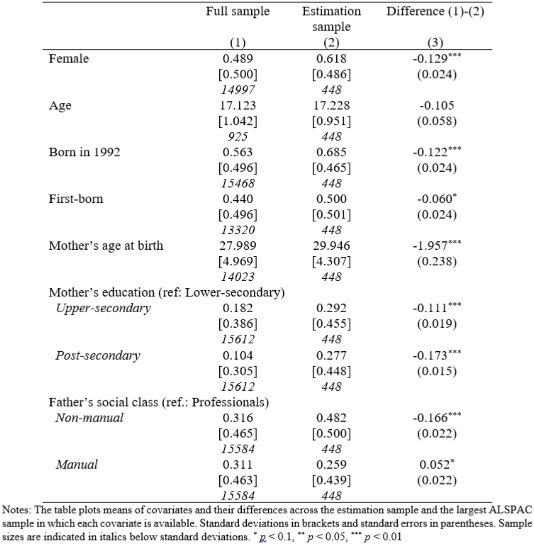
Q&A

Thank you for your attention!



Selection on observables of the estimation sample

Back

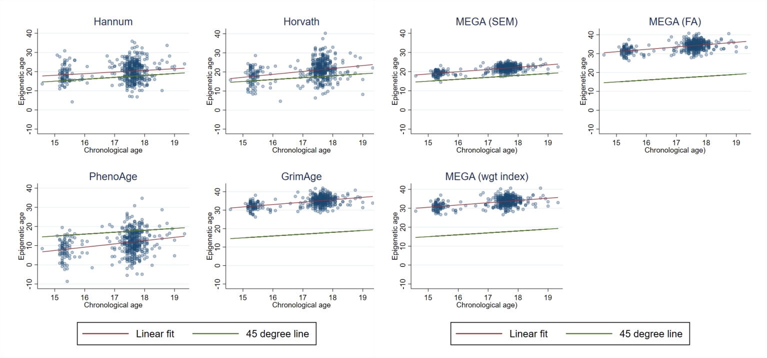


	Full sample (1)	Estimation sample (2)	Difference (1)-(2) (3)
Female	0.489	0.618	-0.129***
	[0.500]	[0.486]	(0.024)
	14997	*448*	
Age	17.123	17.228	-0.105
	[1.042]	[0.951]	(0.058)
	925	*448*	
Born in 1992	0.563	0.685	-0.122***
	[0.496]	[0.465]	(0.024)
	15468	*448*	
First-born	0.440	0.500	-0.060*
	[0.496]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

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

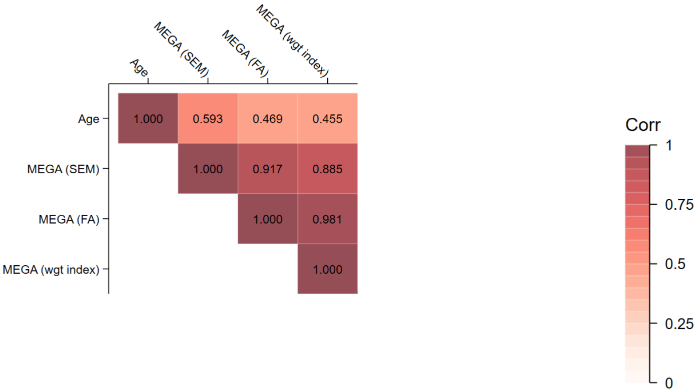
Correlations of chronological and epigenetic age

◂ Back



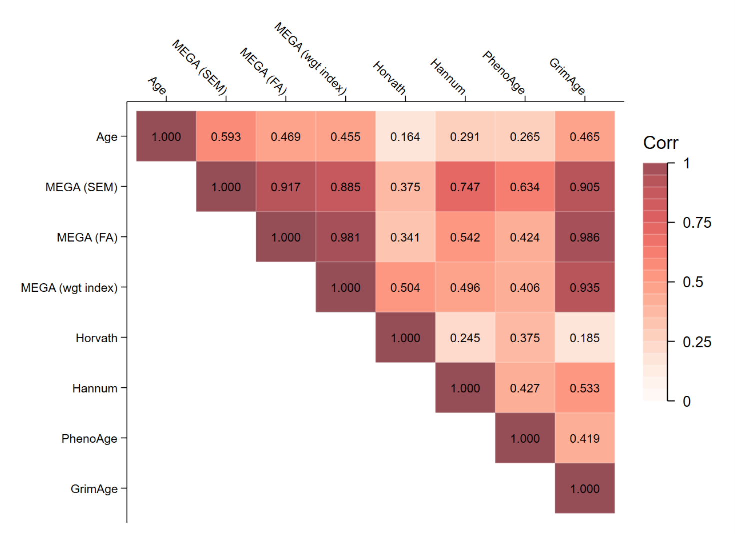
Correlations between clocks

◂ Back



Correlations between clocks

Back



	Age	MEGA (SEM)	MEGA (FA)	MEGA (wgt index)	Horvath	Hannum	PhenoAge	GrimAge
Age	1.000	0.593	0.469	0.455	0.164	0.291	0.265	0.465
MEGA (SEM)		1.000	0.917	0.885	0.375	0.747	0.634	0.905
MEGA (FA)			1.000	0.981	0.341	0.542	0.424	0.986
MEGA (wgt index)				1.000	0.504	0.496	0.406	0.935
Horvath					1.000	0.245	0.375	0.185
Hannum						1.000	0.427	0.533
PhenoAge							1.000	0.419
GrimAge								1.000

Features of the MEGA clock

◂ Back

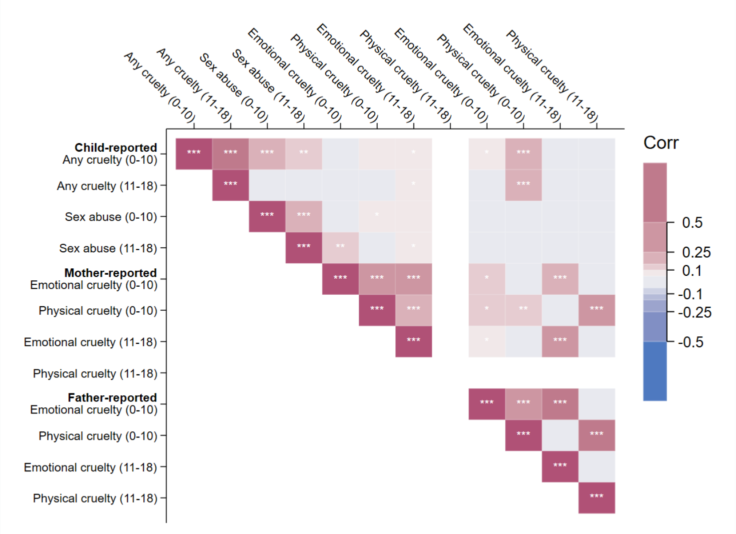
- Exploratory factor analysis (FA) supports a uni-factoral measurement model
    - Only one factor with eigenvalue > 1

Table: Factor analysis - results for MEGA (FA)

Clock	**Factor Loadings**	**Uniqueness**
Horvath	0.410	0.832
Hannum	0.662	0.562
PhenoAge	0.633	0.599
GrimAge	0.639	0.592

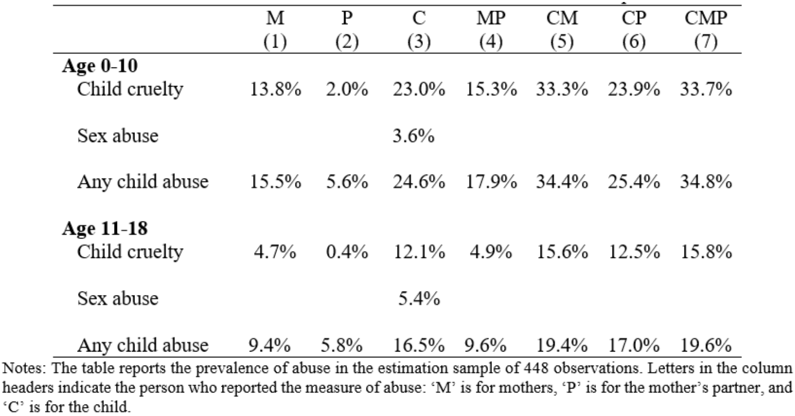
Child abuse in the sample

◂ Back



Prevalence of child abuse in the sample

◂ Back

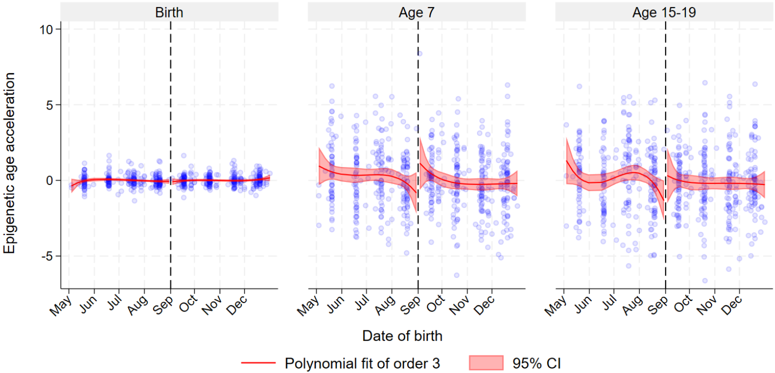


	M (1)	P (2)	C (3)	MP (4)	CM (5)	CP (6)	CMP (7)
Age 0-10							
Child cruelty	13.8%	2.0%	23.0%	15.3%	33.3%	23.9%	33.7%
Sex abuse			3.6%				
Any child abuse	15.5%	5.6%	24.6%	17.9%	34.4%	25.4%	34.8%
Age 11-18							
Child cruelty	4.7%	0.4%	12.1%	4.9%	15.6%	12.5%	15.8%
Sex abuse			5.4%				
Any child abuse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

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

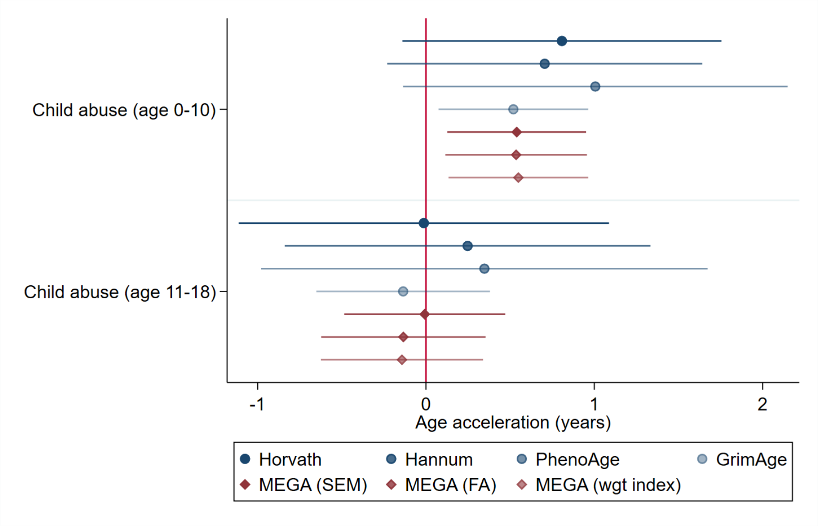
Application 3: Age acceleration (MEGA FA), by MoB

Back



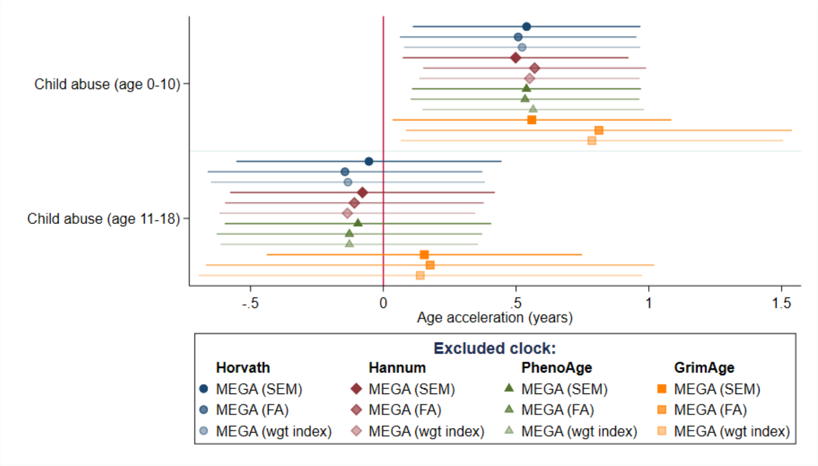
Application 2: MEGA vs single clocks

Back



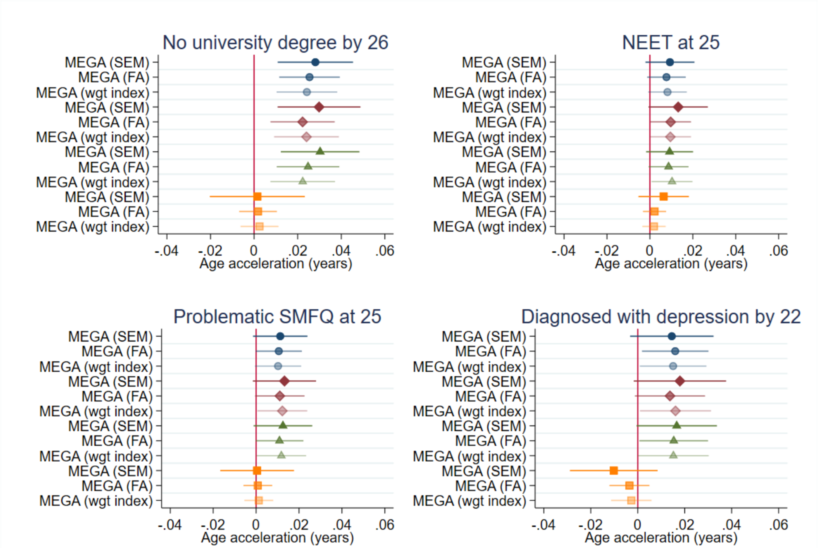
Application 2: Leave-one-out MEGA

◂ Back



Application 3: Leave-one-out MEGA

◂ Back



Age acceleration and blood cell counts by exposure

◂ Back