

University of Nottingham

UK | CHINA | MALAYSIA

Epigenetic age estimates and childhood environment: an interdisciplinary analysis of a Bangladesh-UK cohort

Maria Luisa Evaristo Cursino da Silva

Thesis submitted to the University of Nottingham for the degree of MRes Biosciences

School of Biosciences

September 2023

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions. This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see: http://eprints.nottingham.ac.uk/end_user_agreement.pdf For more information, please contact eprints@nottingham.ac.uk

Table of contents

Title page.....page 1

Table of contents.....page 2

Acknowledgements.....page 5

Abstract.....page 6

List of abbreviations.....page 7

List of figures.....page 9

List of tables.....page 11

Appendix list.....page 12

1. Introduction.....page 13

2. Key issues in ageing, epigenetics and migration.....page 16

2.1. Ageing.....page 16

2.1.1. What is biological ageing?.....page 17

2.1.2. Which factors influence ageing?.....page 17

2.1.3. Stress, cortisol and biological ageing.....page 18

2.1.4. Ways of measuring biological ageing.....page 20

2.1.5. Epigenetics as a way of measuring biological ageing.....page 24

2.2. Epigenetics.....page 25

2.2.1. What is epigenetics?.....page 25

2.2.2. History of epigenetics.....page 26

2.2.3. DNA methylation and CpG sites.....page 27

2.2.4. What is epigenetic ageing?.....page 32

2.2.5. Environmental factors and epigenetic ageing.....page 43

2.3. Migration.....page 52

2.3.1. Migration and paediatric development.....page 52

2.3.2. Migration and epigenetic ageing.....page 56

2.4. Migration, cortisol and epigenetic ageing.....page 58

3. Methods.....page 60

4. Results.....page 66

4.1. Horvath's multi-tissue clock: one clock, three sets of results.....page 67

4.2. PedBE clock.....page 76

4.3. Skin and blood clock.....page 78

4.4. Hannum clock.....page 80

4.5. PhenoAge clock.....page 83

4.6. LinAge clock.....page 85

4.7. Summary of findings.....page 88

5. Discussion.....page 91

5.1. Hurdles with Horvath clock submissions.....page 91

5.2. Support from bioinformaticians.....page 93

5.3. Limitations.....page 93

5.4. Epigenetic significance of our findings.....page 95

5.5. Migration and epigenetic ageing.....page 97

5.6. Migration and its significance in understanding environmental stressors..... page 97

5.7. Cortisol, stress and epigenetic ageing.....page 102

5.8. Key advantages of our study to the research community.....page 102

5.9. Philosophy of science, eugenics and epigenetics.....page 103

6. Conclusion.....page 105

7. References.....page 106

Appendix.....page 118

Acknowledgments

I am incredibly grateful to all who have supported me towards getting to this stage. In particular, my academic supervisor, Reinhard Stöger, for his support throughout my MRes degree and for his expertise in the field of epigenetics, which meant I could effectively communicate our research findings despite having a limited background in Biological Sciences. Thank you to my examiners, Simon Welham and Dianne Ford, for their thorough and constructive feedback, which have certainly improved the quality of the work presented in this thesis. Thank you to Richard Emes, my second supervisor, as well as to Rachel Clifford and Poojitha Rajasekar for their support and bioinformatics skills, which supported the collection of results from epigenetic clocks, a central component of the findings reported in this thesis. I also want to thank the wider School of Biosciences postgraduate team, particularly Ravinder Anand-Ivell and Mandy Walker, for their pastoral support during challenging times as I balanced my MRes studies with a full-time job at BBC News. The conclusion of these studies would not have been possible without my family and friends, to whom I owe my immense gratitude. Thank you to my boyfriend, Nicholas Elmer, who supported the findings of this thesis by discussing the use of data science tools, and for his constant care for my wellbeing. Thank you to my parents for their unconditional love and support and for being an endless source of inspiration. Science was all around me growing up, and I thank them for teaching me that knowledge is wealth – without pressuring me to follow in their footsteps.

Abstract

Background: Ageing is not a negative phenomenon but ageing prematurely can lead to adverse health outcomes. This study investigates the rate of epigenetic ageing across migrant and non-migrant populations of girls living in the UK. First-generation Bangladeshi migrant girls are the focus of this study as we pursue a better understanding of whether migration as a socio-economic factor can affect epigenetic ageing. The thesis builds on the work by Houghton and colleagues, who previously investigated the relationship between socio-economic factors and pubertal timing across a larger cohort than the one studied in this thesis. **Methods:** For this thesis, the DNA methylation data from the saliva samples of a subset of data donated from Houghton et al.'s study (115 girls) was analysed to predict epigenetic age estimates based on the methylation levels of buccal epithelial cells. DNA methylation-based epigenetic clocks were used as a proxy to predict the biological ages of the cohort, which was composed of Bangladeshi girls born and raised in Sylhet Town, Bangladesh (n=43), first-generation Bangladeshi girls who left Bangladesh to live in the UK (n=24), second-generation Bangladeshi girls born in the UK to immigrant parents (n=27), and white British girls born in the UK (n=21). **Results:** Across the six clocks used in this study, epigenetic age estimates from the PedBE clock yielded the most closely correlated results to the populations' chronological ages. The skin and blood clock also delivered similarly correlated results. According to data harvested from most clocks used in this study, first-generation migrant Bangladeshi girls had a higher epigenetic age acceleration rate. However, ANOVA tests concluded that the discrepancy was not statistically significant with the sample size available. **Conclusion:** The PedBE clock's accuracy is linked to the tissue-type and age range from this study's cohort, which met the specificity parameters used to design the PedBE clock. This study identified migration as a socio-economic stressor that may contribute to accelerated biological ageing. Although the divergence across groups was not statistically significant, it alludes to a trend corroborated in later life among migrant Bangladeshi women in the UK. The work undertaken here implores further investigation and action from policymakers to better support immigrant communities to the UK, particularly those coming from the Global South.

List of abbreviations

ABBY: Adolescence Among Bangladeshi and British Youth

ACEs: Adverse childhood experiences

ACGT: DNA bases; adenine (A), cytosine (C), guanine (G) and thymine (T)

ACTH: Adrenocorticotropic hormone

AFAR: American Federation for Aging Research

AgeAccel: Age acceleration rate, used in the context of DNA methylation-based epigenetic clocks

BMI: Body mass index

C-glyTrp: C-glycosyl tryptophan

CpG: Cytosine and Guanine separated by a phosphate

CRH: Corticotrophin releasing hormone

DHEAS: Dehydroepiandrosterone sulfate

DNA: Deoxyribonucleic acid

DNAm: DNA methylation

DNMTs: DNA methyltransferases

EPIC array: Illumina HumanMethylationEPIC BeadChip array

Et al.: And others

GrimAge: DNA methylation-based age mortality prediction

HGPS: Hutchinson Gilford progeria syndrome

HPA axis: Hypothalamic-pituitary-adrenal axis

LinAge: DNA methylation-based age informed by research from Lin and colleagues

PhenoAge: DNA methylation-based age prediction based on phenotypic biomarkers

PedBE: Paediatric buccal epigenetic clock

PRISMA: Preferred reporting items for systematic reviews and meta-analysis

qPCR: Real-time polymerase chain reaction

RNA: Ribonucleic acid

SAM: S-adenyl methionine

SNPs: Single nucleotide polymorphisms

TET enzymes: A family of ten-eleven translocation (TET) methylcytosine dioxygenases

TSS: Transcription start site

TTS: Transcription termination site

UK: United Kingdom

5hmC: 5-hydroxymethylcytosine, a demethylated form of the DNA base cytosine (C)

5mC: 5-methylcytosine, a methylated form of the DNA base cytosine (C)

List of figures

- **Figure 1**. Age acceleration measured across three epigenetic clocks (Horvath, Hannum and Levine) plotted against clinical biomarkers and environmental factors
- **Figure 2.** PedBE. Linear regression graph shows the DNA methylation-based age plotted against the chronological age of paediatric subjects in the training dataset of the PedBE clock.
- **Figure 3.** Estimated epigenetic ages from the PedBE clock plotted against the chronological ages in years of children involved in Dammering et al.'s study.
- **Figure 4.** Box plot graph that shows the covariance of standardised residuals of the PedBE clock age estimates regressed against chronological age in children with and without internalising disorder as part of Dammering et al.'s study.
- **Figure 5.** Age acceleration based on data from **Horvath's** epigenetic clock, using raw data.
- **Figure 6.** Age acceleration based on data from **Horvath's** epigenetic clock, using normalised data.
- **Figure 7. Horvath's** raw clock epigenetic age estimates plotted against individuals' chronological ages.
- **Figure 8. Horvath's** normalised clock epigenetic age estimates plotted against individuals' chronological ages.
- **Figure 9.** Age acceleration based on data from Horvath's multi-tissue epigenetic clock.
- **Figure 10.** Horvath clock epigenetic age estimates plotted against individuals' chronological ages.
- **Figure 11.** Age acceleration based on data from PedBE (paediatric buccal epigenetic) clock.
- **Figure 12.** PedBE clock epigenetic age estimates plotted against individuals' chronological ages.
- **Figure 13.** Age acceleration based on data from Skin & Blood clock.
- **Figure 14.** Skin and Blood clock epigenetic age estimates plotted against individuals' chronological ages.

- **Figure 15.** Age acceleration based on data from Hannum clock.
- **Figure 16.** Hannum clock epigenetic age estimates plotted against individuals' chronological ages.
- **Figure 17.** Age acceleration based on data from PhenoAge clock.
- **Figure 18.** PhenoAge clock epigenetic age estimates plotted against individuals' chronological ages.
- **Figure 19.** Age acceleration based on data from Lin Age clock.
- **Figure 20.** Lin Age clock epigenetic age estimates plotted against individuals' chronological ages.
- **Figure 21.** Horvath's 2013 DNA methylation-based epigenetic clock's weighted methylation average plotted against chronological age in years.
- **Figure 22.** Linear regression of DNA methylation-based age estimates plotted against first-generation population's chronological ages.
- **Figure 23.** Table demonstrating socio-economic variables and sample characteristics of the four groups of populations analysed in the study carried out by Houghton and colleagues.

List of tables

- **Table 1.** Summary of findings from epigenetic clocks.

Appendix list

- **Appendix A. Table A**. ANOVA statistical results for six epigenetic clocks, with three entries on Horvath's multi-tissue clock.
- **Appendix B. Figure B**. Linear regression of DNA methylation-based age estimates plotted against the white British population's chronological ages.
- **Appendix C. Figure C.** Linear regression of DNA methylation-based age estimates plotted against the Bangladeshi population's chronological ages.
- **Appendix D. Figure D.** Linear regression of DNA methylation-based age estimates plotted against the second-generation population's chronological ages.

1. Introduction

Ageing in and of itself is not a negative characteristic but ageing prematurely can lead to several adverse health outcomes, including a higher predisposition to diseases like cancer (Zheng, Widschwendter and Tischendorf, 2016). This thesis will give the reader a better understanding of biological ageing, focusing much of the analysis on epigenetic ageing and migration as a socio-economic factor that affects physiological compositions in Global South populations migrating to the Global North.

Each chapter will build on the principles explained and developed in the previous section. Methodically, this study will provide insight into the biological mechanisms in place to assess ageing. Given the ongoing debate about the best way to measure biological ageing, this thesis will explain why the decision was made to focus on results gathered from DNA methylation-based tools to measure epigenetic ageing (epigenetic clocks). While explaining the benefits of epigenetic clocks, other well-documented forms of predicting age acceleration rates will be discussed - including telomere length, inflammatory biomarkers, DNA damage and repair information, to name a few.

Much of the discussion of this thesis builds on the previous work by Houghton and colleagues, who carried out a bio-cultural analysis of the impact of migration and other socio-economic factors on the pubertal timing of Bangladeshi girls who had moved to the UK during their childhood, contrasted with their non-migrant counterparts both in the UK and Bangladesh (Houghton et al., 2014a). A subset of saliva samples collected for Houghton and colleagues' study is being used in this thesis to assess DNA methylation levels across the cohort and provide a better understanding of the epigenetic age acceleration rate measured from the methylation levels at specific CpG sites - dinucleotides within DNA, where a cytosine nucleotide is followed by a guanine nucleotide. The sample data was donated by Gillian Bentley and colleagues at the University of Durham for the purpose of this study.

DNA methylation-based epigenetic clocks were first developed in 2011 (Bocklandt et al., 2011). Since then, several other epigenetic clocks have been developed, enhancing the research

community's understanding of the correlation between methylation levels at specific CpG sites near gene promoters, and clinical biomarkers, chronological age, health outcomes as well as mortality.

DNA methylation levels of 115 girls aged between 5 and 16 years were calculated across six different epigenetic clocks. The girls were categorised into four groups: Bangladeshi girls who were born and raised in Sylhet (non-migrant); first-generation migrant Bangladeshi girls who moved to the UK during a critical phase of their childhood development, as some moved before they reached adrenarche (migrant); second-generation Bangladeshi girls born in the UK to Bangladeshi immigrant parents (non-migrant); and white British girls who were born in the UK to Caucasian parents (non-migrant). We hypothesised that migrant communities from the Global South were exposed to environments that led to epigenetic age acceleration.

Epigenetic age predictions for this study's samples were gathered using Horvath's multi-tissue epigenetic clock, McEwen and colleagues' PedBE clock, which is specific for buccal samples from a paediatric cohort. Other clocks used include Levine and colleagues' PhenoAge clock, Horvath and colleagues' skin and blood clock, Lin et al.'s epigenetic clock, and Hannum's clock (Horvath, 2013; McEwen et al., 2020, Levine et al., 2018; Horvath et al., 2018; Lin et al., 2016; Hannum et al., 2013). Results from the PedBE clock were most accurate at predicting the chronological age of our cohort, as discussed in the results and discussion chapters. Epigenetic age estimates from DNA methylation data were used as a proxy to infer the biological age of individuals whose samples were analysed.

Our work explored the relationship between migration, stress, cortisol levels and epigenetic ageing. In doing so, the thesis concluded that although epigenetic age acceleration findings were not statistically significant, the demographical context in which first-generation migrant Bangladeshi girls were growing up may trigger heightened cortisol levels due to stress, and subsequently impact methylation levels at CpG sites, which can in turn lead to accelerated biological ageing due to the silencing or overexpression of specific genes (Hunter and McEwen, 2013; Belsky et al., 2015).

In short, the work comprised in this thesis hopes to detangle the broader relationship between stress and ageing, while working to better understand migration as a stressor that may contribute to accelerated biological ageing. Due to the interdisciplinarity of this study, the aim is that readers from both the sciences and humanities will have a clear understanding of key themes in the field of epigenetics, biological ageing, and the interaction of those two fields with socio-economic factors. Key terms necessary for the reader's understanding of the wider paper will be defined to facilitate dialogue and knowledge exchange across disciplines.

A deeper understanding of the connection between undue stress, heightened cortisol levels and gene silencing presents itself as a framework capable of detangling the relationship between the environment and accelerated epigenetic ageing in specific communities. Our findings show that individuals born in Bangladesh, who have migrated to the UK during their formative years, age more rapidly than their non-migrant counterparts - both in Bangladesh and the UK. Mortality is a guaranteed phenomenon across all living organisms. Still, our research findings raise alarm bells as they signal that migrant populations are experiencing an unfair physiological burden in response to the environment they are relocated to during their childhood. Thus, the findings presented in this thesis implore action from policymakers to better support communities in transit, particularly those from the Global South, as they experience the legacy of coloniality, and historical hostile environments inflicted upon them.

2. Key issues in ageing, epigenetics and migration

Ageing, epigenetics and migration have been historically discussed in isolation. We aim to spell out the iterative relationship between these three fields of study while providing an interdisciplinary analysis on the relationship between the physical and the social. Methodically, we will first outline key discussions in the field of ageing research. Following an exploration of the key factors that influence ageing, and biological mechanisms used to measure ageing across the human population, we will move on to discuss epigenetics as a tool that provides well-informed age estimates. This will lead us to consider environmental factors that influence epigenetic ageing, focusing on migration and its association with stress and heightened cortisol levels.

2.1. Ageing

One of the few guarantees during an individual's lifespan is their mortality. Many fear it and work tirelessly to avoid, or obfuscate, bodily manifestations of ageing – from plastic surgeries to hair transplants. However, during key developmental stages, ageing is an extremely positive sign. Across different species, to see your offspring able to fend for themselves ignites pride, joy, satisfaction, among other behavioural manifestations. Although humans exert these manifestations when their new-borns achieve key developmental milestones, like crawling and uttering sounds that resemble part of the vocabulary they are exposed to, the prospect at the tail-end of our lifespans – death – is unnerving, intrusive and researchers across different disciplines are working tirelessly to halt, reverse and better understand ageing. Mammalian ageing can be delayed by genetic, dietary and pharmacologic approaches but as humans continue to live for longer, a more coherent understanding of factors that may accelerate ageing in early life will prove useful years to come. Given the relationship between ageing and some chronic diseases, a deeper insight into ageing could also improve quality of life, as well as ensuring individuals live longer. Labelled geroscience, researchers across multiple disciplines – from epigeneticists to inflammation specialists – have combined forces to provide a more cohesive understanding of the field of ageing (Kennedy et al., 2014).

2.1.1. What is biological ageing?

Biological ageing is a term given to processes in living organisms' bodies that can be observed by researchers and clinicians. It is often assessed based on key biomarkers, which tell a story about an organisms' behavioural patterns, nutrition, immune system, proneness to diseases and general health outcomes. Often discussed as biogerontology, the scientific field works at the intersection of biology, genetics, medicine and gerontology, and provides explanations to the mechanisms and biochemical reactions responsible for why (and how) we age (López-Otín et al., 2013). Kennedy and colleagues present the 7 pillars of ageing: metabolism, macromolecular damage, epigenetics, inflammation, adaptation to stress, proteostasis, and stem cells and regeneration (Kennedy et al., 2014). This study will focus on epigenetics and adaptation to stress as factors that impact ageing, without overlooking these other pillars.

The biomarkers of ageing are not too dissimilar to biomarkers of general health. But there is a distinction between molecular biomarkers of ageing, based on DNA and RNA information, and phenotypic biomarkers, which are based on clinical factors such as blood pressure and cholesterol levels (Jylhävä, Pedersen and Hägg, 2017). The American Federation for Aging Research (AFAR) formulated the criteria for ageing biomarkers (Johnson, 2006), which suggests that they must: predict the rate of ageing and accurately inform where individuals are in their total lifespan; monitor basic processes that underlie ageing in particular, not disease; be tested repeatedly without using intrusive techniques (samples must originate from blood or saliva, for example); and biomarkers should be functional tools in humans and in laboratory animals, to ensure that measures are only validated in humans (Jylhävä, Pedersen and Hägg, 2017). However, as argued by Jylhävä and colleagues, and still remains the case 6 years later, there is no consensus across the field and literature on what an ageing biomarker is. The subsequent discussion will delve into the current state of play in terms of biological ageing biomarkers.

2.1.2. Which factors influence ageing?

Time is one of the most obvious factors to influence ageing, but controlling it is not a possibility. Hence environmental, behavioural and clinical factors have received close attention in the field of ageing. The main motivation being that if we can build a coherent and empirically sound

understanding of what processes accelerate epigenetic ageing, then we might be able to employ preventative measures and work alongside policymakers to build an environment conducive to decelerated ageing.

There are myriad environmental factors that can be traced back to impact biological ageing. Exposure to air pollution is an example, as it can be a constant for many who live in urban areas. Ward-Caviness and colleagues have studied the relationship between exposure to air pollutants and accelerated biological ageing. The team assessed exposure to various air pollutants and measured the rate of biological ageing on several fronts. Ward-Caviness et al. used data on telomere length and DNA methylation age acceleration, a measure of epigenetic ageing, and went on to correlate it with the immune cell count of individuals (Ward-Caviness et al., 2016). The conclusion from their findings was that long-term exposure to air pollution was linked to accelerated biological ageing (Ward-Caviness et al., 2016).

2.1.3. Stress, cortisol and biological ageing

Stress is another factor that influences biological ageing. The relationship between stress, cortisol and biological ageing – epigenetic ageing in particular – is significant because heightened cortisol levels are largely caused by stressors in an individual's environment. In certain situations, these stressors lead to positive longevity outcomes as they warn the body of danger, or limited resources. However, sustained and heightened levels of stress can be detrimental to an individual's physiology and health outcomes because it can impact DNA methylation levels at key CpG sites.

The hypothalamic-pituitary adrenal (HPA) axis, for example, is sensitive to stressors in the environment and it is vital in the human body's response to stress (Perroud et al., 2011). Glucocorticoids, a class of corticosteroids produced primarily in the adrenal glands, are secreted as a response mechanism to signals from the hypothalamus and pituitary gland. The axis between these three glands is important to the neuroendocrine system because it regulates the production and release of stress hormones, like cortisol. The hypothalamus also secretes corticotrophin-releasing hormone (CRH) when it responds to stressors like physical or emotional stress. The CRH then makes its way to the pituitary gland. In receipt of the hormone, the pituitary

gland releases adrenocorticotropic hormone (ACTH) into the body's bloodstream. The adrenal glands – which typically develop in girls when they are aged between 6 and 8 years – are stimulated by ACTH to produce and release cortisol (Smith and Vale, 2022).

The fight or flight analogy is helpful to illustrate the body's physical response to environmental stressors. That is because when individuals see a threat or endure a traumatic experience, a chain of physiological mechanisms are triggered and carried out by the sympathetic nervous system and the HPA axis itself. The biochemical chain reaction is kicked off by the sympathetic nervous system, which enables the fast release of epinephrine (known as adrenaline) and norepinephrine (noradrenaline). The main bodily manifestations are increased blood pressure and heart rate, as these signals are preparing the body to leap into immediate action. After these hormones are in flow, the HPA axis releases cortisol, which prolongs the body's stress response to ensure individuals are ready to face whatever the stressor may be (Alexander et al., 2018).

In addition to preparing the body to respond to environmental stressors, cortisol plays an important role in modulating gene expression. Epigenetic mechanisms such as DNA methylation, histone modification and non-coding RNAs are all impacted by cortisol, which can promote silencing of certain gene expressions.

2.1.4. Ways of measuring biological ageing

Jylhävä, Pedersen and Hägg carried out a review to assess advances made in the field of molecular biology to identify biomarkers that can be used as biological age predictors (Jylhävä, Pedersen and Hägg, 2017). Jylhävä and colleagues considered six types of biological age predictors and concluded, in a similar way to this study, that the most promising biological age predictor was the epigenetic clock. However, the team emphasised that more work needs to be done to determine its value as a biomarker of ageing, adding that longitudinal studies would support this. The team compared DNA methylation-based age predictor of biological age with telomere length, transcriptomic age, Glycan age, protein-derived age, C-glyTrp, metabolic age score and composite biomarkers. Other biological age predictors include DNA damage and repair, physical performance and other metabolic parameters (Jylhävä, Pedersen and Hägg, 2017).

Telomere length has been used as a biological age predictor to determine age-associated health outcomes such as cancer, cardiovascular disease, Alzheimer's disease, physical function, cognition and mortality. Telomeres are repetitive DNA sequences at the natural ends of eukaryotic chromosomes that cap and protect chromosome ends from degradation, recombination, and fusions (Maddar, Ratzkovsky and Krauskopf, 2001). Each time cells divide, telomeres shorten as a result and analysing their length was one of the most popular markers of cell ageing in the 2000s (Blackburn et al., 2006). To measure telomere length, qPCR (real-time polymerase chain reaction) is carried out and for each experimental sample, the ratio of the result is the relative telomere to single copy gene (T/S) (Cawthon, 2002). Research carried out by Gardner et al. and by Lapham et al. indicates that women, on average, have longer telomeres than men (Gardner et al., 2014; Lapham et al., 2015).

Telomere shortening activates DNA damage response pathways, which go on to modify epigenetic biomarkers associated with ageing. Telomeres are dynamic and their length can vary in response to genetic and environmental factors. They can be indicative of an individual's and a helpful biomarker, as factors such as stress can contribute to accelerated telomere shortening. It is one of the many factors that promote **cellular senescence**, a state of irreversible cell cycle arrest and thereby a stage in the cell cycle linked to ageing. Nevertheless, the dynamism in which

telomere lengths change across different cell types makes it less reliable than other biomarkers when it comes to assessing its correlation with chronological age. DNA damage, metabolic alterations and epigenetic changes can also lead to cellular senescence. In this state cells lack replicative capacity and cannot work towards homeostasis. However, cells that have reached cellular senescence are still active, and they secrete pro-inflammatory cytokines (Coppé et al., 2010; Lowe, Horvath and Raj, 2016). **Inflammation** has been identified as one of the causes for many diseases and morbidities. Chronic inflammatory diseases and oxidative stress are linked to chronological age as they can often activate signalling pathways that impact key biomarkers, and subsequently impact gene expression. Thus, cellular senescence, telomere length and inflammation are among well-accepted biomarkers of ageing (Fransquet, Wrigglesworth and Ryan, 2023).

Jason Lang and colleagues published a paper in 2020 that analysed the link between adverse childhood experiences and epigenetic ageing by measuring telomere length. The paper carried out a systematic review using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines, which is a 27-item checklist to ensure researchers carry out transparent reporting of systematic reviews and meta-analyses. The study hoped to understand what epigenetic, telomeric and associated biological changes relate to an individual's exposure to adverse childhood experiences (ACEs) before they are 12 years old (Lang et al., 2020). Lang et al. selected 23 high-quality scientific papers on the subject, from a pool of 190 papers. From the analysis carried out, Lang et al. observed that scientific papers which investigated such correlation concluded that there was a more powerful effect on telomere length if physical neglect was involved (Lang et. Al, 2020). The team identified that most study participants had been adults, and methylation level assessments were the chosen method to measure epigenetic modification. Lang et al. argued that more research is need on non-coding aspects of the epigenome, as well as on coding aspects beyond DNA methylation levels (Lang et al., 2020).

As discussed, Jylhävä and colleagues investigated six marker categories for ageing across human populations: epigenetic clock, telomere length, transcriptomics, proteomics, metabolomics and multi-biomarker.

Transcriptomics is a field of molecular biology that investigates an organism's set of RNA transcripts, molecules, notably messenger RNA (mRNA), non-coding RNA (ncRNA) and other types of RNA. The study of transcriptomics is often either cell-specific, tissue-specific or organism-specific – depending on the dataset and the quality of samples researchers have at their disposal. Transcriptomics aides our understanding of the regulation of gene expression and it has been identified as a way of studying biological ageing. As transcriptomics focuses on different types of RNA and the subsequent impact on gene expression, it is a useful tool when researchers want to identify which genes are expressed at higher levels (upregulated) and which ones are expressed at lower levels (downregulated) while an individual ages (López-Otín et al., 2013). RNA sequencing and microarrays are techniques that provide datasets where researchers can analyse the set of RNA transcripts – either on the entire organism or specific tissue types – and go on to compare gene expression profiles across paediatric and adult cohorts (Peters et al., 2015). The field of transcriptomics has informed researchers investigating DNA methylation-based age estimates, and although different approaches to measuring ageing are continuously evolving, there is an iterative relationship across the field of molecular biology as the quest to better understand ageing continues. Hannum et al., whose epigenetic clock we will discuss in the upcoming pages, used transcriptomics to identify key age-related biomarkers (Hannum et al., 2013). Studies of the RNA transcript have also contributed to the body of work focused on which biological processes and pathways are most impacted by chronological ageing, identifying gene expression pathways that relate to oxidative stress, inflammation, DNA repair and cell cycle regulation (Franceschi et al., 2000).

Proteomics is another subfield of molecular biology, which studies the complete set of proteins expressed by an organism, or within a particular tissue – depending on the parameters and focus of the scientific endeavour researchers are focusing on. A study carried out by Tanaka, Biancotto and Moaddel investigated and identified specific proteins whose gene expression levels change with age (Tanaka, Biancotto and Moaddel, 2018). Tanaka and colleagues were able to use elastic net regression models to create a proteomic signature of age that is based on the concentrations of 76 proteins identified as highly correlating to chronological age from samples originating from 240 men and women aged between 22 and 93 years. In total, 1,301 proteins were measured in

plasma from these individuals' samples. The correlation between the proteomic predicted age and individuals' chronological ages had an R value of 0.94 (Tanaka, Biancotto and Moaddel, 2018).

Metabolomics is the study of metabolites, small molecules involved in bio-chemical reactions in a living organism's cells and tissues. There are a range of metabolites – primary metabolites focus on basic functions like energy production, cell structure and growth, for example – and each are significant in ensuring organisms can carry out biological processes seamlessly. As individuals age, they experience several metabolic changes, including alterations in how their bodies make use of nutrients in their diet. A review into metabolomics by Johnson, Ivanisevic and Siuzdak explained that homeostatic controls that require certain metabolites can become compromised with age and disease, which would in turn lead to an organism's functional decline and failure to return to a steady state (Johnson, Ivanisevic and Siuzdak, 2016). Also in 2016, Chaleckis and colleagues were able to identify age-related metabolite concentration changes. In their research, Chaleckis et al. Used high-resolution liquid chromatography-mass spectrometry to perform non-targeted quantitative metabolomics analysis based on the blood samples of 15 young individuals and 15 elderly individuals – 126 blood metabolites were obtained from all 30 subjects. From the data the team harvested, they were able to conclude that age differences were partly linked to a decrease in antioxidant production and increased inefficiency of urea metabolism among the older subjects included in the analysis.

Cell passage number can aid our understanding of biological ageing, as it is closely correlated with the chronological age of a cell. That is because cell passage number means the number of times a cell population has been passaged or undergone sub-culture in a lab. It is also an important way to monitor the reproducibility of cell populations and their propensity to genetic stability – as well as to alert researchers of whether the cell population is damaged, as it will reproduce more slowly or cease to reproduce altogether, reaching cellular senescence (Lowe, Horvath and Raj, 2016). The number of times a population of cells has been sub-cultured since originally isolated can help explain at which stage of development particular cells are in. Earlier, the term telomere shortening was explored and the significance of cell passage number is not too dissimilar – somatic cells have limited capacity for replication and as cells divide, they

experience changes to gene expression, telomere length and they also accumulate cellular damage. Therefore, to understand biological ageing from cell passage, researchers can infer that the number of passages a cell culture has undergone is linked to the number of times cells have divided and it can indicate how long they have been around for (their age) (Hayflick and Moorhead, 1961).

2.1.5. Epigenetics as a way of measuring biological ageing

This study decided to focus on epigenetic ageing as a way of measuring biological ageing because of its accuracy across large cohort studies, as well as individuals' entire lifespans. DNA methylation-based algorithms, known widely as epigenetic clocks, were chosen as a method to assess biological age because of their ability to provide insight into tissue-specific methylation levels at key CpG sites. These mathematical algorithms were chosen as biomarkers of ageing because of their strong correlation with chronological age, as well as their ability to identify individuals whose epigenetic age does not match their chronological age. In instances where individuals have a higher epigenetic age than their chronological age, other biomarkers have been used alongside to suggest worse health outcomes for those with accelerated ageing (Fransquet, Wrigglesworth and Ryan, 2023).

As outlined in the landmark epigenetic clock paper by Horvath, the DNA methylation-based tool could even estimate an age close to 0 years in embryonic and induced pluripotent stem cells. In those circumstances, the age estimates correlated with cell passage number – which as described above is a relevant biomarker of biological ageing (Horvath, 2013; Lowe et al., 2016). The rate that individuals biologically age is not consistent across different tissue types, therefore epigenetic clocks tailored at assessing multiple tissues look at many CpG islands, which are stretches of DNA that have a higher CpG density than the rest of the genome (Bird; 1986). As an example, the female breast tissue was identified as ageing more rapidly than other tissues (Sehl, 2017). Unlike other biomarkers, some epigenetic clocks can predict all-cause mortality independently of risk factors such as smoking, alcohol use and physical activity. Chen and colleagues carried out a meta-analysis of more than 13,089 individuals which corroborated the

notion that some epigenetic clocks can predict all-cause mortality independent of these risk factors (Chen et al., 2016; Jylhävä, Pedersen and Hägg, 2017). As a tool for age prediction, the epigenetic clock – in all its variations – accurately predicts a spectrum of ageing outcomes as a pervasive and systems-level mechanism. It is not yet the gold standard for measuring biological ageing, but it is the most advanced biological mechanism to date, capable of identifying clinical biomarkers and with additional time it will continue to be calibrated (Li, Koch and Ideker, 2022). Nevertheless, there are some unresolved questions on whether the epigenetic clock's so-called ticking rate can be modifiable and whether methylation changes observed with age and ageing phenotypes trigger changes to gene expression, or if they represent the work of other genomic control mechanisms like histone regulation (Jylhävä, Pedersen and Hägg, 2017). Given the tissue type of our study's samples, and the interdisciplinary approach laid out in this thesis, the unresolved concerns are not sufficient justification to avoid epigenetic clocks as a tool to measure biological ageing across populations in this study's cohort.

2.2. Epigenetics

2.2.1. What is epigenetics?

Epigenetics studies the impact of heritable characteristics, and exposure to environmental factors, on an individual's gene expression, without any changes to their DNA sequence. Key epigenetic mechanisms include, but are not limited to, DNA methylation, histone modifications, and non-coding RNAs. It investigates how developmental signals, the environment and behaviours can impact the way our genes work. Unlike genetic changes, epigenetic changes can be reversible and do not change an individual's DNA sequence. However, they can change how an organism's body interprets a DNA sequence. Contrary to popular science, epigenetic research spans beyond psychological trauma and how the body stores these stressors – although insight into how socio-economic stressors may lead to epigenetic changes is valuable and will be explored further in the discussion chapter (Van der Kolk, 2014; Centers for Disease Control and Prevention, 2020).

As defined by Berger and colleagues, epigenetics involves the heritability of a phenotype passed on through mitosis or meiosis. According to Berger et al.'s findings, an epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence. Epigenetic mechanisms can be distinguished and separated into three categories of signals: the epigenator, which reads environmental cues and communicates via an intracellular pathway; the epigenetic initiator, which responds to the cue triggered by the epigenator and defines the precise location and environment for the establishment of the epigenetic pathway; and the epigenetic maintainer, which sustains the chromatin environment for generations to come (Berger et al., 2014). The structure and organisation of chromatin, like histone modifications and DNA methylation patterns are central to the field of epigenetics, particularly epigenetic regulation. In their paper, Berger et al. explain that the epigenator signal, or trigger, which is informed by the environment is received and converted into a signalling pathway that could be an interaction between proteins or a modification-based phenomena leads to the more dormant activity of the epigenetic initiator. The initiator in question, as explained by Berger and colleagues, could be a DNA-binding protein, a non-coding RNA or another protein capable of identifying the location of the chromatin structure that needs to be composed. The subsequent maintenance stage includes DNA methylation (the focus of this thesis), as well as histone variants, nucleosome positions and histone modifications. The main characteristic of maintainers is to carry epigenetic signals through cell cycles (Berger et al., 2014).

2.2.2. History of epigenetics

Epi from the Greek word for above, or over; and gene, from the Greek génos, meaning birth or generation. Above the gene. The term 'epigenetics' was first coined in the 1940s (Waddington, 1942). The main goal of the discipline was to provide a more explicit link between genotype and phenotype. Epigeneticists investigate the interplay between a living organism's environment and their biological composition; the field differs starkly from assumptions that motivate genetic determinists to carry out their line of work (Waddington, 1942). Waddington's conceptualisation of epigenetics proliferated at a time where the field of genetics was booming, and the study of inheritance revolved around Mendelian principles. And that is not without reason, Gregor

Mendel's work in the 19th century has been credited with laying down the foundation for modern genetics and the collective understanding of how traits are passed down from one generation to the next.

The Epigenotype, Waddington's three-page scientific paper, focused particularly on bridging the study of embryology with the field of genetics. As put forward by Waddington, an individual's environment acts as cues that go on to shape their development. The developmental landscape referred to metaphorically by Waddington suggests that cells roll down a multidimensional landscape guided by both genetic and environmental factors – until they reach stability during their development, which would correlate to becoming a specific cell type (Waddington, 1942). As well as coining epigenetics, canalisation and genetic assimilation are two important concepts put forward in the 1942 paper. Canalisation being the ability of consistent phenotype production during development despite environmental or genetic disturbances, while genetic assimilation is when a phenotypic characteristic, usually impacted by the environment an organism resides in (body hair, for example), becomes genetically encoded and subsequently an inherited trait. Waddington also put forward an argument that suggests traits acquired during the span of an individual's lifetime via environmental exposure could go on to become advantageous, as well as genetically inheritable. This idea was called the Baldwin Effect (Waddington, 1942). In his work Waddington aimed to summarise a new branch of biology focused on the relationship between gene and protein expression (Peixoto et al., 2020). Since the 1940s, the term epigenetics was used fleetingly by the scientific community – up until a surge in the 2000s and a peak in publications including the term in 2021 (PubMed data; retrieved March 2023).

2.2.3. DNA methylation and CpG sites

There are four types of bases in a DNA molecule: ACGT – adenine (A), cytosine (C), guanine (G) and thymine (T). DNA methylation (DNAm) works by adding a methyl-group from S-adenyl methionine (SAM) to the fifth carbon of a cytosine residue within the DNA, which forms 5mC. Typically, this base-modification is added to specific genomic regions on the DNA, where it blocks proteins that attach to the DNA, and it is responsible for interpreting the information in the gene. DNA methylation is the best-understood epigenetic memory system to date, and tools have been

developed to measure genome-wide levels of DNA methylation in humans (Horvath and Raj, 2018).

5mC can be removed through a process called demethylation. The process does not interfere with what protein is made, rather it impacts the gene expression. DNA methylation can lead to gene silencing. It involves the addition of cytosine bases in the DNA – at CpG dinucleotides. CpG islands are sections of the DNA rich in CpG dinucleotides that are placed near gene promoters (Christensen et al., 2009). CpG islands associated with gene promoters are rarely methylated (Moore et al. 2013). Hypermethylation of CpG sites can lead to gene silencing, whereas hypomethylation can lead to gene expression (Rakyan et al., 2010). The human genome has 28 million CpG dinucleotides. Technology developed to assess DNA methylation has prompted the identification of sets of CpG sites that best account for tracking biological aging (Horvath and Raj, 2018).

Gene promoters are specific DNA sequences that serve as initiation sites for the transcription of genes into RNA. They contain paramount regulatory components, which include the transcription start site (TSS) and binding sites for transcription factors. Their main role is to initiate the transcription of genes into RNA. DNA methylation in gene promoter regions can influence gene expression. High DNA methylation levels at gene promoter regions can lead to gene silencing or reduced gene expression. Focusing analysis on the interaction between CpG sites and gene promoter regions gives us a better understanding of how DNA methylation – and other epigenetic changes – can impact gene regulation and, potentially, health outcomes in individuals. **Enhancers** are short DNA sequences in non-coding parts of the genome, and they have an important role in regulating gene expression. These are important parts for the EPIC array to focus on because DNA methylation at enhancer regions can impact how accessible enhancers are to transcription factors. DNA methylation at these sites impacts gene regulation, which in turn affects other biological process and an individual's predisposition to negative (or positive) health outcomes. **Gene bodies** are regions within a gene that span from the TSS to the transcription termination site (TTS). They encompass both coding sequences (exons) and non-coding sequences (introns). Introns play a role in gene regulation but do not code for protein, whereas exons contain the information that is needed to synthesise a functional protein. The main

functionality of looking at gene bodies in DNA methylation analysis is to decipher methylation patterns within TSS and TTS. In addition to this, gene body methylation can impact splicing, which in turn may have a knock-on effect on the diversity of proteins generated from any given gene.

In their review of the discipline of epigenetics, Peixoto and colleagues highlighted that DNA methylation (5mC) is one of the better-known epigenetic markers because it is also the most stable epigenetic marker. DNA methylation was first described in 1965 but has since been researched intensively within the field of epigenetics (Scarano et al., 1965). In their review Peixoto et al. methodically explain the historical advancements made in the field of DNA methylation research, while explaining complex mechanisms involved in gene regulation and epigenetic processes. Their review recites the discovery of DNA methyltransferases (DNMTs) in 1988: DNMT1. DNMT1 plays an important role in maintaining DNA methylation levels stable during DNA replication and it also ensures the accurate transmission of its own epigenetic mark to daughter cells (Hervouet, Vallette and Cartron, 2010). Peixoto and colleagues also describe the significance of the discovery of 5mC binding proteins in the late 1980s and 1990s (MeCP2 and MBD1-4), as these proteins were identified to coordinate gene expression by interpreting the presence of 5mC. Disruption in processes have since been associated with disorders like Rett syndrome (Kriaucionis and Bird, 2003; Peixoto et al., 2020).

Other DNA methylation features such as de novo DNA methyltransferases have also played a significant role in the advances made to the contemporary landscape of epigenetics. It was only in 1998 that DNMT3A and DNMT3B – which are responsible for adding methyl groups to DNA independently of replication (de novo methylation) – were identified. Soon after this discovery, Okano and colleagues found that the two features are essential not only for de novo methylation, but also for mammalian development as it ensures gene regulation during embryonic development (Okano et al., 1999; Peixoto et al., 2020).

The discovery of DNA demethylation in mammals in 2009 was made possible by the discovery of 5-hydroxymethylcytosine (5hmC). Contemporarily, there is a shared understanding that DNA demethylation does – in very simplistic terms – the opposite of DNA methylation. That means, in the same way that increased DNA methylation levels at specific gene promoter regions may lead

to gene silencing, DNA demethylation in gene promoter regions can lead to the over-expression of some gene expressions that were either latent or silenced. Ito and colleagues discovered the ten-eleven translocation (TET) methylcytosine dioxygenases family of enzymes, which contribute to active DNA demethylation (Ito et al., 2010; Peixoto et al., 2020).

The enzymes described above are responsible for DNA methylation and demethylation, respectively. DNMTs are responsible for DNA methylation, as these enzymes catalyse the methyl group transfer (CH3) from cofactor SAM to the 5-carbon cytosine position. DNMT1 enzyme is responsible for DNA methylation patterns for DNA replication, whereas DNMT3A and DNMT3B enzymes are responsible for DNA methylation patterns during an individual's lifespan after key developmental stages (Ying and Chen, 2023). Three TET proteins (TET1, TET2 and TET3) have distinct roles in DNA demethylation during development and they are still present in some adult tissues (Ying and Chen, 2023). Environmental factors can affect the methyl group donors behind epigenetic changes and translocate epigenetic enzymes, such as TETs and DNMT, from cytoplasm to nucleus. These factors may also impact gene function, cell fate and epigenetic memory, through the changes made that impact gene transcription (Resendiz et al., 2023). In short, methylation turns genes 'off', while demethylation turns genes 'on'.

Altered DNA methylation can lead to some diseases associated with ageing. Methylation enzyme DNMT3B, for example, is known as an oncogene that can transform a healthy cell in a tumour cell and has been reported in colon cancer. Meanwhile methylation DNMT3A enzyme is a tumour suppressor in hematologic cancers, like lymphoma, myeloma and leukaemia. Also in hematologic malignancies, demethylation enzyme TET2, when altered, has been identified as a precursor to the onset of diseases in ageing populations (Field et al., 2018).

Other epigenetic mechanisms, as outlined earlier, include histone modification and non-coding RNA. Histone modifications occur when DNA wraps tightly or loosely around histones – proteins. Histones are always wrapped around histones, with the exception of DNA replication. When DNA is wrapped tightly around histones, it cannot be accessed by proteins that read the gene. Genes wrapped around histones are silenced, while genes that are not wrapped around histones are more pronounced. But the process is more complex than an on and off switch, as the tightness

can be regulated, and gene expression thereby fine-tuned. Chemical groups can be added or removed from histones and change whether a gene is unwrapped or wrapped - 'on' or 'off'. However, histone modifications are not always the result of epigenetic changes. Many histone modifications impact more dynamic biological processes such as transcriptional induction and DNA repair (Berger et al., 2009). DNA is used as a template to make coding and non-coding RNA. Coding RNA makes proteins, whereas non-coding RNA controls gene expression by attaching to coding RNA as well as other proteins to break down the coding RNA in a way that it can no longer be used to make proteins. In addition to this, non-coding RNA can recruit proteins to modify histones to turn genes 'on' or 'off'.

This study focuses on cytosine methylation (5mC) because it is a covalent, epigenetic modification that is not easily lost during the isolation process of DNA. Platforms have been developed to interrogate DNA methylation on samples from many individuals. This has led to standardised, robust data sets that lend themselves for statistical analysis. As outlined above, DNA methylation functions by adding a chemical group (5mC) to the DNA. The addition of this chemical group can be made directly to specific coordinates of the DNA, and it can inhibit the proteins that attach to the DNA to correctly identify the information from the epigenetic initiator or the epigenetic maintainer (Berger et al., 2014).

This mechanism is most useful for measuring DNA methylation levels at specific CpG sites near gene promoters relevant towards assess ageing. In addition to this, DNA methylation is the most studied modification in human population epigenetics. Its information content can be explored in two principal ways: epigenome-wide association studies and epigenetic age.

Bisulfte sequencing is a method of detecting DNA methylation. There is no alteration in the DNA sequence itself in epigenetics. DNA methylation is one of the most abundant phenomena in epigenetics. For most animals, DNA methylation occurs on the cytosine molecule (Bird, 2002). Acknowledging that there are four types of bases in a DNA molecule, bisulfite sequencing is a tool used to detect DNA methylation levels at single-base resolution. The main reaction in bisulfite sequencing is the chemical conversion of unmethylated cytosines (C) into uracils (U).

The conversion is carried out by treating DNA with sodium bisulfite, which reacts to convert cytosine into uracil. The method dates back to 1992, with the Frommer method (Frommer et al., 1992). Methylated cytosines (5mC) remain unaltered in this process. It is used to assess methylation levels in specific tissues. The process provides the basis for larger scale arrays, such as the Illumina EPIC array that was used in this study. Bisulfite-converted DNA is used in the EPIC array, and nuanced DNA methylation levels at thousands of CpG sites are assessed (Lister et al., 2009; Dedeurwaerder et al., 2014).

What are CpG sites?

CpG sites describe dinucleotides within DNA, where a cytosine nucleotide is followed by a guanine nucleotide in the linear sequence of bases along its 5' → 3' direction. CpG sites occur with high frequency in genomic regions called CpG islands (Bird, 1986). Bird's paper contributed towards the now commonplace understanding of CpG sites. The work put forward by Bird focuses on the concept of CpG-rich islands, and their functionality as regions of the genome that are enriched in both CpG sites and associated with gene promoter regions. Promoter regions are important for the epigenetic initiator signal that discussed earlier, they also play an important role in influencing whether genes are expressed or silenced (through demethylation and methylation, respectively). CpG islands are often heavily studied because of potential links that may be drawn between DNA methylation patterns and diseases. CpG sites selected to be analysed by epigenetic clocks are dependent on the main research aim from the group that develops the mathematical model. If researchers are interested in investigating the epigenetic age estimates across a particular tissue type, or to draw conclusions about the correlation between lifestyle factors and epigenetic age acceleration, then different epigenetic clocks will be more suited for each endeavour (Bird, 1986; Horvath and Raj, 2018).

2.2.4. What is epigenetic ageing?

Epigenetic clocks are mathematical computational models that use DNA methylation sites to predict age (Johnson et al., 2022). An individual's epigenetic age is based on biomarkers, such as patterns of DNA methylation, and it is often distinct from their chronological age (measured from when they are born onwards). This discrepancy is often associated with an individual's lifestyle

and the environment they grew up in – or are exposed to during key developmental phases in their life cycle.

Epigenetic ageing aids our understanding of the molecular mechanisms behind biological ageing. DNA methylation plays an important role in assessing epigenetic ageing. However, other biomarkers such as telomere shortening, cellular senescence and chronic inflammation and oxidative stress can help build a more comprehensive picture of the ageing processes. (Franceschi et al., 2000). The more diverse the sources of information about an individual's biological ageing, the more comprehensive the picture of that individual's health outcomes will be. For instance, if both the DNA methylation levels based on an individual's buccal cells and the telomere length of their blood cells are known – then we have a more comprehensive notion of their biological composition had we only been privy to their DNA methylation information.

What is epigenetic age acceleration?

Epigenetic age acceleration is a form of describing the discrepancy between an individual's epigenetic age estimate – based on DNA methylation on specific CpG sites – and their chronological age. Those whose epigenetic age is higher than their chronological age exhibits positive epigenetic age accelerations, and those whose epigenetic age is lower than their chronological age experience negative epigenetic age acceleration (Horvath, 2013). However, we must not be deceived by the linguistic association of the two terms, as those who exhibit positive epigenetic age acceleration may be more prone to morbidities and be exposed to environments that lead to premature ageing. Interesting observations based on the DNA methylation measured across different tissues and cell types: female breast tissue is older than other parts of the body (Sehl et al., 2017), while the cerebellum ages more slowly than other parts of the brain (Horvath et al., 2015a).

DNA methylation as a tool to assess epigenetic ageing

DNA methylation has been identified as the most accurate way to measure epigenetic age estimation. Epigenetic clocks are the culmination of sets of CpG sites that were identified as relevant biomarkers to track biological ageing, and a mathematical algorithm that estimates age

in years based on what the model estimates the biological age of the data at its disposal to be (Horvath and Raj, 2018). It operates on a supervised machine learning method, therefore the more information that is processed by a clock, as researchers enter diverse datasets into the programme, the more accurately it will be able to estimate epigenetic ages based on DNA methylation levels at specific CpG islands (Christensen et al., 2009). Each epigenetic clock focuses on different CpG sites and unlike other biomarkers of ageing, it is not a specific CpG site that provides information on the biological age of an individual, rather it is the culmination of a group of CpG sites that – when analysed collectively – can provide a comprehensive understanding of the correlation between relevant DNA methylation information and age. As described in the pages that follow, Horvath's clock looks at 353 CpG sites while Hannum's clock provides predictions based on methylation levels at 71 CpG sites, and the PedBE clock looks at 94 CpG sites. These clocks are machine learning mathematical models developed based on training datasets particular to the parameters and biomarkers deemed most important for each study.

DNA methylation-based age estimators require the statistical prediction method, outcome measure and the covariates. Statistical prediction methods include penalised regression. A penalised regression is used to improve the accuracy of DNA methylation age prediction mathematical models. By using the least absolute shrinkage and selection operator (lasso), epigenetic clocks can exclude CpG sites that do not add as much information – by shrinking some coefficients to 0 - when it comes to calculating age estimates. It also helps epigenetic clock extend predictions from training models to new datasets (Tibshirani, 1996). Lasso penalised regression is one of the main components used to develop epigenetic clocks because it allows for a more accurate, interpretable, and generalised prediction of age estimates across populations. Given the large number of CpG sites comprised in DNA methylation data, Lasso penalised regression allows for more efficient selection of the CpG sites that will be analysed in any given clock. It also prevents the overfitting of irrelevant data and ensures clocks are less sensitive to variations in DNA methylation data, that is because it was designed to predict an outcome dependent on the input features – which in this case are the DNA methylation levels for each CpG site. The penalised regression model ensures that the most informative CpG sites, for the purpose of

accurately predicting age estimates, are selected. This in turn means that the output of the clock through supervised machine learning analysis gives a set of CpG sites and a corresponding mathematical algorithm that provides an age estimate based on DNA methylation levels (Horvath and Raj, 2018).

Outcome measure is a surrogate marker of biological age; a quantifiable parameter that can substitute an individual's biological age. Covariates are additional variables, also known as biomarkers, that are considered alongside other factors in the analysis of subsets of CpG sites within DNA methylation-based clocks. Depending on the criteria of a specific clock, the other factors can include gender, cell type, smoking status among other specific biomarkers. Epigenetic clocks aid our understanding of how prone an individual is to experience biological impacts of ageing – disease and mortality rates – despite their chronological age. Epigenetic clocks are often described as the most impressive of biomarkers of ageing (Ryan, 2021; Caudill et al., 2020).

First-generation epigenetic clocks

Bocklandt and colleagues developed the first DNA methylation age estimator. The mathematical model was calculated based on DNA data gathered from monozygotic twin humans' saliva. The team identified 88 CpG sites in or near 80 genes where the degree of cytosine methylation was significantly correlated with the age in the buccal samples of male identical twins aged between 21 and 55 years (Bocklandt et al., 2011). Titled the epigenetic predictor of age, Sven Bocklandt and colleagues developed a machine learning model after quantifying the methylation status of 27,578 CpG loci covering at least 14,000 genes at single-nucleotide resolution in saliva samples from 34 identical twins. The team used Illumina HumanMethylation27 microarrays and for each CpG site on the microarray calculated the beta value, which expresses the fraction of methylated cytosines in each location examined. The first of its kind, and setting the precedent for epigenetic clocks to come, Bocklandt et al. determined that completely methylated CpG sites on both alleles in all cells would have a beta value of 1 and completely unmethylated sites would have a value equal to 0.

In their quest to identify the CpG sites relevant to investigating the relationship between DNA methylation levels and ageing, Bocklandt and colleagues calculated q-values to correct multiple

comparisons and better explain in which loci the methylation values correlate positively or negatively with age. The findings returned 88 CpG sites correlated with age and corresponding to 80 genes spread over many of the modules. 19 of the probes were negatively correlated with age, while 69 were positively correlated with age. 73 of the 88 probes were within CpG islands – which are CpG sites near gene promoters and therefore relevant locations to investigate the impact of DNA methylation on gene expression. The clock developed by Bocklandt and colleagues uses Lasso penalised regression as well, which was performed using the 'penalized' package of the R coding programme. Bocklandt et al. carried out all statistical analyses and data processing using the statistical package of R, which in 2011 meant version 2.11.1.

Hannum et al.'s clock was developed based on DNA methylation in blood – based on 71 CpG sites. Age estimates from Hannum's clock are meant to account for adult blood samples. Although we ran our dataset through Hannum's clock, we predicted that age estimates would not be as accurate as those from Horvath's and McEwen's clocks – as the latter are made either for all ages or paediatric samples in particular, respectively (Horvath, 2013; McEwen et al., 2020). The findings from our study, which corroborate such hypothesis, can be found in the results chapter. Hannum's epigenetic clock shows high age correlations and has an R value of 0.91. The mathematical model used to predict epigenetic age estimates in Hannum's clock was developed based on samples from 656 individuals from ethnically diverse populations based on CpG site information gathered from the Illumina 450K array (Hannum, 2013).

The first multi-tissue epigenetic clock

Evidence suggests that although it is possible to have accurate DNA methylation age estimates from clocks based on a restricted number of CpG sites, there is a large body of evidence that indicates larger CpGs sets can produce better age estimates – more accurate and robust (Horvath and Raj, 2018, Lin et al., 2016).

Some 8,000 open-source microarray samples from more than 30 different tissue and cell types were collected from both children and adults to develop **Horvath's clock** (Horvath, 2013). The

study found that a penalised regression model, using a transformed version of chronological age showed that the rate of change in the DNA methylation age estimate is faster during growth and development (Horvath and Raj, 2018; Horvath 2013). The clock provides epigenetic age estimates based on 353 CpG sites. From those CpG sites, DNA methylation levels in 193 CpGs were positively related to ageing and get hypermethylated with time, while 160 (of the 353 CpG sites) are negatively correlated with age and get hypomethylated with time. The 193 positively correlated with age CpGs are likely to be in poised gene promoter regions, while the negative-correlated CpGs are likely in weak promoter regions or strong enhancers. The clock was trained on data generated by Illumina 27K and 450K arrays, but it remains applicable to the Illumina EPIC array despite the absence of 17 CpG sites (Horvath and Raj, 2018; Horvath 2013). The clock applies to all tissues in the human body that contain DNA, except for sperm cells. It has high age correlations and has been categorised as one of the most robust predictors of chronological age based on DNA methylation analysis (Horvath, 2013; Jylhävä, Pedersen and Hägg, 2017).

A second-generation of clocks: environmental factors and clinical characteristics

Lu et al. embarked on a study to predict mortality based on DNA methylation data on 59 CpG sites, the clock was called **GrimAge.** It is a composite biomarker of seven DNA methylation surrogates and a DNA methylation-based estimator of smoking pack-years. The GrimAge clock identified that in most instances the DNA methylation-based surrogate biomarkers, such as smoking pack-years, were better predictors of mortality than the observed and self-reported biomarker. DNA methylation-based GrimAge was correlated with lifestyle factors and the research group found that the age estimates from their 59-CpG strong mathematical model correlated with lifestyle factors and other age-related conditions. Lu et al. Also found that DNAm GrimAge is associated with age-related changes in blood cell composition and leukocyte telomere length. The training dataset for the clock was based on blood samples from 2,356 individuals (Lu et al., 2019; Föhr, 2021; Noroozi et al., 2021).

The PhenoAge clock was developed by Levine et al. on a weighted average of 10 key clinical characteristics – chronological age, albumin, creatinine, glucose and C-reactive protein levels, lymphocyte percentage, mean cell volume, red blood cell distribution width, alkaline

phosphatase and white blood cell count (Levine et al., 2018; Horvath and Raj, 2018). The age estimates are based on DNA methylation from 513 CpG sites. As illustrated in **figure 1,** it goes one step further than the first-generation epigenetic clocks because it differentiates morbidity and mortality risk from individuals that have the same chronological age (Horvath and Raj, 2018).

The surrogate marker of biological age in the PhenoAge clock is not chronological age. Although largely successful, surrogate measures of biological age that are not chronological age can be challenging due to the dangers of confounding (Horvath and Raj, 2018; Newman, 2015). The age prediction from the PhenoAge clock is linked to lifestyle and demographic variables like physical exercise, income, systolic blood pressure, body mass index (BMI) and nutritional intake (Horvath and Raj, 2018; Levine et al., 2018). Age acceleration as measured by PhenoAge has strong correlations with certain gene expression levels. For example, positive PhenoAge acceleration was linked to an increased activation of pro-inflammatory and interferon pathways and decreased activation of the DNA damage response and nuclear mitochondrial signatures (Horvath and Raj, 2018; Levine et al., 2018).

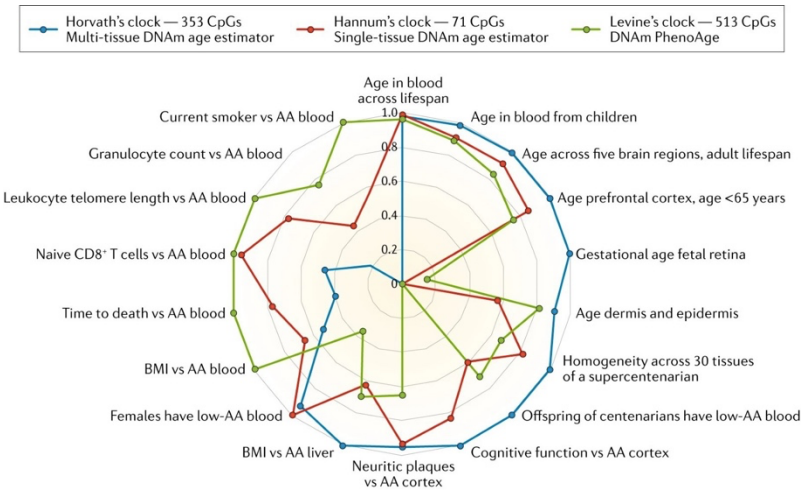


Figure 1. Age acceleration measured across three epigenetic clocks (Horvath, Hannum and Levine) plotted against clinical biomarkers and environmental factors. The figure illustrates that Horvath's multi-tissue clock more accurately predicts features such as the age of the dermis and epidermis, and the age in blood from children, whereas Levine's PhenoAge clock more accurately predicts whether the sample analysed comes from a smoker and what the BMI of the individual might be. Hannum's clock, as illustrated above, most accurately indicates that females have low age acceleration based on blood samples and more broadly it also predicts the age in blood across lifespan. The figure was adapted from Horvath and Raj's 2018 paper (Horvath and Raj, 2018).

Also in 2018, the **skin and blood** DNA methylation-based epigenetic clock was designed. The clock's main aim was to address the challenge posed by sub-optimal accuracy in DNA methylation-based age predictions for fibroblasts and other cell types used in ex vivo studies across other clocks – which were trained with tissue and organ data. Horvath et al.'s skin and blood clock was described by the team as a highly robust epigenetic age estimator, as it accurately tracked the dynamic ageing of cells cultured ex vivo and was able to reveal that their proliferation was accompanied by an incremental increase in epigenetic age. It was based on 391 CpG sites (Horvath et al., 2018).

Once developed, the skin and blood clock was applied to Hutchinson Gilford Progeria Syndrome (HGPS), as well as ex vivo studies. HGPS is a rare genetic disorder characterise by unusually rapid ageing during an individual's formative years. Children who have HGPS exhibit symptoms associated with old age, like wrinkled skin, cardiovascular problems, skeletal abnormalities and hair loss. In developing a clock tailored to fibroblasts, as well as other cells used in ex vivo studies, Horvath et al. hoped to gain a better understanding of the ageing process linked to HGPS, while also investigating the underlying epigenetic mechanisms that underly the syndrome. The key findings from the study demonstrated that individuals who had HGPS had much higher DNA methylation-based ages than their chronological age and the age acceleration value was also considerably high – particularly high in skin tissue (Horvath et al., 2018).

PedBE clock

Horvath's clock predicts epigenetic ages based on DNA methylation data and works with impressive accuracy across most tissues. In adults, the age acceleration results it produces can be associated with various environments and the health of individuals. However, it is not very accurate when it concerns paediatric samples, which is the focus of this study, namely because DNA methylation changes at a faster rate in children (McEwen et al, 2020).

The PedBE clock works as a precise tool to estimate the epigenetic age of paediatric individuals based on their buccal samples. The clock is tailored especially for buccal epithelial cells, which

are known to predict epigenetic ageing more accurately (McEwen et al., 2020). The clock has the potential to become the standard reference for epigenetic studies relevant to child development across the spectrum of health and disease. One potential outcome of this clock is that it could help shape our collective understanding of how environmental and socio-economic factors can shape the which genes are on or off during child development – which can in turn give us a better picture of a child's health and proneness to disease and premature mortality.

Part of the reason for the increased accuracy of the PedBE clock comes from it building on from the predictor in a single target tissue within a focused age range. Unlike the PedBE clock, Horvath's epigenetic clock is agnostic to tissue type, which sacrifices some precision when estimating age from samples gathered from a single tissue type – like buccal epithelial cells (McEwen et al, 2020). But the PedBE clock is not without its limitations. It looked at a limited sample size. There was a distinction in the sample used to produce the clock - Horvath's clock looked at 8,000 samples, whereas McEwen et al.'s PedBE clock focused on buccal epithelial cell samples gathered from 1,721 healthy individuals. The 1,721-genome wide DNAm profiles were from 11 different cohorts of developing individuals aged 0 to 20 years old. Horvath's epigenetic clock looked at 353 CpG sites, whereas the PedBE clock looked at 94 CpG sites (McEwen et al., 2020; Horvath, 2013).

To validate the tool McEwen et al. applied the PedBE clock to the test dataset, which revealed a correlation between chronological age and paediatric DNA methylation age. The study was largely successful. The R value in two scenarios were 0.99 and 0.98, which indicates that the clock is incredibly accurate at predicting the age of samples based on DNA methylation levels from the samples' buccal epithelial cells.

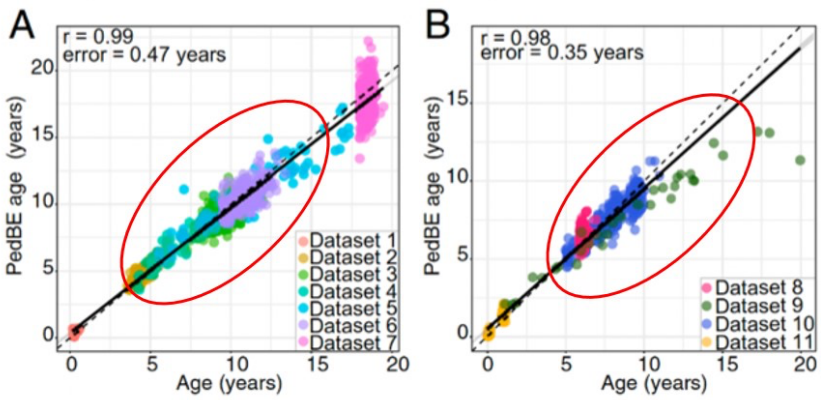


Figure 2. Linear regression graph shows the DNA methylation-based age plotted against the chronological age of paediatric subjects in the training dataset of the PedBE clock. The figure shows the distribution of age predictions using the paediatric buccal DNA methylation clock across 11 different datasets. Illustrated in two graphs due to the volume of data, the graph on the left (A) has a strong R value of 0.99 and the error margin is 0.47 years. These statistics are slightly different when we look at datasets 8 to 11, with the R value of 0.98 and an error margin of 0.35 years. The solid red oval shape focuses on the chronological age range (5 to 16 years) relevant to the study carried out in this paper. The solid black and dashed lines show the linear regression and perfect correlation, respectively. The figure above was adapted from McEwen et al.'s 2020 paper (McEwen et al., 2020).

The PedBE clock managed to reduce variation when compared to Horvath's clock. The median absolute difference between DNAm age and chronological age in Horvath's clock was 1.73 years, while the PedBE clock's median absolute difference between DNAm age and chronological age was 0.35 years, as shown in **figure 2**. This scenario is exactly how researchers desire the data from their control group to behave, which is why it becomes so interesting to use the clock to measure for different factors relevant to an individual's development – such as the impact of migration, as well as internalising disorders and maltreatment categories (Dammering et al., 2021). The difference between PedBE age and chronological age in the test dataset correlated significantly with that of the pan-tissue Horvath DNAm age difference (McEwen et al., 2020; Horvath, 2013).

Key questions for researchers who choose to further research epigenetic ageing and better understand the role of DNA methylation-based epigenetic clocks in current gerontology studies

concerns the significance of clock CpG sites. As discussed by Horvath and Raj in their review, now that there is a robust body of literature and mathematical models focused on the investigation of epigenetic ageing based on DNA methylation at key CpG sites, we must understand whether methylation changes in these CpG sites are ***driving*** ageing or if they are ***driven by*** ageing. Longitudinal studies can help address these questions, as well as in vitro studies and studies in model organisms.

As outlined in the description of the different epigenetic clocks above, to measure the rate of epigenetic ageing, these clocks analyse a set of CpG sites that have been identified over time to best predict the relation between DNA methylation and ageing. A recent study by Pipek and Csabai identified five CpG sites, present in Horvath's clock, that are highly indicative of individuals' chronological ages. As they presented their findings for four separate mathematical models to predict ageing, cg00864867, cg06993413, cg14424579, cg16241714 and cg22736354 were hypermethylated in the elderly and hypomethylated in younger people. Interestingly, the five CpG sites are located in genes involved either in apoptosis, autophagy, metabolic processes, transcription or cell differentiation – PAWR, DPP8, AGBL5, CEBPD and NHLRC1 respectively (Pipek and Csabai, 2023).

Identifying common CpG sites across different epigenetic clocks is an important task, as it can provide researchers with a better understanding of which particular CpGs, if any, are relevant towards predicting the biological age of individuals. A study published in 2022 by Johnson and colleagues identified common CpG sites across 11 epigenetic clocks; some of these clocks were used for the analysis of this thesis (Johnson et al., 2022). Johnson et al. managed to identify common CpG sites across Horvath's clock, the skin and blood clock, the PedBE clock and the PhenoAge clock:

- 12 common CpG sites were used in the PedBE clock and the skin and blood clock
- 1 common CpG site present in Horvath's clock and the PedBE clock
- 0 common CpG sites present in the PedBE clock and the PhenoAge clock
- 41 common CpG sites in Horvath's clock and the PhenoAge clock
- 58 common CpG sites in the PhenoAge clock and the skin and blood clock

- 60 common CpG sites in the skin and blood clock and Horvath's clock

Horvath's clock, the skin and blood clock and the PhenoAge clock all identified cg04084157 as one that becomes hypermethylated with age (Johnson et al., 2022; Horvath, 2013; Horvath et al., 2018; Levine et al., 2018). In the PedBE clock and the skin and blood clock, common CpG sites cg16867657 and cg21572722 were identified to undergo hypermethylation with age and they are linked to the ELOVL2 gene, which lengthens long-chain polyunsaturated fatty acids, has been identified as an important biomarker of ageing (Johnson et al., 2022; McEwen et al., 2020; Horvath et al., 2018). Although the focus tends to be on the hypermethylation of CpG sites, there are instances where a hypomethylated probe has been associated with ageing. Common CpG site cg05442902, present in Horvath's, skin and blood and PhenoAge clocks, is an example of this phenomenon (Johnson et al., 2022; Horvath, 2013; Horvath et al., 2018; Levine et al., 2018).

Transcription factors

Transcription factors are effectors of DNA methylation, and their DNA binding dynamics as well as activation and repression are incredibly varied, dependant on the presence of CpG in cognate binding sites, the density of CpGs, and the relative level of methylation of these CpGs (Héberlé and Bardet, 2019). Proteins belonging to the zinc finger family are among the most studied transcription factors sensitive to CpG methylation. These transcription factors control a vast range of cellular processes in mammalian cells, from proliferation to apoptosis and increased metabolic rate to cellular senescence, all of which may play a role in epigenetic, and therefore biological, ageing (Zhou et al., 2018).

2.2.5. Environmental factors and epigenetic ageing

Stress impacts age acceleration on DNA methylation-based clocks; but in a targeted way. That is to say, metabolic stress due to obesity puts an additional strain on the liver tissue, and HIV infection accelerates the rate of epigenetic ageing in blood and brain tissues (Quach et al., 2017; Horvath and Levine, 2015; Horvath et al., 2015a). Meanwhile, semi-supercentenarians, individuals aged between 105 and 109 years, were identified representing the gold standard of successful human ageing, according to a study carried out by Horvath and colleagues. The focus

of the study was on Italian semi-supercentenarians and the findings were based on epigenetic age estimates carried from a DNA methylation-based clock. Horvath et al. concluded from DNA methylation levels data of peripheral blood mononuclear cells from Italian families that the offspring of semi-supercentenarians have a lower epigenetic age estimate than age-matched controls (Horvath et al., 2015b). Their findings suggest that successful ageing can be hereditary.

Higher education levels and income are linked to lower extrinsic age acceleration, these characteristics are often linked to an individual's environmental factors and may impact their extrinsic age acceleration. Environmental factors do not impact intrinsic age acceleration in the same way, and Quach et al. suggest that is because the intrinsic age acceleration of individuals is under strong genetic control (Quach et al., 2017; Lu et al. 2019). A recent review carried out by Mc Auley investigated the relationship between advances in the field of epigenetics and biogerontology – an interdisciplinary field of research that bridges geroscience studies with medicine, public health and sociology. Mc Auley's review found that there is a demand for a more pluralistic perspective when assessing the intersection between ageing, evolutionary theories and epigenetics. That is because the shared understanding and clear dialogue across disciplines will allow for an improved understanding of epigenetics and where it sits within conceptual theories of ageing – particularly the evolution of ageing (Mc Auley, 2023).

Adverse childhood experiences (ACEs) and epigenetic ageing

Joshi et al.'s research, published in 2023, suggests that early life adversity has the potential to become biologically embedded and lead to accelerated aging (Joshi et al. 2023). The researchers used GrimAge and PhenoAge epigenetic clocks to assess in a longitudinal study whether the rate of epigenetic aging in individuals was impacted by what they could disclose about their childhood experiences. Utilising self-reported adverse childhood experiences (ACEs), the study looked at the DNA methylation data of 1,445 individuals, who were aged between 45 and 85 years. ACEs exposure prior to an individual reaching 16 years was assessed in self-reported questionnaires in the cohort. Interestingly, Joshi et al.'s findings reported that there was a correlation between ACEs and accelerated ageing when measuring the DNA methylation using the GrimAge clock –

but not with the PhenoAge clock. The PhenoAge clock predicts the phenotypic age of individuals based on DNA methylation by looking at 10 measures, which include lymphocyte percentage, albumin and glucose levels, and chronological age (Bergsma and Rogaeva, 2020; Levine et al., 2018).

Since the development of McEwen's PedBE clock, several subsequent studies – including the work carried out in this thesis – have worked towards gaining a better understanding of how childhood environments can impact the epigenetic age acceleration rate among children of the same chronological age, living under significantly divergent circumstances (Joshi et al., 2023; McGill et al., 2021; Musci et al., 2023; Dammering et al., 2021). The environment organisms develop in is an important factor throughout the entire lifespan, across flora and fauna. Thereby, McGill and colleagues embarked on a study to assess the impact of maternal prenatal anxiety on their offspring, and to gain a better understanding of the foetal origins of epigenetic ageing. McGill and colleagues used McEwen's PedBE clock to provide a robust index of the biological characteristics of prenatal adversity (McEwen et al., 2020). The study looked at longitudinal cohorts from the Netherlands and Singapore. Both groups provided data on prenatal maternal anxiety and longitudinal assessments of buccal DNA samples and their methylation levels. The findings suggest that prenatal maternal anxiety predict child PedBE epigenetic age acceleration across both groups – but the effects were restricted to males. The PedBE epigenetic clock predicted increased externalising symptoms in males from the Netherlands as they reach their mid to late childhood. McGill and colleagues concluded from these findings that it is important to provide timely and effective mental health support to pregnant women, as the impacts may permeate on to their offspring (McGill et al., 2021).

Musci and colleagues published a systematic review that investigates the role of epigenetic clocks as tools to characterise biological ageing studies of children and childhood exposures. The research proposed that biological age be measured via epigenetic clocks, in a similar way to this thesis. Their work focused on how useful epigenetic clocks are for prevention scientists and it found that the majority of the studies it reviewed – focusing on the Hannum clock, the Horvath clock and the PedBE clock – found significant associations between the epigenetic clock used and the health outcome of interest. From their findings, Musci et al. argued that epigenetic clocks

should be used as a tool to understand the impact environmental exposures have on biological ageing pathways and development in early years of life. They also argue that these clocks can aid the monitoring of the effectiveness of preventative care interventions (Musci et al., 2023).

Internalising disorder and accelerated epigenetic ageing

Dammering et al. applied the PedBE clock to investigate whether there is any significant ageing acceleration in children with internalising disorder and maltreatment exposure Dammering et al. 2021) - the clock identified significant ageing acceleration in children living under those circumstances. As part of an interdisciplinary study, Dammering et al. looked at the impact of adverse childhood experiences on the epigenetic age and subsequent health and well-being of individuals. The research team hoped to establish a correlation between internalising disorders, maltreatment severity paediatric samples experienced and their epigenetic ages. From the data collected, researchers hope to inform and devise early targeted interventions for children and vulnerable to maltreatment (Dammering et al. 2021).

Internalising disorders are characterised by anxiety, depressive and somatic symptoms. Meanwhile, externalising disorders are characterised by impulsive, disruptive conduct, substance use, and other addictive symptoms (American Psychiatric Association, 2013: 13). Adverse childhood experiences include trauma; maltreatment includes neglect, physical abuse, and emotional abuse. To carry out their study, Dammering et al. used a paediatric buccal epigenetic clock – devised by McEwen's research team – to predict DNA methylation-based ageing deviation in children with and without internalising disorder. Once they had this data, Dammering and colleagues looked to assess the effect of maltreatment exposure (Dammering et al., 2021).

Dammering et al. isolated DNA samples from the saliva of 158 paediatric individuals: 73 girls and 85 boys. The mean age of the samples was 4.25 years. The research team assessed the contribution of glucocorticoid signalling to results from the PedBE clock. As mentioned earlier on in this paper, the age estimate was based on DNA methylation across 94 CpG sites of the PedBE clock. The epigenetic ages from samples were analysed against the children's chronological ages, as well as differences between children with and without internalising disorder. Maltreatment was coded in three different severity levels and entered in a moderation model. Genome-wide

dexamethasone-responsive CpG sites were derived from an independent sample and enrichment of these CpG sites within the PedBE clock was identified. As outlined earlier a thorough understanding of the relationship between stress, cortisol and epigenetic ageing can provide helpful explanations on the impact of stressors on biological ageing (Suor et al., 2015).

Dammering et al.'s study found that children with internalising disorder exhibited significant acceleration of epigenetic ageing, in contrast to children without internalising disorder. This association was also proportionate to the maltreatment severity. Children with internalising disorder who were not exposed to maltreatment did not show epigenetic ageing acceleration. The study also found that significant enrichment of dexamethasone responsive CpG sites relevant to the PedBE clock. Among the PedBE clock's 94 CpG sites 18 (19%) were responsive to dexamethasone (Dammering et al. 2021).

Glucocorticoids, stress and epigenetics

Glucocorticoid hormones are the main effectors of stress response and abnormal glucocorticoid signalling has been associated with an increased risk for psychiatric and mood disorders, including schizophrenia, post-traumatic stress disorder and depression (Mourtzi et al, 2021). Dexamethasone-responsive CpG sites are responsive to environmental stressors and heightened cortisol levels. DNA methylation patterns can be shaped by cortisol, particularly genes and CpG islands associated with stress response (Turecki and Meaney, 2016).

Cortisol interferes with histone modifications and can modulate gene expression by triggering histone modification processes such as methylation, acetylation, phosphorylation and ubiquitination. In the case of non-coding RNAs, cortisol-induced changes to microRNA expression can in turn impact biological processes (Hunter and McEwen, 2013). Dysregulation of core epigenetic processes can contribute to the development of stress-related disorders, which can in turn accelerate biological ageing.

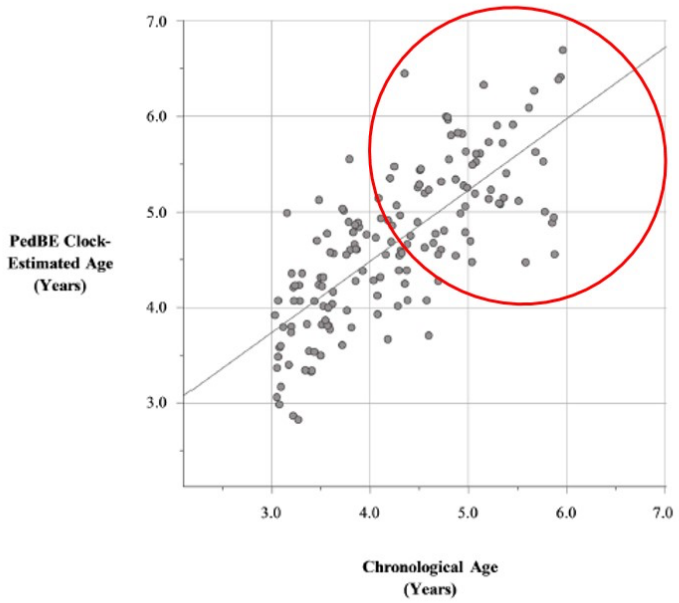


Figure 3. Estimated epigenetic ages from the PedBE clock plotted against the chronological ages in years of children involved in Dammering et al.'s study. The subjects analysed in the study carried out by Dammering et al. had varying degrees of internalising disorder and exposure to adverse childhood environments. The bold red shape focuses on the data gathered from individuals that were aged 5 years or more, as the information from that age group more closely connects to the age group analysed in this study (5 to 16 years). The measurements were carried out in a controlled setting and the PedBE clock was designed to factor the discrepancy between epigenetic age estimates and chronological ages during early years. The R value for Dammering et al.'s research was 0.745. Adapted from Dammering et al.'s 2021 research paper (Dammering et al., 2021).

As outlined above, Dammering et al.'s study looked at the correlation between exposure to maltreatment – which can increase cortisol levels due to stress – and epigenetic ageing across a paediatric cohort. As illustrated in **figure 3**, the PedBE clock was used in Dammering et al.'s study to measure the epigenetic age estimates of a cohort of children with varying degrees of internalising disorder and propensity to age more prematurely than their counter parts. The adapted figure has a bold circular red shape to highlight the data coming from individuals aged 5 and older, as that more closely correlates to the cohort in this study. It is interesting to note that the age estimates largely correspond to the chronological age, but the R value was 0.745 (Dammering et al., 2021). **Figure 4** shows that children in Dammering et al.'s study identified to

have internalising disorder exhibited a wider, and positive, range of epigenetic ageing deviation from their chronological ages. Whereas their counterparts without internalising disorder had a smaller and negative range of ageing deviation.

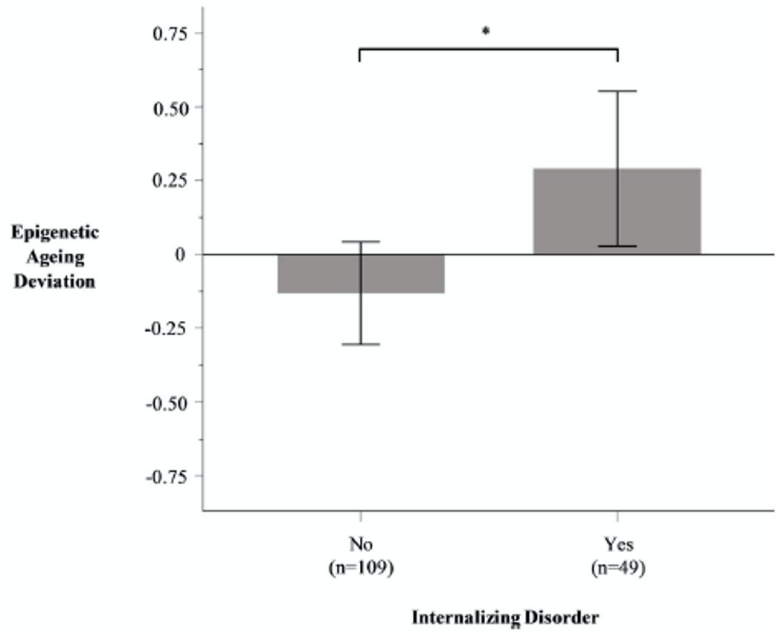


Figure 4. Box plot graph that shows the covariance of standardised residuals of the PedBE clock age estimates regressed against chronological age in children with and without internalising disorder as part of Dammering et al.'s study. Figure adapted from Dammering et al.'s 2021 research paper (Dammering et al., 2021).

Covariance of standardised residuals. Covariance means the mean value of the product of the deviations of two variates from their respective means. Standardised residuals are a measure of the strength of the difference between observed and expected values. Observed values are the epigenetic age, whereas chronological values are the expectation. This is particularly relevant in the PedBE clock context because it illustrates the discrepancy between biological ages across individuals with different lived experiences. The P-value is probability of obtaining test results at least as extreme as the results that were observed, under the assumption that the hypothesis put forward is accurate. The higher the number under the maltreatment category, the more alarming the discrepancy between an individual's chronological age and their epigenetic age. Out of the pool of 158 children - 73 girls, 85 boys - 77 experienced no maltreatment. Some 55 fell

under categories 1 or 2 and 26 fell under category 3. The categories were: "No maltreatment (value = 0, n = 77), 2) Mild to moderate maltreatment: 1 to 2 maltreatment categories (value = 1, n = 55), and 3) Severe maltreatment: 3 or more maltreatment categories (value = 2, n = 26)" (Dammering et al, 2021).

Although the PedBE clock can predict epigenetic ageing in a paediatric sample more accurately than Horvath's clock, the interpretation Dammering et al. draw from their findings appears to differ significantly from McEwen et al.'s. In adults, positive DNAm age acceleration from the Horvath pan-tissue clock has been associated with increased risk for certain diseases, mortality, frailty, and other negative outcomes, DNAm age acceleration in children may not follow a similar pattern, in that accelerated age deviation could potentially reflect positive outcomes. McEwen et al. suggest that age acceleration in paediatric samples may be an indicator of progressive development through key milestones, whereas deceleration is seen as a negative feature, and it might indicate delays in maturation.

Gender and epigenetic ageing

An important consideration for researchers that would like to focus on longitudinal analyses that follows the development of cohorts is to understand the impact of biological ageing in later life, particularly in a female population. Although there are several empirical findings that suggest women age epigenetically slower than their male counterparts and can go on to live longer lives (Horvath and Raj, 2018), their quality of life in a society that actively erases the elderly may be challenging. A study by Chrisler and colleagues investigated the hazardous impact of ageism on women's health outcomes and how sexism and stereotypes of older women play out in the healthcare system (Chrisler, Barney and Palatino, 2016). The intersection of misogyny, ageism and frailer bodies may lead to detrimental and unsatisfactory diagnosis and prognosis, thereby negating the slower pace of epigenetic ageing across the female population as prejudicial societal structures can decrease confidence in health professionals and lead to negative health outcomes. In addition to these societal pressures, a study by Levine and colleagues found that menopause

accelerates biological ageing based on DNA methylation-epigenetic clock analyses carried out using both blood and saliva samples (Levine et al., 2016).

Ethnicity and epigenetic ageing

In 2016 Horvath et al. embarked on an analysis of the relationship between ethnicity, sex, and coronary heart disease (CHD) (Horvath et al., 2016). The study found that epigenetic ageing rates were significantly linked to sex, ethnicity, and coronary heart disease – although CHD correlations were to a lesser extent. The research team chose to categorise extrinsic and intrinsic levels of epigenetic ageing. As defined by the researchers investigating these correlations, intrinsic measures of epigenetic age acceleration in blood concern the residual from a multivariate regression model of DNA methylation age against chronological age and certain blood immune cell counts. Horvath et al. recognise that it is an incomplete measure because it does not account for age-related changes in the composition of blood cells – like the decrease of naïve CD8+ T cells (Horvath et al., 2016). Extrinsic measures of acceleration in blood apply to the entire blood and measures epigenetic ageing in immune-related components, while intrinsic measure focus on the numbers of specific immune cells change as individuals age (Horvath, 2013). The study used Horvath's multi-tissue epigenetic clock to assess whether there were any discrepancies across groups, and it found that Hispanics and Tsimane Amerindians had a lower intrinsic epigenetic age but a higher extrinsic epigenetic age than their Caucasian counterparts. Whereas African Americans had lower extrinsic epigenetic ageing rates than both Caucasians and Hispanics, but there were no significant differences for the intrinsic measure (Horvath et al., 2016).

In a similar study, Galanter et al. concluded that different methylation levels between ethnic groups can be explained by their shared genetic ancestry. But the team added that environmental factors associated with ethnicity – such as exposure to racist and hostile environments – are likely to explain ethnicity-associated methylation changes that cannot be encompassed solely by genomic ancestry (Galanter et al., 2017).

Disease and epigenetic ageing

Research undertaken by Zheng, Widschwendter and Teschendorff looked into epigenetics and cancer, with the aim of understanding whether the gradual modification cells undergo with age – known as epigenetic drift – is associated with a propensity to have cancer (Zheng, Widschwendter and Teschendorff, 2016). Cancer is a very complex disease, where cancerous cells grow at an uncontrolled rate and negatively impact other cells, tissues, organs and even whole-body systems. Given the detrimental impact of cancer on quality of life, it is a widely researched field and innovations and solutions are always being devised in the quest to better understand the disease. Epigenetics lends a rather useful hand to these research endeavours because certain epigenetic changes such as DNA methylation and histone modifications can lead to the activation of oncogenes (genes that promote cancer growth) and silencing tumour suppressor genes that would inhibit cancer. In their analysis of the relationship between epigenetic changes and cancer risk, Zheng and colleagues concluded that monitoring molecular changes associated with cancer risk factors in normal cells may offer new and improved risk prediction and disease prevention strategies (Zheng, Widschwendter and Teschendorff, 2016).

2.3. Migration

2.3.1. Migration and paediatric development

In this paper we are committed to building on the work previously undertaken in the field, while offering a fresh perspective. The dataset used in this thesis comes from Houghton et al.'s 2014 paper titled: Childhood Environment Influences Adrenarcheal Timing among First-Generation Bangladeshi Migrant Girls to the UK. In the 2014 paper, pubertal development of girls living in the UK and Bangladesh. The project looked at first-generation, British-Bangladeshi children and second-generation British-Bangladeshi children whose parents emigrated from Bangladesh to pursue a life in the UK. Houghton and colleagues went on to compare these cohorts with Bangladeshi girls who stayed living in Sylhet, Bangladesh, as well as children of white European origin that attended the same schools in east London, UK, as the first and second-generation Bangladeshi girls (Houghton et al., 2014a).

There are some overlaps in key characteristics and the populations analysed in Houghton's earlier work, the ABBY (adolescence among Bangladeshi and British youth) project, which investigated

the juvenility, puberty and adolescence among Bangladeshi and British youth. Houghton's work in 2013 focused on the relationship between migration and growing up from a biocultural perspective (Houghton, 2013). That is, an interdisciplinary approach that focuses on the intersection between how an individual's biological composition, and the cultural environment they grow up in, impact them. Houghton analysed the relationship between migration, ethnicity, ecology and the timing and tempo individuals reached adrenarche, thelarche, pubarche and menarche. The four stages of development relate to key milestones reached as humans age in their formative years.

Adrenarche is the maturation of the adrenal glands, and typically happens when girls are aged between 6 and 8 years. This stage of development is associated with the production and secretion of adrenal androgens hormones, their secondary function includes growth of axillary and pubic hair. Adrenarche is clinically defined when dehydroepiandrosterone sulfate (DHEAS) levels exceed 40 to 50 μg/dl in serum and the corresponding level in saliva was converted by Houghton et al. using the 0.1% conversion factor (Houghton et al., 2014a).

Thelarche refers to the beginning of breast development in girls when they are aged between 8 and 13 years, it also marks the beginning of puberty. **Pubarche** is the development of pubic hair driven largely by an increased production of sex hormones like oestrogen. It overlaps with the same period of thelarche. It is another indicator of puberty. **Menarche** is most commonly known as the first-time girls menstruate. This developmental milestone has many cultural implications dependent on the environment an individual is raised in, and biologically speaking it means young girls now have the ability to get pregnant – although given that they are aged between 10 and 16 at this point and not yet fully matured, pregnancy is ill-advised as it may hinder other developmental stages not linked to their menstrual cycle (Campbell, 2006).

The earlier study published in 2013 focused on 488 girls of similar ages to the cohort analysed in this thesis - aged between 5 and 16 years. The girls were either Sylheti, first-generation British Bangladeshi, second-generation British Bangladeshi or white British. The findings concluded that adrenarche happened two years earlier among first-generation British Bangladeshi girls. Houghton argues that this discrepancy is linked to a change in ecological factors, which then lead

to a more rapid juvenile onset (Houghton, 2013). The timing in which individuals reached menarche did not diverge significantly between the different groups, the findings concluded. Houghton's findings suggest that it is stressful for Bangladeshi girls who are first-generation migrants to grow up in London, UK. In addition to this, the study suggests that the extended period of plasticity among girls who experienced socio-ecological changes during their transition to adolescence may be an adaptive response – which might in turn ensure better tracking of the current socio-ecological conditions they are exposed to and make better predictions for future environmental changes (Houghton, 2014a).

Also in 2014, Houghton et al. published findings from a study that looked into the impact an individual's childhood environment had on their adrenarche timing, particularly among first-generation Bangladeshi migrant girls to the UK. The study compared whether girls who emigrated Bangladesh and moved to the UK during their key developmental years reached adrenarche earlier than their non-migrant counterparts both in the UK and Bangladesh. In this study 419 healthy girls were recruited from the same four groups – Bangladeshi girls living in Sylhet, Bangladesh, first-generation Bangladeshi migrants in the UK, second-generation Bangladeshi migrants in the UK, and white European British girls. The girls were aged between 5 and 16 years. The team collected a wide range of data, which included: height, weight, waist circumference, saliva samples, information on socioeconomic status, family history of migration, self-reported ethnicity and other characteristics. The data was collected between 2009 and 2011 (Houghton et al., 2014a).

Saliva samples to measure DHEAS were collected when the research team interviewed participants. The samples were collected in 5 ml polystyrene tubes using gum base as a stimulant and collection tubes were immediately placed in a cooler with ice until transported to storage laboratories. They were stored at –20C. For the purpose of Houghton et al.'s 2014 study, the samples were analysed using salivary DHEAS enzyme-linked immunosorbent assay. DHEAS (dehydroepiandrosterone sulphate) is a male sex hormone that plays an important role in making the female sex hormone oestrogen, and the male sex hormone testosterone (Houghton et al., 2014a).

Socioeconomic findings from Houghton et al.'s research

Houghton and colleagues found that the median age of arrival for first-generation Bangladeshi migrants to the UK was 4 years. According to the study's findings, families of girls in the UK had lower employment rates but higher educational qualifications than families of girls living in Bangladesh. Fathers of first-generation Bangladeshi girls had slight lower employment rate (67%) compared to fathers of second-generation Bangladeshi girls (77%) or European girls (76%), but Houghton and colleagues found the divergence negligible, as the P-value was 0.2 (Houghton et al., 2014a).

Mothers, on the other hand, had low employment rates across the board. But they were incrementally higher as the population moved away from Bangladesh – 19% of Sylhetis interviewed reported their mother was employed, compared to 21% of first-generation Bangladeshi migrants, and 26% of second-generation Bangladeshis in the UK. In contrast, 44% of mothers of European girls in London were employed. House ownership was lower among first-generation parents when compared to second-generation and Sylhetis – but it remained higher than rates among parents of European girls.

Houghton et al. also found that girls were heavier once living in the UK. The weight-for-age z score of first-generation girls was 0.1, while the score for second-generation girls was 0.3 and the score for European girls was 0.8. The BMI z-score of first-generation girls was 0.7, similar to the BMI-Z for second-generation girls (0.6). European girls meanwhile had a BMI-Z of 1. DHEAS levels detected in the saliva samples increased with age across all groups, which was expected by researchers as the girls were undergoing key developmental changes. Among girls aged less than 9.5 years, DHEAS levels were higher among the groups living in the UK – that means all except Sylhetis. First-generation Bangladeshis had the highest DHEAS levels in the saliva samples collected. But Houghton and colleagues found that there were no differences in age adjusted DHEAS levels by age at migration or time since migration. The median age at adrenarche happened earlier for first-generation Bangladeshis at 5.3 years on average. European girls reached adrenarche at 7.1 years, Sylheti girls at 7.2 years and second-generation Bangladeshi girls at 7.4 years (Houghton et al., 2014a).

Taking ABBY project's findings forwards

Given the conclusions from the study, and a shared understanding that accelerated or premature development can lead to health implications, this thesis hopes to provide further insight and analysis towards understanding the potential correlations between accelerated pubertal timing and epigenetic ageing within populations, particularly any changes or observations that can be drawn from focusing on first-generation migrant Bangladeshi girls. The work of this thesis aims to bridge the literature and gain a better understanding of whether there are significant developmental differences between Bangladeshi girls that migrate during their childhood and their Sylhet-Bangladeshi and white British counterparts that do not migrate during their childhood. A critical analysis that focuses on whether an individual's migration status impacts their health outcomes in contrast to their non-migrant counterparts will be carried out in the discussion section of this thesis.

The samples collected for the ABBY project were frozen and retrieved for the purpose of this thesis. They were analysed using the EPIC Illumina array DNA methylation research, which measure the methylation levels across the samples' 850,000 CpG sites.

2.3.2. Migration and epigenetic ageing

Stöger et al. investigated whether childhood environments influence an individual's epigenetic age and the methylation concordance of a CpG clock locus. The team looked at British-Bangladeshi migrants in particular and the research and analysis undertaken for the purpose of this thesis builds on their research findings (Stöger et al., 2023).

The research team analysed key features of DNA methylation in adult – albeit young – British Bangladeshi women who experienced different environments during their formative childhood years. The study looked at DNA methylation data from migrants who grew up in Bangladesh and were exposed to comparatively higher pathogens and more limited access to healthcare, as well as second-generation British-Bangladeshi women born in the UK to immigrant parents.

Buccal DNA was used to estimate the DNA methylation-base age from 25 individuals in total: 14 migrants and 11 second-generation British-Bangladeshi women. The subjects were aged between 18 and 35 years.

AgeAccel data showed that women who had spent their childhood in Bangladesh had a higher age acceleration rate than their UK counterparts. Stöger et al. chose to make the comparison using AgeAccel because it measures DNA methylation age independently of chronological years. In their study, Stöger et al. went on to evaluate the preference for concordant DNA methylation at the luteinizing hormone/choriogonadotropin receptor (LHCGR/LHR) locus. They chose this region because it has one of the 353 CpG sites from Horvath's epigenetic clock. From this analysis, Stöger et al. also saw that both individual and double stranded DNA molecules indicate higher DNA methylation stability at the LHCGR/LHR locus among samples from women who grew up in Bangladesh.

From these research findings, Stöger et al. concluded that childhood environments may trigger subtle physiological changes that are detectable after exposure has occurred, which explains altered activity of the epigenetic maintenance system and difference in the proportion of cell types in buccal tissue (Stöger et al., 2023).

Methodology from Stöger's research

To carry out their research Stöger et al. collected DNA methylation data using the Illumina HumanMethylationEPIC BeadChip array. Genomic DNA data was isolated from buccal cell samples using the DNeasy Blood & Tissue kit by Qiagen. The team proceeded to employ multidimensional scaling plots to best assess whether significant bath effects were skewing the datasets. The data went on to be processed with the Bioconductor/minfi package. CpG probes are short sequences of single-stranded DNA, often labelled with a fluorescent tag. The probes associated with single nucleotide polymorphisms (SNPs) were removed from the analysis as they can interfere with DNA methylation measurements (Stöger et al., 2023). CpG probes on both X and Y chromosomes were retained. Probes with a detection probability lower than 0.01 were excluded from further analysis. The methylation beta values were then analysed using Horvath's clock and the skin and blood clock.

Methylation data ranging from 0 to 1 was analysed using some normalisation procedures – such as SWAN and Quantile – and the data was also analysed raw, with no normalisation (Stöger et al., 2023). Different normalisation procedures are needed to answer specific research questions. The team made the methylation dataset available on the Gene Expression Omnibus – a repository for genomic data hosted by the National Center for Biotechnology Information. To obtain epigenetic age estimates the team used the beta values from the EPIC array in Horvath's clock to measure both DNA methylation age and age acceleration residuals. Stöger et al. chose to normalise the data when inputting it on Horvath's online calculator because by normalising it the data becomes comparable to the training dataset Horvath used to devise the clock. The output of the clock provides epigenetic age estimates, age acceleration values, predicts an individual's sex based on the DNA information and it also predicts the tissue or cell source of the samples.

2.4. Migration, cortisol and epigenetic ageing

Migration is an unquestioned stressor, and we will delve into the implications of migration and being a first-generation immigrant from the Global South in the discussion section of this thesis. Forced migration as a result of conflict or economic factors can expose individuals to high levels of stress, whether that is because they are fleeing traumatic events, or because they arrive in a country where living conditions are uncertain, discrimination is rife and the wider family network struggles to adapt to a new environment.

Work undertaken by Perroud et al. investigated DNA methylation patterns within cohorts of trauma-exposed subjects, and it found a link between trauma exposed individuals and differential DNA methylation patterns. Their work focused on the transgenerational impact of maternal stress, with an emphasis on epigenetic mechanisms and physiological manifestations of the HPA axis (Perroud et al., 2011). Transgenerational epigenetic inheritance is well-documented and findings from Yehuda et al. demonstrate that the impact of traumatic experiences is long-lasting, and the consequences are intergenerational. In their 2016 paper, Yehuda et al. looked at epigenetic changes to DNA methylation in individuals who survived the Holocaust themselves, or were descendants from Holocaust survivors (Yehuda et al., 2016).

Lam, Raket and Soehl investigated the impact of discrimination on DNA methylation across a racially diverse sample. Their findings concluded that environmental factors like discrimination were linked to adverse DNA methylation patterns across an Asian American cohort (Lam, Raket and Soehl, 2018). Discrimination and exposure to environments that do not welcome an individual when they require support or are particularly vulnerable can be extremely detrimental.

Data from the cohort of Bangladeshi first-generation girls, which forms the basis of our study, was collected between 2009 and 2011 - the beginning of overt hostile environment initiatives by the British government to deter immigration. The discussion section of this thesis will explore the impact of attrition and punitive measures on the wellbeing of migrants arriving in the UK, particularly during the past decade (Essex et al., 2022).

3. Methods

The research aim of this thesis was to use at least one epigenetic clock to examine the potential impact of childhood environments on differences across chronological age and biological age. The buccal samples analysed in this study were donated by Gillian Bentley and colleagues from Durham University who worked on the ABBY project. The donated data samples come from girls aged less than 18 years at the time of collection and permission from their guardians was sought by Bentley and colleagues at the time. The sample IDs were anonymised to protect the individuals' privacy. The findings provided in this thesis are innovative because it is the first-time epigenetic age estimates have been provided for the ABBY project cohort. It is also the first opportunity to coordinate and identify analytical insights from the cohort's epigenetic ageing data and their socio-economic data.

Previous work identified strong correlations between childhood environmental conditions and adult reproductive function (Bar-Sadeh et al., 2020; Murphy et al., 1995). In particular, British-Bangladeshi migrant women of reproductive age, who were born and raised in the Sylhet region of Bangladesh and now live in London. These individuals were found to have consistently lower levels of the hormone progesterone - and experience overall accelerated reproductive ageing when compared to Bangladeshi women who moved to the UK prior to age of eight and women who were born in London to first-generation Bangladeshi immigrants. Using buccal cell DNA samples from these British-Bangladeshi women, Stöger et al. recently identified genome-wide, altered DNA methylation levels in individuals who migrated as adults to the UK (Bar-Sadeh et al., 2022) and accelerated epigenetic age (Stöger et al. 2023). DNA methylation data (derived from the Illumina EPIC array) was used to establish epigenetic age estimates, using Horvath's epigenetic clock (2013; Horvath and Raj, 2018).

Buccal samples donated from Houghton et al.'s 2014 study

The datasets of our study originate from the EPIC array, which was carried out with information from the ABBY project. We chose to use this data because there was also data available in relation to our samples socio-economic background, as well as their hormonal levels and proneness to

early puberty. Consolidating these different sets of information, alongside the novel information on epigenetic age estimates, allowed for a more comprehensive analysis of the relationship between biological age and environmental factors.

Buccal samples were collected by Houghton et al. from 10 schools in east London, UK, between September 2009 and December 2010, and from 7 schools in Sylhet Town, Bangladesh, between January 2011 and April 2011. Houghton et al. chose London schools in areas that had a high proportion of immigrants from Sylhet, Bangladesh. Meanwhile, the recruitment in Bangladesh focused on schools that were likely to have pupils from middle class families, as most families who emigrate Bangladesh and move to the UK are middle class (Houghton et al. 2014a; Gardner, 1994). Buccal samples collected by Houghton et al. were frozen in quick succession to the data collection - to preserve the DNA quality - and stored at the University of Durham (Houghton et al., 2014a).

EPIC array data

The Illumina Infinium MethylationEPIC array BeadChip (850K) was carried out by the Epigenomic Services from Diagenode (Cat nr. G02090000). The buccal cell DNA data samples from 115 girls aged between 5 and 16 years was entered on the Illumina 850 MethylationEPIC array platform. In one instance, the chronological age from an individual pertaining to the white British population group was not shared with this study, which led to the use of the mean chronological age for the group of white British girls. That was the case in solely this instance.

The decision to use the EPIC array was largely motivated by its breadth and depth. It offers a comprehensive coverage of DNA methylation across the human genome and assays over 850,000 distinct CpG sites – locations where DNA methylation takes place. The EPIC array covers gene promoter regions as well as gene bodies, CpG island shores and enhancers – please refer to the previous chapter for an in-depth explanation of these terms.

The EPIC array is an evolution of the HumanMethylation450 BeadChip. It is compatible with many different sample types and capable of analysing genomic DNA data from cell lines, blood and tissues. The dual-colour BeadChip technology means the EPIC array can use fluorescent labels to

distinguish between methylated and unmethylated cytosines at different CpG sites. The fluorescence, a feature that was also a key part of the Illumina HumanMethylation450 BeadChip, allows for quantitative methylation data and the ability to make pointed comparisons between samples (Dedeurwaerder et al., 2014).

One limitation of the EPIC array is that early career scientists with restricted access to high quality DNA samples may be unable to yield the most accurate results from the sophisticated tool, that is because the quality of the DNA impacts the output the EPIC array can provide. In addition to this, there is a significant and sizeable cost to obtain such comprehensive information – which the EPIC array provides – and run large-scale experiments. Given the volume of data the EPIC array can return, data normalisation was carried out by bioinformaticians, and results were harvested from six distinct epigenetic clocks. The large data returned by the EPIC array on its own does not support epigenetic ageing research; it is only when we apply it to epigenome-wide association studies that we can identify correlations between DNA methylation patterns and diseases, phenotypic traits, or environmental factors. In conjunction, the information can allow studies to provide informed findings about links between DNA methylation and epigenetic ageing, cancer, neurodevelopmental disorders, cardiovascular disease and age acceleration.

The analysis in this thesis goes further than previous work, as the study looks at the epigenetic age of our samples across different epigenetic clocks and compares it to the individuals' chronological ages, as well as other socio-economic factors.

Horvath clock

In a similar way to Stöger et al., we collected DNA methylation data using the Illumina HumanMethylation EPIC BeadChip Array. The array enabled us to assess the DNA methylation levels at several CpG sites across the genome. We used Horvath's method to establish the DNA methylation-based age and age acceleration residuals (AgeAccel) from the values obtained from the Human MethylationEPIC BeadChip Array. Horvath's clock uses algorithms to estimate epigenetic age based on DNA methylation patterns, it focuses on 353 CpG sites. We ran our data using an open-source tool, which works as an online DNAm age calculator, and it is available at: https://dnamage.genetics.ucla.edu/home. The default 'Normalize Data' setting was applied

during the online submission process to enhance predictive accuracy, as it made the input data comparable to the training data of Horvath's clock.

The output from Horvath's calculator provided estimates of both DNAm age and AgeAccel but it also predicts the sex and tissue or cell source of the samples. All samples were from females and buccal (saliva) tissue was the source used. In the first couple of runs through the clock, Horvath's tool was able to recognise that our samples came from buccal cells, but the gender readings returned were either 'male' or 'unsure'. Given the inaccuracy, the decision was made to extract the raw DNA methylation data values as well as normalising the data. We normalised the data by carrying out quantile normalisation, which is a method widely in genomics research. Quantile normalisation standardises and aligns methylation values for CpG sites across different samples and ensure that each sample has the same distribution of methylation values. To carry this out, average methylation value at each quantile is assessed and then implemented in each of our samples (Bolstad et al., 2003).

Tribulations with both raw and normalised data entries, with incorrect age predictions, led us to revisit Horvath's tutorial guidance. Data was then normalised in accordance with the guidance provided by Horvath in the tutorial information to use his online epigenetic clock, in which Horvath explains that data should be normalised using the online tool itself. The instruction was that by normalising the DNA methylation data using the mathematical model, it would make it comparable with the training data used to develop the epigenetic clock. As presented in the results chapter that follows, this guidance was very helpful and led to far more accurate epigenetic age predictions – in comparison to our cohort's chronological ages.

PedBE clock

To compare results, we moved on to use the PedBE clock, which was designed as a paediatric buccal epigenetic mathematical model to predict individuals' ages based on DNA methylation information. The mathematical model was designed using the coding programme R and McEwen et al. made the PedBE clock script available on GitHub, an open-source platform where data

analysts can share coding scripts. Prior to using the clock, it was necessary to install and load the required R packages like 'minfi' for data processing and the R script for the PedBE clock itself.

A thorough data cleansing process was undertaken to ensure all Excel files had the correct column and row labels for ease and clarity. The submission to all epigenetic clocks requires clearly labelled Excel files that include the sample IDs, CpG sites and in some cases the chronological age. For the PedBE clock the beta methylation values (ranging from 0 to 1) for the CpG sites were prepared and the data was organised in the layout of sample IDs as rows and CpG sites as columns on the spreadsheet. The DNA methylation data was pre-processed to ensure it met the requirements of all clock analyses. For the PedBE clock analysis it included removing any CpG sites with missing values, normalising the data where necessary and adjusting for batch effects. To harvest these data, our study was supported by colleagues at the University of Nottingham's School of Medicine (Poojitha Rajasekar and Rachel Clifford) who have a better domain of the R coding language. After the data had been pre-processed, the team ran the script by applying the 'calculate PedBE' function from the 'PedBE clock' package, which returned the PedBE age estimates for each of the samples collected from our cohort.

Epigenetic age estimates from the PhenoAge clock, LinAge, skin and blood clock and Hannum's clock, the candidate counted on the support from her second supervisor. Richard Emes, who co-supervised this MRes thesis alongside Reinhard Stöger, supported the findings from this study by providing bioinformatics insight and enabling the data harvest from some epigenetic clocks.

All epigenetic clock data presented in this study was obtained either by the MRes candidate herself, or with the support of her academic supervisors. As explained above, in one instance guidance was sought from colleagues within the University of Nottingham to obtain results from the PedBE clock.

Data visualisation

The Python coding language was used to visualise the DNA methylation-based epigenetic age predictions plotted against the populations' chronological ages in linear regression graphs with

multiple plots. Whisker box plots were designed based on the age acceleration values from the different clocks' residuals.

Statistical analysis

To assess the statistical significance of the results ANOVA tests were carried out to measure the p-value of the age acceleration residuals across the different populations measured within this study. To do this the Brown-Forsythe ANOVA test and the Welch ANOVA test were used.

4. Results

As previously outlined, the research findings presented in this study are the result of novel epigenetic ageing estimates from donated saliva samples from 115 girls, which was originally collected between 2009 and 2011. The data used for the purpose of this thesis was a subset from 419 buccal samples collected from girls for Houghton and colleagues' 2014 study (Houghton et al., 2014a). The 115 buccal samples analysed in our study were composed of four populations: Bangladeshi girls (n=43), first-generation girls (n=24), second-generation girls (n=27) and white British girls (n=21). All girls were aged between 5 and 6 years when samples were collected. The samples were processed by the EPIC array to measure DNA methylation levels across 850,000 CpG sites. Data results from the array were then properly categorised and stored, and methylation information on CpG sites relevant to specific epigenetic clocks was retrieved in accordance with the instructions for each mathematical model.

The hypothesis ahead of data collection from the six epigenetic clocks was that well-established clocks, which are widely respected and have been used vigorously over the past 12 years since the first clock was developed (Bocklandt et al., 2011), would return the most stable and closely correlated epigenetic age predictions to the populations' chronological ages. However, based on the findings discussed below, the 2020 PedBE clock provided more closely correlated results to our cohort's chronological age. While better-known clocks like Hannum's produced age estimates that hugely deviated from individuals' chronological ages, across the populations (Horvath and Raj, 2018; McEwen et al., 2020). As well as looking at age predictions, the age acceleration rate was also the focus of our study's analysis. It is helpful to remind the reader that positive epigenetic age acceleration in later life is linked to age-related morbidities, like cancer and coronary heart disease. However, a faster developmental pace in early life can be beneficial for immediate health outcomes (Belsky et al., 2015; Dammering et al., 2021). The findings from this study may also aid the understanding of whether there is a correlation between girls who age at a faster pace epigenetically and those who reach key developmental stages, such as pubertal timing, before their non-migrant counterparts (Houghton et al., 2014a). The discussion chapter will expand on these issues further.

To calculate if there were statistically significant differences in age-estimates between the groups, we used ANOVA tests, with a threshold value, $P < 0.05$. P-value results were gathered for both Welch's ANOVA test and the Brown-Forsythe ANOVA test. The findings presented are from Welch's ANOVA test because Glantz and colleagues recommend the Welch test in most situations, as it has more power and maintains alpha at its desired level (Glantz, Slinker and Neilands, 2016). Results for the Brown-Forsythe ANOVA test, which are recommended when the data is skewed, can be found in the appendix.

4.1. Horvath's multi-tissue clock: one clock, three sets of results

Results yielded from Horvath's clock had a large discrepancy dependent on (i) whether the datasets were normalised, and (ii) how they were normalised. Our study unveils three sets of results from the same Horvath multi-tissue clock, in the quest for transparency with the reader. One result, illustrated in **figure 6** and **figure 8**, contains the epigenetic age estimates harvested from raw DNA methylation data processed by Horvath's clock. Another, shown in **figure 5** and **figure 7**, shows age estimates from data that was normalised by ourselves, independently of the mathematical model, and processed by the clock as raw data (details on how we carried out data normalisation can be found in the previous chapter). And lastly, the most significant results that can be seen in **figure 9** and **figure 10**, are from raw DNA methylation data processed and normalised by Horvath's epigenetic clock itself, in accordance with its training data.

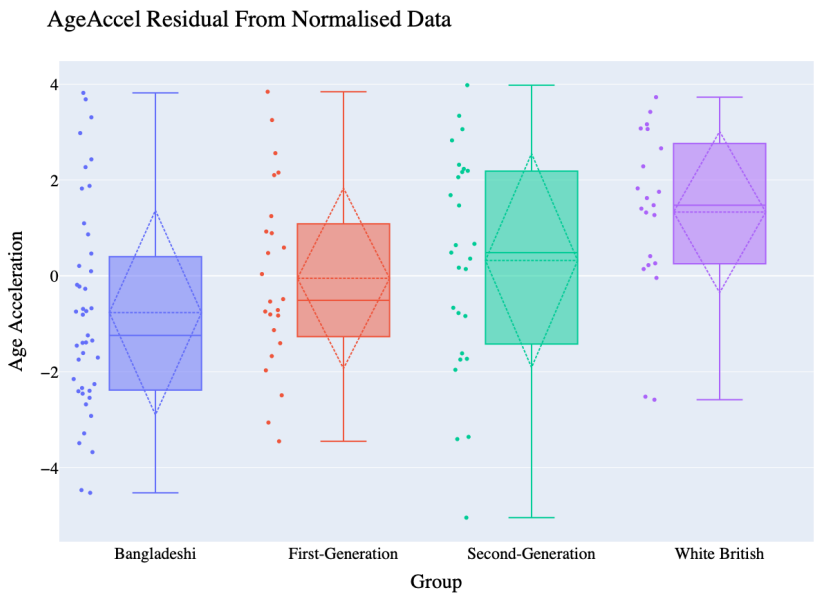


Figure 5. *Age acceleration based on data derived from Horvath's multi-tissue epigenetic clock, using normalised data.* Box plot shows the differences in pace of epigenetic ageing against four populations of girls aged between 5 and 16 years. Bangladeshi (blue); first-generation (orange); second-generation (green); white British (purple). The median values are indicated by a darker horizontal line in that goes through each of the different group's box plot. A positive or negative age acceleration value indicates that the estimated epigenetic age of the samples is higher or lower, respectively, than expected. All four populations matured at a similar rate. The P-value for the age acceleration data is based on Horvath's clock using normalised methylation data was **0.0015**; statistically significant.

Figure 5 and **figure 6** show the age acceleration, illustrated in box plots, for the data input on Horvath's clock utilising the tool's 'raw data' selection, but **figure 5** shows the acceleration for when we normalised the data ourselves and **figure 6** shows the acceleration rate for raw DNA methylation data. Age acceleration among Bangladeshi girls had an upper fence of 5.13 and a minimum of –5.05. We highlight the upper fence because it is a significant factor in statistical analysis as it relates to outliers in the dataset. Identifying which results were outliers helps to visualise the full range of values across any given dataset and enables an enhanced interpretation of these results alongside averages, such as medians and means. In turn, a richer and more complete understanding of the results collected can be determined.

The median age acceleration among Bangladeshi girls who live in Sylhet was –0.14. For first-generation Bangladeshi girls who migrated and now live in the UK, the median age acceleration value was -0.67, while the upper fence of acceleration is 5.81 and the minimum was -5.66. Second-generation Bangladeshi girls had minimum age acceleration values of –6.06, an upper fence of 5.83 and a median value of 0.22. Data collected from their white British counterparts meanwhile produced median values of 0.44, an upper fence of 4.96 and a minimum age acceleration value of –3.01.



Figure 6. *Age acceleration based on data derived from Horvath's multi-tissue epigenetic clock, using raw data.* Box plot shows the differences in pace of epigenetic ageing against four populations of girls aged between 5 and 16 years. Bangladeshi (blue); first-generation (orange); second-generation (green); white British (purple). The median values are indicated by a darker horizontal line in that goes through each of the different group's box plot. A positive or negative age acceleration value indicates that the estimated epigenetic age of the samples is higher or lower, respectively, than expected. All four populations matured at a similar rate. The P-value for the age acceleration, which are based on *Horvath's multi-tissue* clock using raw methylation data was **0.5847**; not statistically significant.

Age acceleration from the normalised dataset shows a similar pattern to the one illustrated in **figure 5**, as the age acceleration rate among Bangladeshi girls from Sylhet trends lower (-1.24) than the rest. But the exception in this scenario is that the median age acceleration values for

whit British girls is significantly higher (1.47) than that for first-generation Bangladeshi girls who have moved to the UK (-0.51). Second-generation Bangladeshi girls born in the UK to immigrant parents had a median age acceleration value of 0.48.

If we focus on each group methodically, some trends can be observed although only the results for the raw data are statistically significant.

Given that the mean chronological age among Bangladeshi girls was 7.7 years, the values returned by Horvath's clock when it processed both normalised and raw data largely deviated from our predictions. Hence why the age acceleration values are such an important tool – as it helps put our findings among a cohort into context. As illustrated in **figure 8**, the mean epigenetic age for Bangladeshi girls using the raw data on Horvath's clock was 16.9 years. For the normalised data on the same clock, the mean value was 26.1 years.

For first-generation Bangladeshi migrants to the UK, the mean chronological age was 10.2 years. The DNAmAge for normalised data again deviated largely from the chronological value for each of the girls and it returned that the mean epigenetic age among the first-generation cohort was 28.7 years. For raw data, the mean epigenetic age was 18.9 years. While second-generation Bangladeshi girls had a mean chronological age of 7.7 years. The results obtained by running the normalised DNA methylation data on Horvath's clock was 27.2 years. The epigenetic age estimate for the raw data was 16.9 years. Data from white British girls of a similar age group, who had the mean chronological age of 8.8 years, returned similarly skewed results. The DNAmAge for normalised datasets had a mean epigenetic age estimate of 29 years, while the DNAmAge for raw data returned an estimate age of 18.6 years.

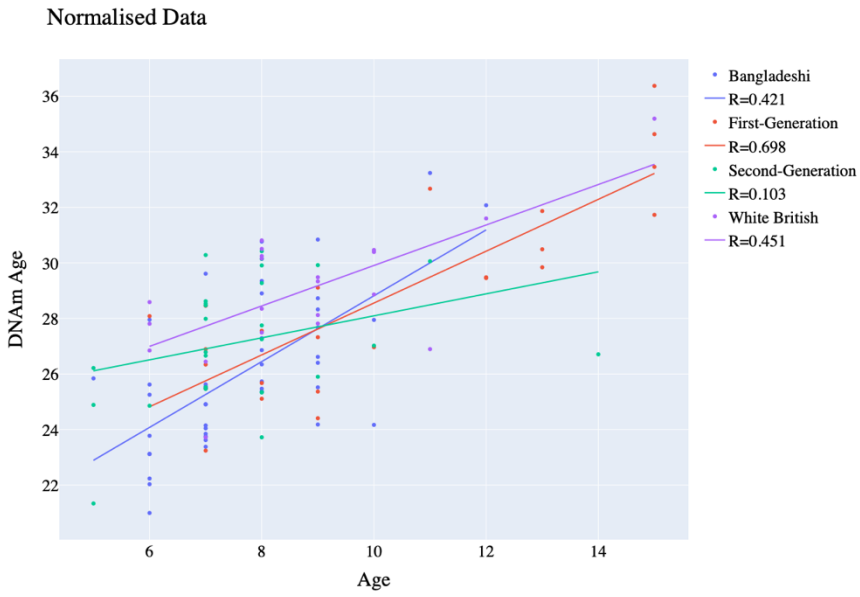


Figure 7. DNAmAge normalised clock epigenetic age estimates plotted against individuals' chronological ages. Graph shows DNA methylation-based age in years against chronological age in years for four populations of girls aged between 5 and 16 years. Bangladeshi (blue), R=0.421; first-generation (orange), R= 0.698; second-generation (green), R=0.103; white British (purple), R=0.451. The four populations did not mature at a similar rate. The gradients for the four populations have a wide range, with a large discrepancy between the gradient for the second-generation group and the population of first-generation group. The results suggest the clock is not well-calibrated to predict an individual's age based on DNA methylation data.

The aim was to obtain the best results from the samples available. To begin, the datasets from the EPIC array were normalised before Horvath's epigenetic clock was used. The reasons for doing this were largely influenced by discussions across the literature in epigenetic ageing, which suggest that to obtain the best results from epigenetic clocks designed to fulfil a function that may not be specific to the area of study you are investigating you should normalise the data. Normalising the data often shrinks the margin for error as the data you input in the mathematical model used to measure epigenetic age estimates should in theory be closer to the data quality from the training dataset used to develop the machine learning method. In the upper range of the diverging predictions, 4 girls who were aged 15 and belonged to both the first-generation migrants' group and the white British group were estimated to have epigenetic ages ranging from 31.7 years to 36.4 years. The lowest epigenetic age prediction, for second-generation

Bangladeshi girls aged 5, was 21 years. While other second-generation 5-year-old girls were predicted epigenetic ages of 24.9 years and 26.1 years. Given that the majority of epigenetic age estimates from these calculations diverged starkly from the chronological ages of the individuals whose data we processed, we decided to use the raw DNA methylation data instead. Given the skewed results for the normalised data, as shown in **figure 7**, the decision was made to carry out the same tests to assess epigenetic age estimates using raw datasets.

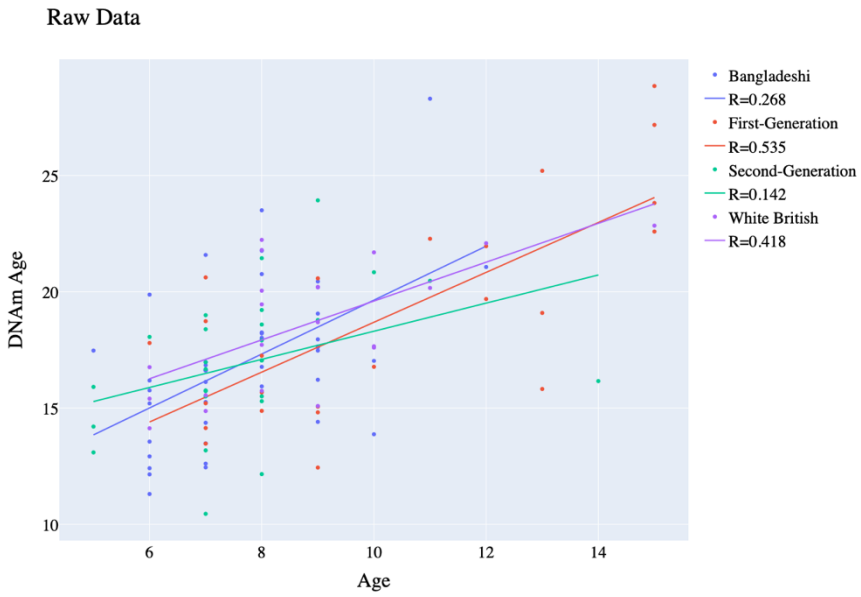


Figure 8. DNAmAge raw clock epigenetic age estimates plotted against individuals' chronological ages. Graph shows DNA methylation-based age in years against chronological age in years for four populations of girls aged between 5 and 16 years. Bangladeshi (blue), R=0.268; first-generation (orange), R= 0.535; second-generation (green), R=0.142; white British (purple), R=0.418. All four populations matured at a similar rate. The four populations have a similar gradient, but the majority of epigenetic age estimates do not correlate to the chronological age in years of the populations. The clock does not appear to be well-calibrated to predict our samples' epigenetic ages based on DNA methylation data from buccal samples.

The raw DNA methylation data delivered less skewed results across all four groups. However, the estimated epigenetic age output from Horvath's clock persisted to deviate from their chronological ages. Bangladeshi girls (n=43) were aged between 5 and 12 years when their buccal

epithelial cells were collected as part of the wider study carried out by Houghton et al. on the impact of childhood environments in key developmental stages among migrant and non-migrant girls in the UK and Bangladesh (Houghton et al., 2014a). The epigenetic age estimates from this clock for some of the samples collected from second-generation Bangladeshi girls aged 5 years predicted their epigenetic age to range from 13.1 years to 15.9 years. Although it remains far greater than their chronological ages, the results deviated far less than those discussed above from the normalised datasets. This was a theme across the three other groups as well.

To understand the statistical significance of the results, the Welch ANOVA test was used. For the normalised results the ANOVA test returned a P-value of 0.0015. Given that the P-value was lower than 0.05, the difference between groups was deemed to be statistically significant. For raw data, the Welch ANOVA test delivered a P-value of 0.5847, which was not deemed statistically significant.



Figure 9. *Age acceleration based on data derived from Horvath's multi-tissue epigenetic clock.* Box plot shows the differences in pace of epigenetic ageing against four populations of girls aged between 5 and 16 years. Bangladeshi (blue); first-generation (orange); second-generation (green); white British (purple). The median values are indicated by a darker horizontal line in that goes through each of the different group's box plot. A positive or negative age acceleration value indicates that the estimated epigenetic age of the samples is higher or lower, respectively, than expected. All four populations matured at a similar rate. The P-value for Horvath's clock based on normalised methylation data was **0.5479**; not statistically significant.

Horvath's world-renowned multi-tissue epigenetic clock was one of the first tools we turned to as we embarked on a quest to make sense of epigenetic ageing among our cohort, as we hoped it would help build on the work already carried out by Loughton, Bentley, Begum, Stöger et al., as we detangle the links between biological compositions, cultural environments, and the constant dialogue between the two components that shape human identity. Hence, when the results described in **figures 5, 6, 7, 8** yielded diverging epigenetic age estimates from the chronological age, we decided to further consult the tutorial notes for Horvath's clock submission tool. As explained earlier, the third entry data entry reflected guidance from Horvath's submission service that we should choose to normalise our datasets within the R tool to allow for consistency with the training dataset used to originally develop the clock. Results from the third, and successful, data entry are illustrated in **figure 9** and **figure 10**.

We compared our results with the mean chronological ages across our cohort (**Bangladeshi:** 7.7 years; **first-generation**: 10.2 years; **second-generation**: 7.7 years; and **white British**: 8.8 years), as we reproduced the same DNA methylation information across different epigenetic clocks – focusing on the CpG sites identified by the clock developers as important towards predicting epigenetic ageing. The mean epigenetic age returned by Horvath's clock for Bangladeshi girls was 10.7 years. For first-generation Bangladeshi migrants to the UK the mean estimated age was 12.1 years. While for second-generation Bangladeshi girls born to immigrant parents in the UK, the mean epigenetic age estimate was 10.8 years and for their white British counterparts the estimated age from Horvath's clock was 11.9 years.

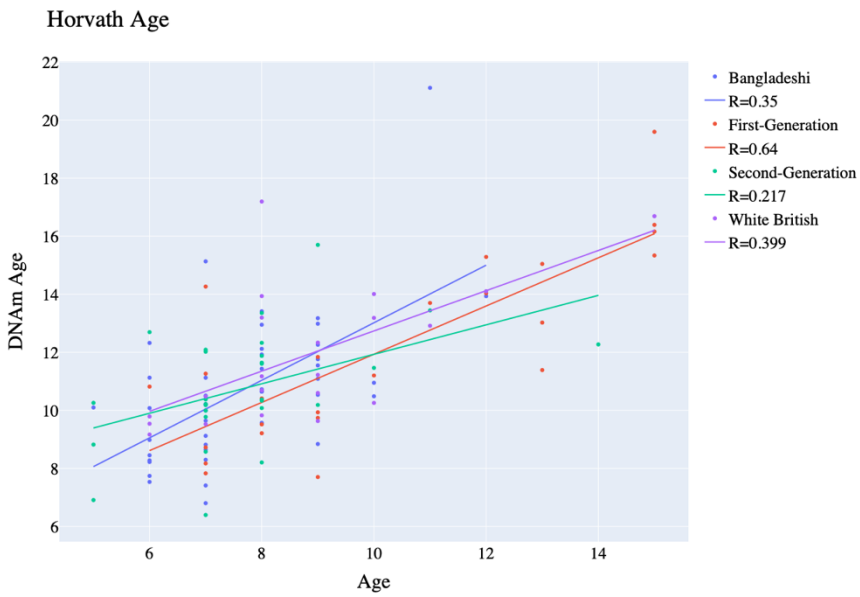


Figure 10. Horvath clock epigenetic age estimates plotted against individuals' chronological ages. Graph shows DNA methylation-based age in years against chronological age in years for four populations of girls aged between 5 and 16 years. Bangladeshi (blue), R=0.35; first-generation (orange), R= 0.64; second-generation (green), R=0.217; white British (purple), R=0.399. All four populations matured at a similar rate. The four populations have a similar gradient. The epigenetic age predictions more closely correlated to the populations' chronological ages, which suggests the clock is calibrated to predict an individual's age based on DNA methylation data.

The statistical significance of the results was calculated with the Welch ANOVA test, which returned a P-value of 0.5479. Given that the P-value was higher than 0.05, the difference observed between populations was not deemed to be statistically significant. These values were calculated based on the age acceleration data, which can be seen in **figure 9.**

Results delivered by Horvath's epigenetic clock, illustrated in **figure 10,** were far closer to the chronological ages from the individuals whose samples we used. As explained in chapter 2 of the thesis, which outlines epigenetic mechanisms and the development of epigenetic clocks as a tool to measure biological age estimates, Horvath's clock was devised to measure ageing from DNA methylation data of different tissue types as well as a wide range of ages. Hence, we were eager to understand the age estimates it could provide.

4.2. PedBE clock



Figure 11. Age acceleration based on data** derived **from PedBE (paediatric buccal epigenetic) clock. Box plot shows the differences in pace of epigenetic ageing against four populations of girls aged between 5 and 16 years. Bangladeshi (blue); first-generation (orange); second-generation (green); white British (purple). The median values are indicated by a darker horizontal line in that goes through each of the different group's box plot. A positive or negative age acceleration value indicates that the estimated epigenetic age of the samples is higher or lower, respectively, than expected. The first-generation population matured at a slightly higher rate than the other three groups. The P-value for the PedBE clock based on normalised methylation data was **0.6500**; hence the changes observed were not statistically significant.

Results returned from the PedBE clock were the most uplifting ones throughout our research process, as the clock was designed to predict epigenetic age estimates for the datasets we had at our disposal. That is to say, the paediatric buccal epigenetic cells clock was made to accurately predict the epigenetic age of children based on samples that contain buccal epithelial cells – like saliva. The age acceleration data from the PedBE clock demonstrates that the median age for all four groups was closer to neutrality – neither positive nor negative age acceleration results – than other clocks were able to predict.

Figure 11 shows that first-generation Bangladeshi girls had the highest age acceleration result, with the median acceleration was 0.24. The highest age acceleration value among first-generation girls was 1.43. The second highest individual age acceleration value was within white British girls (1.35), but the median age acceleration for the white British cohort was far lower (-0.04). Bangladeshi girls had a median age acceleration result of 0.04. For second-generation Bangladeshis, the median age acceleration was –0.07.

The statistical significance of the results was calculated with the Welch ANOVA test, which returned a P-value of 0.6500. Given that the P-value was higher than 0.05, the difference observed between populations was not deemed to be statistically significant.

The rate of DNA methylation-based ageing for each population is well-calibrated to their chronological ages in the PedBE clock. Age acceleration data helps build a better picture of the lived experience across the four cohorts of girls, but alone it cannot answer all the questions. The results from the PedBE clock were encouraging. Among many insights, they confirmed our hypothesis that girls who have emigrated Bangladesh and settled in the UK as first-generation migrants age more rapidly than their non-migrant counterparts.

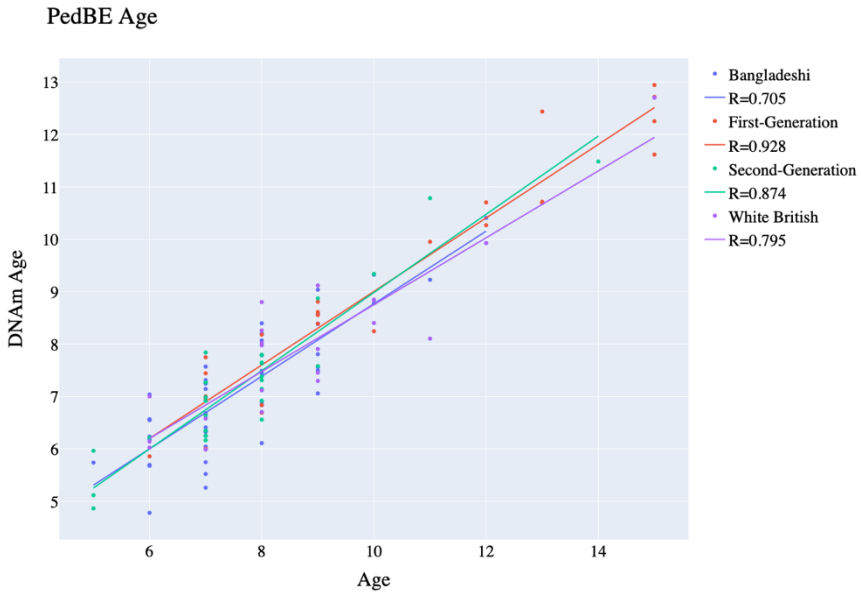


Figure 12. PedBE clock epigenetic age estimates plotted against individuals' chronological ages. Graph shows DNA methylation-based age in years against chronological age in years for four populations of girls aged between 5 and 16 years. Bangladeshi (blue), R=0.705; first-generation (orange), R= 0.928; second-generation (green), R=0.874; white British (purple), R=0.795. All four populations matured at a similar rate. All populations have a similar gradient, which suggests that the clock is a well-calibrated tool used to predict an individual's age based on DNA methylation data.

In addition to this, the information from the PedBE clock had strong R values across the four groups – with the first-generation group attaining the strongest correlation between the epigenetic age estimate from the PedBE clock and their chronological age (R=0.93). This research finding is surprising because given that the median acceleration among first-generation Bangladeshi girls fared higher than the other three groups, we expected the epigenetic age estimates to deviate more from the chronological age. Our findings corroborate the argument put forward by McEwen et al., who designed the PedBE clock to assess datasets like ours. As shown in **figure 12,** datasets from Bangladeshi girls returned a mean epigenetic age estimate of 7.2 years with the PedBE clock. Among first-generation Bangladeshi migrants the mean epigenetic age was 9.2 years. For second-generation girls the mean estimate was 7.3 years, while for white British girls the mean estimated epigenetic age was 8 years.

The PedBE clock was able to accurately predict ages of the individuals in our cohort, and that was likely because the samples analysed were buccal samples and paediatric; the clock was designed to accurately predict the age of paediatric individuals based on buccal cell samples; thus, it fulfilled its mission.

4.3. Skin and blood clock

The skin and blood clock returned age acceleration results that corroborate our hypothesis and much of the argument made thus far in the thesis. **Figure 13** shows that the median age acceleration for first-generation Bangladeshi girls was 0.29. Bangladeshi girls living in Sylhet had an age acceleration median of –0.32, second-generation Bangladeshis tracked a median age

acceleration of –0.14 and DNA methylation data from white British girls produced an age acceleration value of 0.05.

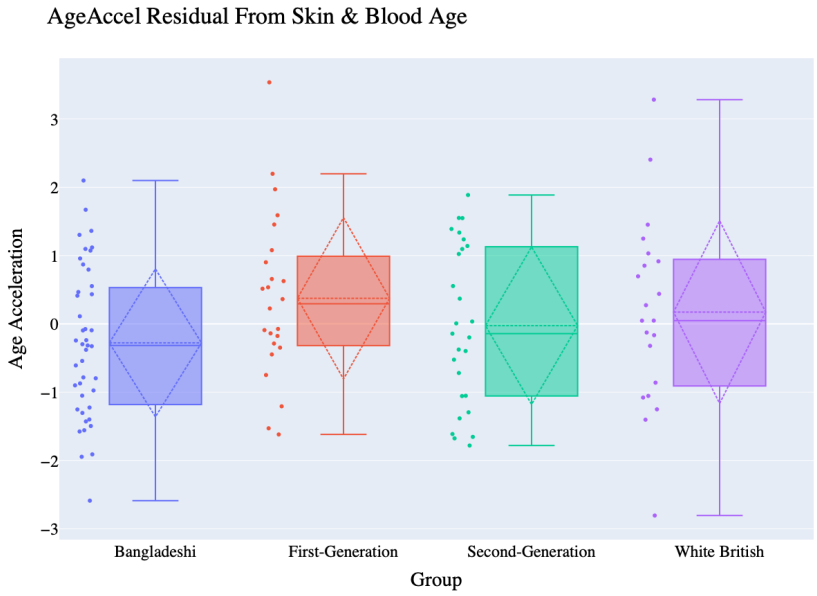


Figure 13. Age acceleration based on data derived from Skin & Blood clock. Box plot shows the differences in pace of epigenetic ageing against four populations of girls aged between 5 and 16 years. Bangladeshi (blue); first-generation (orange); second-generation (green); white British (purple). The median values are indicated by a darker horizontal line in that goes through each of the different group's box plot. A positive or negative age acceleration value indicates that the estimated epigenetic age of the samples is higher or lower, respectively, than expected. All four populations matured at a similar rate, while the range among the white British population was larger. The P-value for the skin and blood clock based on normalised methylation data was **0.1692**; hence the changes observed were not statistically significant.

Bangladeshi girls had a mean epigenetic age estimate of 7.4 years with the skin and blood clock. DNA methylation data from first-generation girls returned a mean epigenetic age estimate of 10.2 years. For second-generation girls the mean was 7.7 years and among white British girls the mean epigenetic age estimate was 8.8 years. R value for results from the skin and blood clock were largely consistent with the correlation across epigenetic age estimates and chronological

age, as illustrated in **figure 14**. The results obtained from this clock across all four populations suggests that it is a well-calibrated tool to predict age based on DNA methylation information.

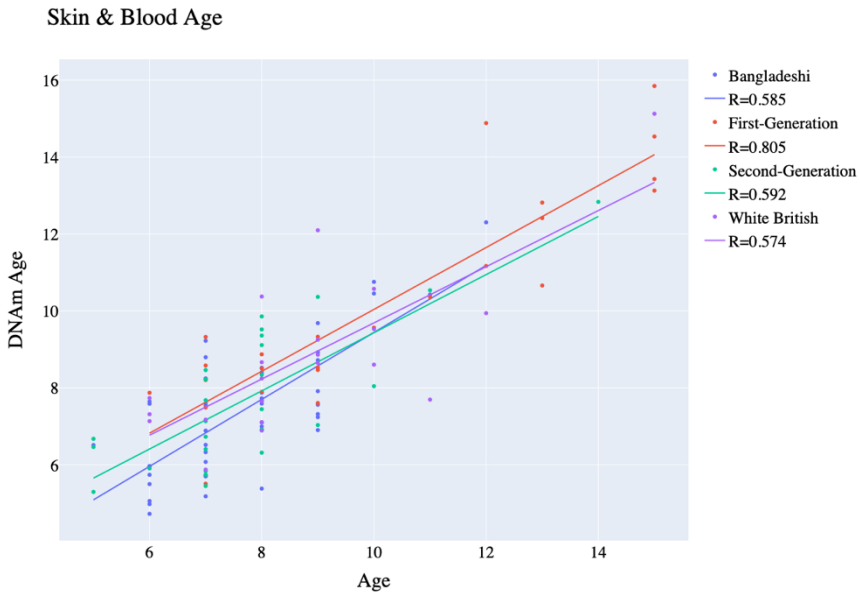


Figure 14. Skin and Blood clock epigenetic age estimates plotted against individuals' chronological ages. Graph shows DNA methylation-based age in years against chronological age in years for four populations of girls aged between 5 and 16 years. Bangladeshi (blue), R=0.585; first-generation (orange), R= 0.805; second-generation (green), R=0.592; white British (purple), R=0.574. All four populations matured at a similar rate. The four groups have a similar gradient, which suggests that the clock is a well-calibrated tool used to predict an individual's age based on DNA methylation data.

The statistical significance of the results was calculated with the Welch ANOVA test, which returned a P-value of 0.1692. Given that the P-value was higher than 0.05, the difference observed between populations was not deemed to be statistically significant.

4.4. Hannum clock

Let us now turn our attention to the Hannum clock and its interpretation of our datasets. As explained earlier on, the Hannum clock's main functionality is to provide epigenetic age estimates based on DNA methylation data. However, the clock is made specifically for a mature and adult

cohort. The training datasets used to develop the Hannum clock included exclusively adults, focused on 71 CpG sites and it was based on blood samples (Hannum et al., 2013).

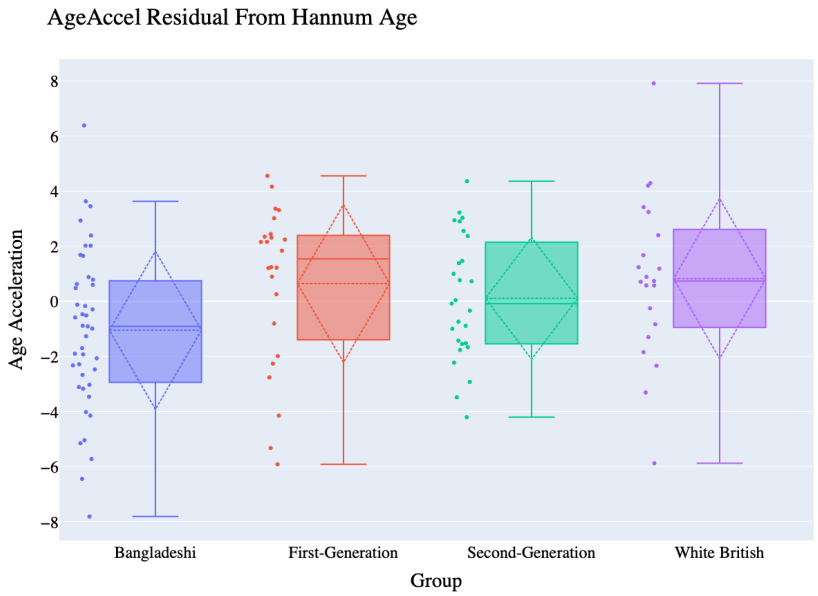


Figure 15. Age acceleration based on data derived from Hannum clock. Box plot shows the differences in pace of epigenetic ageing against four populations of girls aged between 5 and 16 years. Bangladeshi (blue); first-generation (orange); second-generation (green); white British (purple). The median values are indicated by a darker horizontal line in that goes through each of the different group's box plot. A positive or negative age acceleration value indicates that the estimated epigenetic age of the samples is higher or lower, respectively, than expected. All four populations matured at a similar rate. The P-value for the Hannum clock based on normalised methylation data was **0.0524**; changes observed were not statistically significant.

Given the nature of our own samples – paediatric and buccal – we expected the deviation between chronological age and estimated epigenetic age to be stark. Nevertheless, Hannum's clock was able to account for the discrepancy in age acceleration across our different groups, as can be seen in **figure 15**. Not only did Hannum's clock account for it, but it also estimated a similar pattern in terms of age acceleration across the four cohorts of girls. First-generation Bangladeshi girls displayed a positive median age acceleration (1.54), much higher than the other groups of girls. Bangladeshi girls living in Bangladesh exhibited a negative median age acceleration (-0.91).

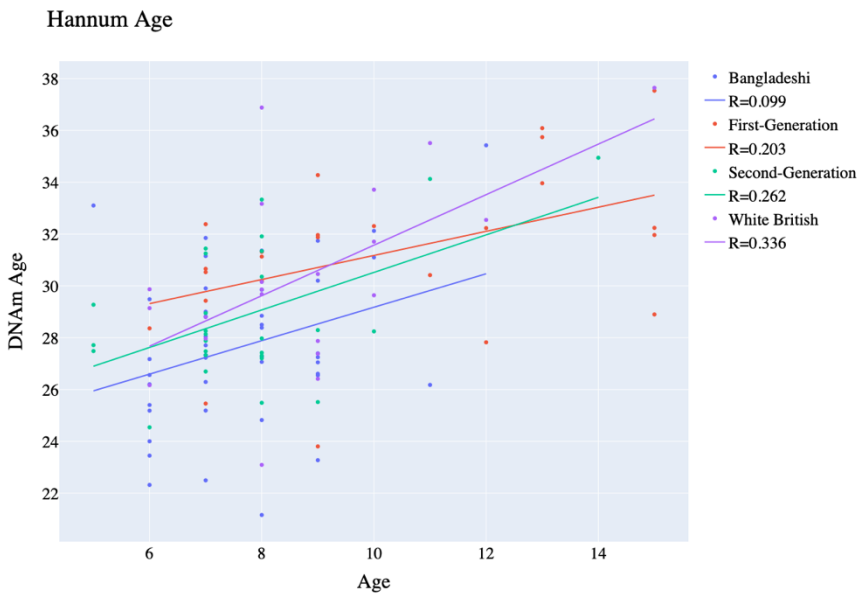


Figure 16. Hannum clock epigenetic age estimates plotted against individuals' chronological ages. Graph shows DNA methylation-based age in years against chronological age in years for four populations of girls aged between 5 and 16 years. Bangladeshi (blue), R=0.099; first-generation (orange), R= 0.203; second-generation (green), R=0.262; white British (purple), R=0.336. The four populations matured at different rates, while the epigenetic age predictions did not correlate to the chronological ages across the populations. The findings suggest the clock is not a well-calibrated tool to predict an individual's age based on DNA methylation data from buccal cell samples.

The mean epigenetic age estimates for Bangladeshi girls based on Hannum's clock was 27.7 years. For first-generation Bangladeshi migrants to the UK, Hannum's clock returned a mean of 31.3 years. While for second-generation Bangladeshi girls born to first-generation immigrant parents, the mean estimated epigenetic age was 28.9 years. White British girls had a mean epigenetic age prediction of 30.4 years.

The statistical significance of the results was calculated with the Welch ANOVA test, which returned a P-value of 0.0524. Given that the P-value was (slightly) higher than 0.05, the difference observed between populations was not deemed to be statistically significant. The Hannum clock might have yielded statistically significant results, but the age predictions were largely inaccurate and far higher than the chronological ages of the populations analysed in this study. Hence, we issue a note of caution to readers that believe there are significant conclusions to be drawn from

these results. Nevertheless, given the statistical significance the clock was able to establish, Hannum's clock could be applied to a larger dataset derived from a more mature population, given that it was designed primarily for the study of adult (and blood) samples. It could also be used in longitudinal studies where samples collected from individuals' (i) blood samples and (ii) collected when they are adults. Given that the age estimates were largely inconsistent with the chronological ages of individuals, the PedBE clock remains the most suitable tool.

As illustrated in **figure 16**, Hannum's epigenetic clock predicted higher epigenetic ages for our cohort than their chronological ages. Given the large discrepancy, and the limitations of this clock due to the reasons outlined above, a comparison among R values can be beneficial towards enhancing our understanding of the DNA methylation data across the four different groups. All four R values were unsatisfactory and closer to 0 than 1 (Bangladeshi: 0.099; first-generation: 0.203; second-generation: 0.262; white British: 0.336). Interestingly, white British girls had the closest – although still miles away – epigenetic age estimate to their chronological age.

4.5. PhenoAge clock



Figure 17. Age acceleration based on data derived from PhenoAge clock. Box plot shows the differences in pace of epigenetic ageing against four populations of girls aged between 5 and 16 years. Bangladeshi (blue); first-generation

(orange); second-generation (green); white British (purple). The median values are indicated by a darker horizontal line in that goes through each of the different group's box plot. A positive or negative age acceleration value indicates that the estimated epigenetic age of the samples is higher or lower, respectively, than expected. All four populations matured at a similar rate. The P-value for the PhenoAge clock based on normalised methylation data was **0.0348**; changes observed were statistically significant.

The PhenoAge clock is well known as one of the trailblazers among the second-generation epigenetic clocks, developed to provide a health-focused measure of biological ageing. It takes into account many health-related biomarkers, including C-reactive protein, which works as an inflammation biomarker; albumin, a protein that illustrates an individual's nutritional status; and creatinine, a biomarker for kidney function. Although Levine et al.'s clock does not explicitly incorporate an individual's cortisol levels, it measures other factors linked to chronically elevated cortisol levels such as increased inflammation, metabolic dysfunction and insulin resistance.

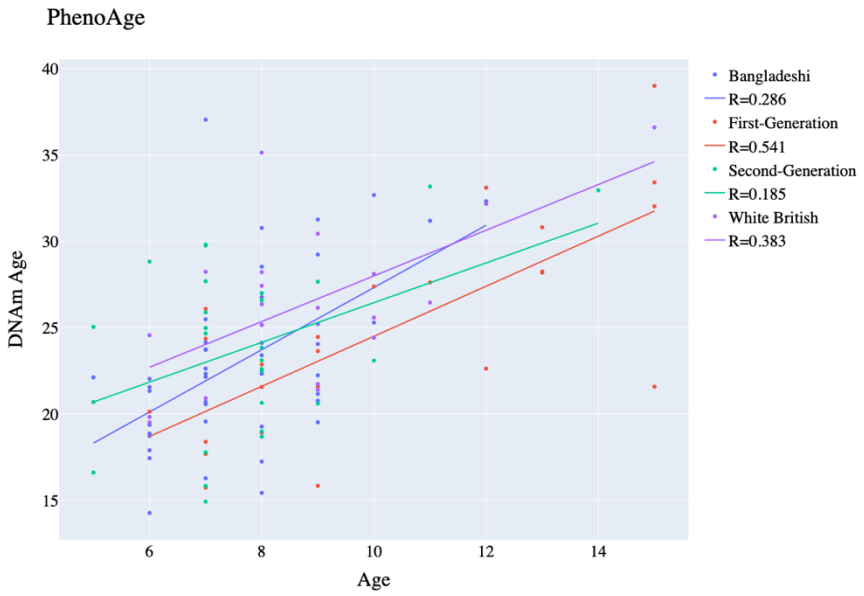


Figure 18. PhenoAge clock epigenetic age estimates plotted against individuals' chronological ages. Graph shows DNA methylation-based age in years against chronological age in years for four populations of girls aged between 5 and 16 years. Bangladeshi (blue), R=0.286; first-generation (orange), R= 0.541; second-generation (green), R=0.185; white British (purple), R=0.383. The four populations matured at similar rates, while the epigenetic age predictions did not correlate to the chronological ages across the populations. The findings suggest the clock is not a well-calibrated tool to predict an individual's age based on DNA methylation data from buccal cell samples.

As illustrated in **figure 17**, positive age acceleration in this epigenetic clock can suggest that an individual is likely to experience negative health outcomes. Second-generation Bangladeshi and white British girls were the only two groups to display positive age acceleration with the PhenoAge clock. The median age acceleration for white British girls was 1.5, while for second-generation Bangladeshi girls the median age acceleration was 0.93. Both Bangladeshi girls and first-generation migrants had negative age acceleration with the PhenoAge clock, with the median age acceleration value for Bangladeshi girls being -0.27 and for first-generation girls being –1.12.

The statistical significance of the results was calculated with the Welch ANOVA test, which returned a P-value of 0.348. Given that the P-value was lower than 0.05, the difference observed between populations was deemed to be statistically significant.

The training dataset used to develop the PhenoAge clock focused on information from 9,926 adults in the NHANES III - an American representative sample (Levine et al., 2018). The epigenetic age estimates, presented in **figure 18**, reflect the lack of paediatric information used to develop the PhenoAge clock. The clock concluded that Bangladeshi girls had a mean epigenetic age of 23.1 years. First-generation Bangladeshi migrants to the UK had a mean epigenetic age of 24.8 years with the clock. Data from white British girls returned a mean epigenetic age estimate of 26.4 years, while second-generation Bangladeshi girls born in the UK to immigrant parents had a mean predicted epigenetic age of 23.8 years, according to the PhenoAge clock.

4.6. LinAge clock

Results from the LinAge clock concluded that Bangladeshi girls had a mean epigenetic age of 25.3 years. For first-generation Bangladeshi migrants to the UK, LinAge returned a mean epigenetic age of 28 years. Data from second-generation Bangladeshi migrants to the UK run by LinAge concluded their predicted mean epigenetic age was 24.8 years. While white British in the UK had an estimated mean epigenetic age of 24.4 years. **Figure 19** shows the dispersion of age

acceleration across the four populations, while **figure 20** illustrates that some of epigenetic age estimates from Lin's clock were higher than 40 years.

Lin et al. developed the clock in their quest to better understand whether measuring DNA methylation levels in a reduced subset of CpG sites can produce similar results outputs in contrast with the methylation levels from a larger set of CpG sites. To carry out this research Lin et al. developed a 99 CpG site epigenetic clock and compared it to a clock that has 3 CpG sites (Lin et al., 2016; Weidner et al., 2014). The conclusion from their findings was that epigenetic clocks looking at few, or even 1, CpG sites can competently provide epigenetic age estimates.



Figure 19. *Age acceleration based on data derived from LinAge clock.* Box plot shows the differences in pace of epigenetic ageing against four populations of girls aged between 5 and 16 years. Bangladeshi (blue); first-generation (orange); second-generation (green); white British (purple). The median values are indicated by a darker horizontal line in that goes through each of the different group's box plot. A positive or negative age acceleration value indicates that the estimated epigenetic age of the samples is higher or lower, respectively, than expected. All four populations matured at a similar rate. The P-value for the Lin Age clock based on normalised methylation data was **0.2560**; changes observed were not statistically significant.

The statistical significance of the results was calculated with the Welch ANOVA test, which returned a P-value of 0.2560. Given that the P-value was not lower than 0.05, the difference observed between populations was not deemed to be statistically significant.

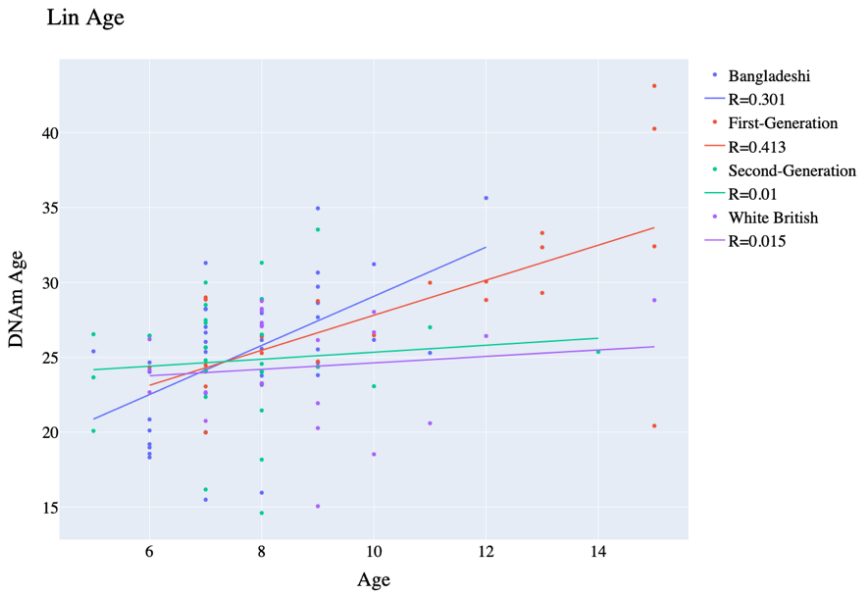


Figure 20. LinAge clock epigenetic age estimates plotted against individuals' chronological ages. Graph shows DNA methylation-based age in years against chronological age in years for four populations of girls aged between 5 and 16 years. Bangladeshi (blue), R=0.301; first-generation (orange), R= 0.413; second-generation (green), R=0.01; white British (purple), R=0.015. The four populations matured at similar rates, while the epigenetic age predictions did not correlate to the chronological ages across the populations. The findings suggest the clock is not a well-calibrated tool to predict an individual's age based on DNA methylation data from buccal cell samples.

4.7. Summary of findings

Epigenetic clock	Parameters assessed	Correlation with chronological age? Yes/No	Observations
Horvath's clock 353 CpG sites *First two data entries used normalised and raw data, respectively but were processed by Horvath's online tool as 'raw'.*	Multi-tissue DNA methylation-based age predictor based on paediatric and adult samples from all cell types – apart from sperm	No	Raw data tests performed better than normalised data attempts, but the DNA methylation-based age estimates deviated largely from the chronological age of our samples.
Horvath's clock 353 CpG sites *Third entry. Guidance sought to understand discrepancy across results. Tutorial guidance for Horvath's online tool explained that datasets should be normalised by the tool to allow for consistency with the training dataset for the clock.*	Multi-tissue DNA methylation-based age predictor based on paediatric and adult samples from all cell types – apart from sperm	Yes	DNA methylation-based results from Horvath's clock were largely equivalent to the chronological age, but not as accurately as we had expected.
PedBE Age 94 CpG sites	Paediatric DNA methylation-based age predictor trained on buccal samples – buccal epithelial cells in particular	Yes	Results from the PedBE clock were encouraging and demonstrated the benefits of carrying out research on a mathematical model tailored to analyse DNA methylation in paediatric buccal samples
PhenoAge 513 CpG sites	DNA methylation-based age predictor	No	The DNA methylation-based age estimates

	that considers clinical biomarkers, such as white blood cell count		provided by the PhenoAge clock were inconsistent with the chronological ages from our samples. It predicted that some 6 to 7-year-old girls were epigenetically older than other 9 to 10-year-olds belonging to the same group.
Skin and blood Age 391 CpG sites	DNA methylation-based age predictor developed using skin and blood tissues from paediatric samples with HGPS	Yes	Epigenetic age estimates from the skin and blood clock largely correlated to the chronological age of our samples. The CpG the mathematical model focused on – although tailored for HGPS patients – were able to inform us as it was designed based on paediatric samples.
Hannum Age 71 CpG sites	DNA methylation-based age predictor trained on blood tissue from adult samples	No	DNA methylation-based age predictions from Hannum's clock deviated from the chronological age across all four groups – with the starkest inconsistencies among white British girls.
Lin Age 99 CpG sites	DNA methylation-based age predictor focused on identifying relevant genetic and environmental risk factors for optimum health	No	DNA methylation-based age predictions from Lin et al.'s clock deviated from the chronological age across all four groups. The findings were not deemed statistically significant.

Table 1. Summary of findings from epigenetic clocks. The table shows a summary of the main findings from results gathered across different epigenetic clocks. The four columns describe the key characteristics of each epigenetic clock, what cell and tissue types were used to develop the tool, whether the correlation with chronological ages was of any statistical significance and key observations from the results regarding the clock's measurements of DNA methylation levels from this study's cohort.

Acknowledging that the statistical power across epigenetic clocks was not great, the results above demonstrate a trend rather than conclusive results.

5. Discussion

The results summarised in **Table 1** were obtained in the duration of the MRes degree at the School of Biosciences, University of Nottingham, provide an independent and original analysis bridging advances in epigenetic ageing research and the information previously collected by colleagues at the University of Durham on the socio-economic characteristics and key features of the populations studied. In the analysis presented in this thesis, trends that corroborate previous conclusions from Houghton and colleagues were observed. They have argued that migrant communities face some biological trade-offs in comparison to their non-migrant counterparts – independent of nationality and ethnic background, this will be explored further in this portion of the thesis (Houghton et al, 2014a).

5.1. Hurdles with Horvath clock submissions

We carried out several submissions to Horvath's open-source online clock several times before harvesting the results presented in this thesis. Horvath's epigenetic clock has clear instructions available on the website but if any of the fields on the submission file are incorrect or incomplete, the mathematical model is unable to return a result and it does not have an alert system to explain the error. For experts who are literate in bioinformatics, the process of submitting DNA methylation data to each clock is straightforward. But given the interdisciplinary nature of the work we endeavoured to carry out, and the limitations due to a lack of domain-specific knowledge, gathering the results to aid our discussion was a difficult task.

Our first submission to Horvath's clock led to some alarming results. The DNA methylation-based age estimates did not correlate to the chronological ages of the individuals whose samples we had used. As discussed in the results section, many other clocks were also unable to provide age estimates that correlated to individuals' chronological ages. But one factor persisted to concern us after we were able to collect a larger range of results from different clocks: the sex prediction. Horvath's 2013 multi-tissue clock returned age acceleration information, a breakdown of the data, tissue or cell type prediction and sex prediction. The clock accurately predicted that the samples used for the analysis were from buccal cells. However, it could not detect that our

samples were female. Instead, it returned results ranging between 'male' and 'unsure'. One possible justification for these results is that low levels of DNA methylation at the key 353 CpG sites impacted the readings. Lower levels of methylation are often associated with male samples, as the array covers CpG sites on the X chromosome. As females have two X chromosomes and males only have one, it is possible that the model inferred that the low levels indicate male samples. The strategy was adapted, and raw DNA methylation data was used instead of the normalised data, which marginally improved results (as demonstrated in **figure 6** and **figure 8**).

The trend observed in Horvath's clock across our paediatric samples is consistent with the weighted average of the 353 CpG sites used by the clock plotted against the chronological age in the training datasets.

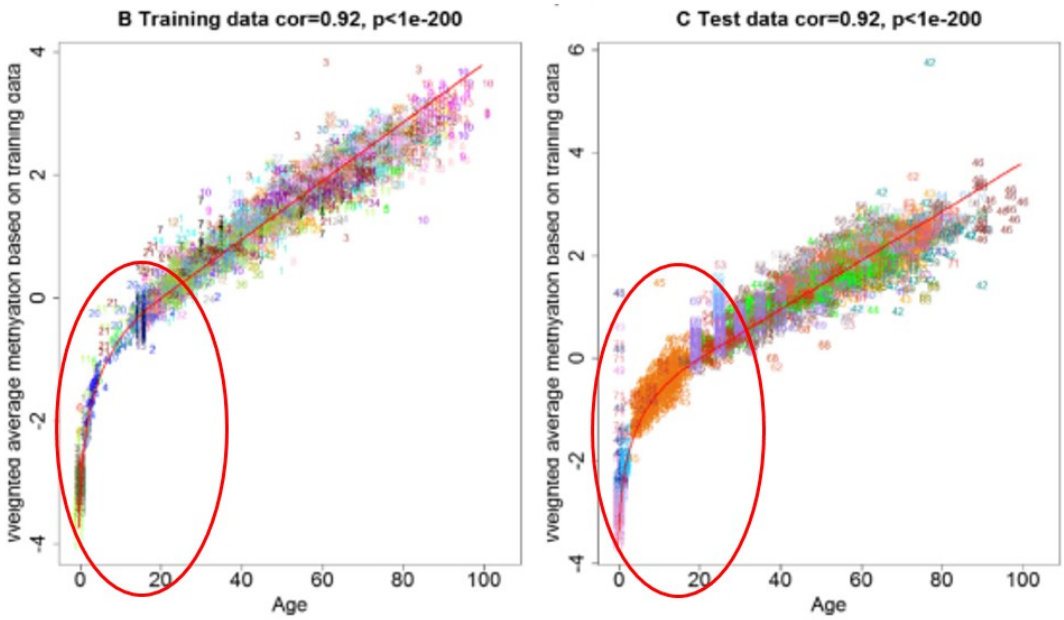


Figure 21. Horvath's 2013 DNA methylation-based epigenetic clock's weighted methylation average plotted against chronological age in years. Figure shows an oval red ring around the 0 to 20 years age bracket, as the age acceleration rate has a steep gradient until subjects are aged 20 or over, which suggests the multi-tissue epigenetic clock designed by Horvath is not as accurate among paediatric populations. Figure adapted from Horvath's 2013 paper: DNA methylation age of human tissues and cell types (Horvath, 2013).

The close analysis of Horvath's findings and weighted methylation averages across individuals aged 0 to 20 years elucidates the interpretation of the results presented in this thesis. As

highlighted in **figure 21**, the age acceleration predictions based on the chosen 353 CpG sites suggest a faster acceleration rate among paediatric individuals and young adults, particularly the age group between 0 and 20 years (Horvath, 2013). Nevertheless, Horvath's epigenetic clock is a trusted tool and has been widely used since its development in 2013. Another observation that can be drawn from **figure 21** is that the age acceleration rate stabilises based on the weighted methylation of individuals aged over 20 – and it goes on to follow a consistent trend across all age groups up until the age of 100 years.

5.2. Support from bioinformaticians

Our research could have been improved by a deeper and more comprehensive appreciation across the team for bioinformatics, in particular the machine learning mathematical models behind the epigenetic clocks we used to collect our DNA methylation-based age estimates. Nonetheless, the MRes candidate's second supervisor is a bioinformatician and he supported the findings from this study through the provision of insights into the discipline and support for data harvesting from lesser-known epigenetic clocks, which enriched the findings presented in this study. Second supervisor Richard Emes supported the candidate in obtaining results from Hannum clock, Skin and Blood clock, PhenoAge clock, Horvath's clock and LinAge clock. While supervisor Reinhard Stöger provided analytical insights and greater understanding of the biological mechanisms in the field of epigenetics, as well as sharing expertise on how best to use well-established tools such as Horvath's clock.

As outlined in the methods section, colleagues Rachel Clifford and Poojita Rajasekar supported our project by providing DNA methylation-based age estimates for the PedBE clock. The data helped deepen our analysis of DNA methylation levels and its relationship with epigenetic ageing, stress and socio-economic environments.

5.3. Limitations

Additional time and funding would have allowed for more research exploration using the EPIC array technology over a longer period, as the results from the array are improved by a larger (and

higher quality) spread of DNA samples from diverse tissue and cell types (Dedeurwaerder et al., 2014).

More time and specialist training would have also contributed to allowing us to carry out analytical investigations into overlapping CpG sites across the different epigenetic clocks. Specialist bioinformaticians can train and equip non-specialist researchers to retrieve the best quality data are differential methylation analysis and functional enrichment analysis, as that aids us to extract meaningful insights from the results.

In this study, it would have been preferable to gather the buccal samples ourselves and ensure the quality of the DNA was as high as possible. The analysis of our hypothesis would also have been improved by access to a larger sample size, as that would contextualise outlier results among our 115-strong cohort. Buccal samples are preferrable among such a young cohort because they are far less intrusive. But access to other tissues from the same individuals would have supported our research findings, as we would have been able to adapt our strategies and deploy blood samples to epigenetic clocks that were developed using blood samples – like the skin and blood clock, as well as Hannum clock.

Going forwards researchers would benefit from carrying out DNA methylation-based age analysis with longitudinal data from the cohort playing a protagonist role in their investigation into epigenetic ageing. A **longitudinal analysis** would provide more pertinent and better-informed information on the use of epigenetic clocks as a biological age predictor. Access to longitudinal data samples gathered for the sole purpose of measuring DNA methylation levels at key CpG sites would assist us in better understanding the value of epigenetic clocks as biomarkers of ageing, as argued by Jylhävä and colleagues in their 2017 paper (Jylhävä, Pedersen and Hägg, 2017). In addition to this, more time would be valuable towards gaining a deeper understanding of the key functionality at CpG islands that overlap across the epigenetic clocks studied in the duration of this thesis, because it could shed light not solely on the most relevant biomarkers of ageing but also enhance the research community's understanding of gene expression at these intersections of CpG sites and gene promoters – among other mechanisms.

5.4. Epigenetic significance of our findings

As articulated in the results section, the findings presented in this thesis inform the research community of a positive age acceleration trend that appears to target migrant communities. **Figure 22**, below, illustrates the different gradients of epigenetic age acceleration depending on the epigenetic clock used to measure the same dataset. The linear regression lines below show the different gradients of DNA methylation-based age predictions plotted against the chronological ages of first-generation Bangladeshi girls who have migrated to the UK during their paediatric developmental stages, and a time when girls reach pubertal milestones such as adrenarche, thelarche, pubarche and menarche.

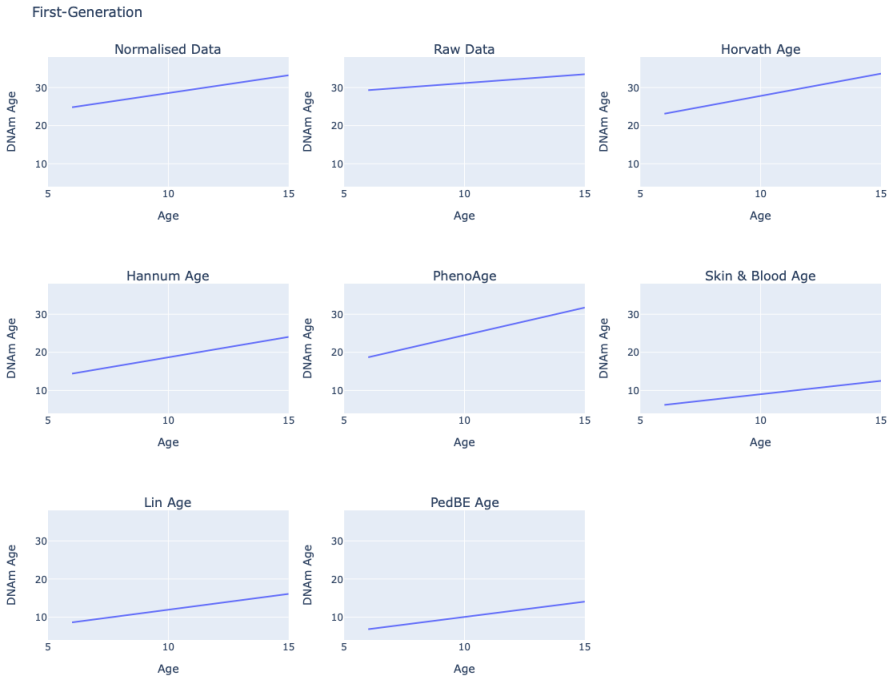


Figure 22. Linear regression of DNA methylation-based age estimates plotted against the first-generation population's chronological ages. The data was obtained from different epigenetic clocks to illustrate how well calibrated the clocks are to assess DNA methylation levels at age informing CpG sites from paediatric buccal samples. From top left across (A) normalised data results from Horvath's clock (B) raw data results from Horvath's clock (C) data results from Horvath's clock (D) data results from Hannum's clock (E) data results from PhenoAge clock (F) data results from Skin and Blood clock (G) Data results from LinAge clock (H) data results from PedBE clock. As illustrated above, the PedBE clock was better calibrated than other clocks to accurately predict the epigenetic age estimates

from this study's cohort. Similar figures showing the linear regression lines across different clocks for the other three groups can be found in the Appendix.

The early development of breast tissue during puberty may be indicative of earlier puberty and therefore an indicator of more rapid ageing, which may in turn contribute towards elevated epigenetic age acceleration levels across first-generation Bangladeshi girls (Song et al., 2015). First-generation Bangladeshi girls who migrate to the UK were identified as reaching thelarche – key pubertal stage that leads to the development of breast tissue – earlier than their second-generation and white British counterparts. The age at which girls reached thelarche was assessed by self-report using a modified version of the pubertal development scale (Houghton et al., 2014b). These findings were outlined in work carried out by Houghton and colleagues, who investigated the relationship between pubertal timing in British-Bangladeshi girls and their risks of breast cancer in later life (Houghton et al., 2014b).

These findings are helpful towards our contextual understanding of the cohort, particularly in terms of epigenetic mechanisms. As discussed by Sehl and colleagues, the DNA methylation levels across CpG sites of female breast tissue indicate that the tissue has an accelerated rate of ageing (Sehl et al., 2017). In understanding that first-generation Bangladeshi migrants to the UK reached thelarche before their non-migrant counterparts and appreciating that female breast tissue has an accelerated rate of ageing, there is a possibility that this key pubertal development stage may be linked to broader epigenetic age acceleration rates. The samples studied in this paper were saliva samples, hence there are no empirical findings to indicate whether the epigenetic age acceleration of breast tissue across the cohort correlates with the findings from Houghton et al.'s study (Houghton et al., 2014b).

Given that the samples studied in this paper are from young girls experiencing several changes, it was not deemed appropriate nor necessary to extend the research to investigate samples that could only be collected through extremely intrusive measures. In addition to several ethical considerations, the study's commitment to due diligence and protecting the welfare of all participants – particularly paediatric ones – means that is an area of study that will not be

explored further. Moreover, as the saliva samples were collected more than a decade ago, the analysis would not be statistically significant – with the sample size available - in our quest to understand the impact of childhood environments on epigenetic age acceleration.

5.5. Migration and epigenetic ageing

Our results illustrated a trend of accelerated epigenetic ageing across first-generation Bangladeshi migrants. But when the results from all clock submissions were compared to first-generation girls' white British, Bangladeshi and second-generation counterparts, the results were not deemed statistically significant. However, they did demonstrate a trend and further investigation with better-quality DNA samples across a longitudinal study could provide more answers. As presented by Stöger et al. in their 2023 paper, adult Bangladeshi women who spent their childhood in Bangladesh and later emigrated and moved to the UK presented higher epigenetic age acceleration rates than their non-migrant second-generation counterparts (Stöger et al., 2023). Therefore, it could be the case that statistically significant age acceleration among migrant communities becomes pronounced in their later years, although the trends can begin to be observed when they are aged between 5 and 16 years – as demonstrated in this study.

5.6. Migration and its significance in understanding environmental stressors

The buccal epithelial cell samples used in this study date back to as early as 2009 (Houghton et al., 2014a). As outlined previously, this thesis bridges the body of work on developmental differences between Bangladeshi girls of differing migration status and their non-migrant contemporaries both in the UK and Bangladesh – while investigating whether their socio-economic background and circumstances impact these physiological changes. Given that the datasets used in the DNA methylation analysis for this paper were a subset of the data harvested and presented in Houghton et al.'s paper, the socio-economic findings presented in their 2014 paper are of particular significance to the analysis of this thesis (Houghton et al., 2014a).

	Sylheti		First-Generation		Second-Generation		European		P for trend; Residency Scale	Pairwise comparison Sylheti vs. UK resident	Pairwise comparison First-generation vs. UK born
N	165		42		162		50				
Age (years)	9.7	(3.0)	10.5†	(3.4)	10.1	(2.7)	9.7	(2.5)	0.13	0.16	0.14
Socioeconomic Variables											
Household											
Household size (people)	6.6	(2.6)	5.8†	(2.2)	5.9†	(1.7)	4.2†	(1.3)	<0.01	<0.01	0.70
Family owns house	100	52%	8†	23%	51	42%	7†	19%	<0.01	<0.01	0.06
Family owns a car	29	17%	20†	50%	109†	71%	28†	57%	<0.01	<0.01	0.01
Single Mother Household	30	17%	2	5%	22	18%	19†	44%	<0.01	0.10	0.00
Employment											
Father (in households with both parents)	133	94%	22†	67%	80†	77%	16†	76%	<0.01	<0.01	0.23
Mother	33	19%	8	21%	39†	26%	20†	44%	0.04	0.01	0.14
Education (college or higher)											
Father (in households with both parents)	29	32%	16†	57%	35	45%	10†	59%	0.02	0.05	0.50
Mother	18	16%	13	39%	56	42%	25	76%	<0.01	<0.01	0.33
Anthropometrics											
Height (cm)											
height for age z-score	−1.2	(1.3)	−0.8†	(1.0)	−0.2†	(1.3)	0.1†	(1.2)	<0.01	<0.01	0.01
<9.5 yr.	116.4	(9.8)	123.6†	(8.9)	124.8†	(9.9)	126.5†	(7.9)	<0.01	<0.01	0.64
>9.5 yr.	144	(11.3)	148.7	(7.7)	147.9	(9.8)	144.9	(11.8)	0.24	0.08	0.63
Weight (kg)											
weight for age z-score	−1.2	(1.6)	0.1†	(1.4)	0.3†	(1.3)	0.8†	(1.2)	<0.01	<0.01	0.17
<9.5 yr.	20.5	(5.8)	28.4†	(10.0)	28†	(8.3)	29.6†	(6.4)	<0.01	<0.01	0.69
>9.5 yr.	37.8	(10.6)	47.1†	(10.2)	45.9†	(12.3)	44.9†	(13.8)	<0.01	<0.01	0.70
Waist Circumference (cm)											
waist for age z-score	−1.8	(9.3)	0.7†	(1.5)	1.0†	(1.2)	1.4†	(1.0)	<0.01	<0.01	0.08
<9.5 yr.	48.9	(6.4)	59.8†	(9.4)	58.7†	(7.4)	61†	(5.4)	<0.01	<0.01	0.76
>9.5 yr.	58.3	(6.9)	65.8†	(8.6)	67.8†	(9.3)	68.1†	(9.1)	<0.01	<0.01	0.35
BMI (kg/m^2)											
BMI for age z-score	−0.6	(1.3)	0.7†	(1.4)	0.6†	(1.2)	1.0†	(1.1)	<0.01	<0.01	0.89
<9.5 yr.	15.0	(3.1)	18.6†	(4.3)	17.7†	(3.2)	18.5†	(2.5)	<0.01	<0.01	0.39
>9.5 yr.	18.0	(3.1)	21.2†	(3.6)	20.5†	(3.6)	21.4†	(4.0)	<0.01	<0.01	0.57

Figure 23. Table demonstrating socio-economic variables and sample characteristics of the four groups of populations analysed in the study carried out by Houghton and colleagues. Characteristics were reported as mean for continuous variables or number and percentage for categorical variables. The solid red rectangles border the two columns relevant to this study: the socio-economic variable measures across first-generation Bangladeshi girls and the pairwise comparison between first-generation and UK born girls. The figure was adapted from Houghton et al.'s 2014 paper, which investigated childhood environment influences on adrenarcheal timing among first-generation Bangladeshi migrant girls to the UK (Houghton et al., 2014a).

The socio-economic factors measured by Houghton and colleagues can be accessed in full via their paper; this discussion will focus on the families' employment and education levels. Interestingly, despite the significantly high education levels among the parents of first-generation migrant Bangladeshi girls, their employment is considerably lower than the other three groups. There are several factors that may influence this phenomenon, and we will consider them in turn shortly, but first let us detangle the education and employment statistics across each population. 19% of the mothers of Sylheti girls living in Bangladeshi are employed, while 94% of these girls' fathers are employed. For second-generation Bangladeshi girls who were born in the UK to

immigrant parents, 26% of their mothers are employed and 77% of their fathers are employed. The figure changes again for white British (labelled European in **figure 23**), as 76% of fathers are employed and 44% of mothers are employed. Across first-generation Bangladeshi girls who moved to the UK during their childhood, the employment rates are overwhelmingly lower – 67% of their fathers are employed while 21% of mothers in this group work (Houghton et al., 2014a).

In contrast, the first-generation group can be described as well-educated, with 57% of fathers from the group having college degrees or higher and 39% of mothers having gone at least to college, if not carried on with further education. The parents of white British girls had the highest education levels (mothers: 76%, fathers: 59%). For Bangladeshi families in Sylhet, 32% of fathers were educated above college level, and 16% of mothers had gone to college or pursued further studies beyond college. Second-generation parents had a more even split among themselves, with 45% of fathers and 42% of mothers having studied at least to college level (Houghton et al., 2014a).

Myriad factors could be behind these statistics. But it is valuable to draw attention to a couple of facts illustrated in **figure 23**. First-generation Bangladeshi fathers fall only short of their white British counterparts in what concerns education levels, as they have the second highest college or above education across the four populations of fathers. Nevertheless, these well-educated men have the lowest employment rate among fathers in the other three groups. This observation becomes even more pertinent when we contrast it to the employment levels for fathers in Bangladesh – a demographic first-generation fathers would have belonged to prior to their migration to the UK – who have the highest employment rate across all groups a staggering 94% employment (Houghton et al., 2014a).

One explanation for this discrepancy is the environment Global South migrants are greeted to when they arrive in the Global North, often in pursuit of a better quality of life for many reasons (better financial prospects, access to healthcare, education opportunities, or fleeing persecution, to name a few). Global South migrants that emigrate countries that have some degree of political stability tend to be dislocating themselves and their families in pursuit of better socio-economic outcomes including, but not limited to, better job and financial prospects as well as better access

to public health infrastructures that may not be as widely available in the country they are emigrating from. The education levels of parents in a position to emigrate their native countries with young children is often higher than their non-migrant counterparts, namely because – as observed by Houghton and colleagues in their study – immigrant families that move to the UK tend to be of sit comfortably in the middle classes of the countries they originate from. Of course, there are exceptions to this, particularly in the case of asylum seekers seeking refuge and families from other European countries who moved to the UK prior to Brexit. The employment requirements to obtain a right-to-work visa in the UK if you come from most Global South countries – and now some EU countries – is prescriptive and the pre-requisites targets and profiles individuals with access to disposable income and high levels of education.

The visa fees alone are significant and more than what most UK households would be able to set aside under usual circumstances. To obtain the visa that allows individuals to work – now known as the skilled worker visa – applicants must: work for a UK employer approved by the Home Office, have a certificate of sponsorship from their employer, and be paid a minimum salary (arbitrarily) determined by the government in accordance with the career path the individual is in. In its guidance to visa applicants, the Home Office clearly states that individuals must not access public funds – most benefits – and they must pay yearly healthcare fees, as well as proving that they have sufficient savings (UK Home Office website, retrieved August 2023).

However, back in 1948, the UK government passed the British Nationality Act which gave people from British colonies the right to live and work in the UK. The passing of this law was largely motivated by the government's need to backfill workers during post-war labour shortages. The dichotomy between the way in which the UK treats and perceives migrants from the Global South – including former colonies – and how heavily it relied on their labour after World War Two is a focal point underlying much of the discussion at the intersection of migration, stress, accelerated ageing and the impact hostile environments have on those three phenomena.

Although the UK did not colonise Bangladesh as we understand it today, it did colonise the territory of present-day Bangladesh when it colonised India. Bangladesh used to be part of British India from the mid-18th century until 1947, when India gained independence from the UK. The

Indian subcontinent was then divided into two nations: India and Pakistan. Pakistan itself was divided into West Pakistan and East Pakistan, which correlate to present-day Pakistan and Bangladesh respectively. It was only in 1971, that Bangladesh became a sovereign independent state after a war of independence (Raghavan, 2013). Bangladesh is a member of the Commonwealth, but as recently demonstrated during the Windrush Scandal, Commonwealth membership no longer guarantees immigration rights or citizenship status in the UK – despite its past colonial dealings that led to the very creation of the Commonwealth. The Windrush Scandal can be used as a focal point to explore the Global South and Global North relations because of the hostility many Caribbeans who had moved to the UK prior to or up until the 1970s – some at Britain's in its rebuilding efforts after World War Two – were wrongly deported to Caribbean nations including Trinidad, Barbados and St Lucia due to Home Office failings in properly recording the details of people granted permission to live and stay in the UK.

A recent study by Essex and colleagues investigated the impact of a decade of hostile environment immigration policies in the UK on the population's health outcomes. The conclusion was largely damming toward the current leadership, as the research group identified that hostile environment tactics including the requirement from healthcare workers and landlords to share individual's migration status with the government, and reporting whether an individual in debt for more than a couple of months at a time, have had detrimental effects to the UK's healthcare system. Essex and colleagues argue that the past decade – led by the Conservative government – has caused destitution and ill-health; they go on to argue that the current state of play undermines the founding principles of the National Health Service (Essex et al., 2022).

5.7. Cortisol, stress and epigenetic ageing

As discussed in chapter 2 of this thesis, the relationship between stress and migration manifests in the physical body and can influence gene expression near key CpG sites. It is well-documented that stable childhood environments work as an antidote to stressors (Suor et al., 2015), which in turn lead to high and sustained cortisol levels that can impact key CpG sites. For many migrant children living in the UK, the employment status of their parents is a stressor. That is because, in a hostile environment Britain, it matters to migrant children whether or not their parents are in work, or able to prove to the government that their earnings continue to meet the minimum criteria, and that their employer continues to sponsor the quality of their work. For many non-migrant communities reaching key developmental milestones in the same geographical environment, parents' employment status and ability to remain in a country they are contributing towards culturally as well as financially are largely irrelevant factors. Hence, the mental – and as expressed in this paper, biological – load on these young individuals can be judged as considerably higher than their non-migrant counterparts.

5.8. Key advantages to the research community based on the data presented in this thesis

The main advantages to the research community based on the findings presented in this thesis are that we have been able to bridge the work carried out in the realm of bio-cultural research and the impact of migration on key hormonal phases in a paediatric female cohort with information about DNA methylation levels. The approach factored in sustainability, as by utilising data collected more than a decade ago, we avoided duplication and reduced waste. The work presented in this thesis also draws on what is already known about the links between cortisol and stress, taking it one step further and factoring in migration and epigenetic ageing.

5.9. Philosophy of science, eugenics and epigenetics

The work undertaken in this thesis was largely inspired by initial observations on the field of epigenetics from a philosophy of science background. The core argument presented at the conclusion of the undergraduate degree in the social sciences (unpublished) argued that there are no inherent biological differences between racialised groups, as is argued by eugenicist discourse. Rather, the dissertation explained, racism and eugenicist frameworks that have permeated through many vectors of society nurture hostile environments for racialised communities, and othered communities (Cursino, 2021). The theory proposed looked at epigenetic mechanisms for their explanatory power towards building a coherent understanding that racialised communities do not have worse health outcomes because they are inherently (or genetically) different than the socio-dominant group in a given scenario (Baedke, 2018); rather these othered groups suffer worse health outcomes because the environment they reside in was constructed to oppress and ensure hierarchical systems stay in place (Cursino, 2021). These findings are transferrable to those presented in this thesis, but instead of racialised groups it appears that migrant communities suffer the consequences of these hostile environments – particularly those that emigrate the Global South in pursuit of a better life in the Global North. However, one of the main limitations of the work carried out within philosophy of science was the lack of appreciation and robust understanding of the epigenetic mechanisms that may (or may not) be able to provide this explanation about socio-economic factors and the treatment of oppressed and marginalised groups in society. Hence why the discourse and iterative relationship between epigenetics and philosophy of science can be hugely beneficial towards building a clearer and more robust understanding of our bodies, the environments they reside in, and the relationship between the two.

Future considerations into the potential misuse of epigenetic age estimators by private companies must be considered, as explained in recent intersectional studies into the bioethics of epigenetics. Saulnier and Dupras argue that there are emerging concerns about the potential reification of race as a concept that is seriously considered and used as justification by those who still carry eugenicist frameworks and apply those to policymaking and access to healthcare

(Saulnier and Dupras, 2017; Saldaña-Tejeda, 2018); Singh, Morrison and Hoy, 2019). In a different study focusing on environmental epigenetics, Dupras and colleagues warn of the misuse of epigenetic clocks and the potential human rights infringements that may arise as a result of their use by ill-intentioned stakeholders (Dupras et al., 2019).

6. Conclusion

The results shared in this paper demonstrate that there is a discrepancy between the age acceleration rate of first-generation Bangladeshi migrant girls and their non-migrant counterparts - both white British populations and Bangladeshi populations. DNA methylation-based epigenetic clocks were used as biomarkers of biological ageing because of their ability to accurately correlate to individuals' chronological ages, as well as its age acceleration predictions that provide warnings on whether the age acceleration rates in a particular group is out of sync with others in the same cohort. The discrepancy across populations was not deemed statistically significant. But the interdisciplinary analysis provided in this thesis indicates that the age acceleration trend, measured across most clocks (particularly the PedBE clock) for first-generation migrant girls, functions as a warning to key stakeholders and policymakers on the detrimental health outcomes migrant communities that move from the Global South to the Global North may endure. The process of migration was identified as a stressor due to the instability associated with moving countries and attempting to integrate in a new community, which encompasses hurdles that include worries that their parents may not have suitable employment. Such stress at a young age may lead to accelerated epigenetic ageing, particularly due to the connection between environmental stressors and heightened cortisol levels that can impact dexamethasone responsive CpG sites and potentially silence or express certain genes. The resilience of migrant racialised communities in the UK is commendable and it would be largely beneficial if early interventions were put in place to reduce the stressors on young individuals going through several chances – from pubertal timing to settling in a new nation away from their well-trodden paths and childhood comforts.

7. References

Alexander, N., Kirschbaum, C., Wankerl, M., Stauch, B.J., Stalder, T., Steudte-Schmiedgen, S., Muehlhan, M. and Miller, R., 2018. Glucocorticoid receptor gene methylation moderates the association of childhood trauma and cortisol stress reactivity. *Psychoneuroendocrinology*, *90*, pp.68-75.

Baedke, J., 2018. *Above the gene, beyond biology: Toward a philosophy of epigenetics*. University of Pittsburgh Press.

Bar-Sadeh, B., Amichai, O.E., Pnueli, L., Begum, K., Leeman, G., Emes, R.D., Stöger, R., Bentley, G.R. and Melamed, P., 2022. Epigenetic regulation of 5α reductase-1 underlies adaptive plasticity of reproductive function and pubertal timing. *BMC biology*, *20*(1), p.11.

Bar-Sadeh, B., Rudnizky, S., Pnueli, L., Bentley, G.R., Stöger, R., Kaplan, A. and Melamed, P., 2020. Unravelling the role of epigenetics in reproductive adaptations to early-life environment. *Nature Reviews Endocrinology, 16(9)*, pp.519-533.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. *Genes & development*, *16*(1), pp.6-21.

Belsky, J., Ruttle, P.L., Boyce, W.T., Armstrong, J.M. and Essex, M.J., 2015. Early adversity, elevated stress physiology, accelerated sexual maturation, and poor health in females. *Developmental psychology*, *51*(6), p.816.

Berger, S.L., Kouzarides, T., Shiekhattar, R. and Shilatifard, A., 2009. An operational definition of epigenetics. *Genes & development*, *23*(7), pp.781-783.

Bird, A.P., 1986. CpG-rich islands and the function of DNA methylation. *Nature*, *321*(6067), pp.209-213.

Blackburn, E.H., Greider, C.W. and Szostak, J.W., 2006. Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nature medicine, 12(10),* pp.1133-1138.

Bocklandt, S., Lin, W., Sehl, M.E., Sánchez, F.J., Sinsheimer, J.S., Horvath, S. and Vilain, E., 2011. Epigenetic predictor of age. *PloS one*, *6*(6), p.e14821.

Bolstad, B.M., Irizarry, R.A., Åstrand, M. and Speed, T.P., 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics, 19*(2), pp.185-193.

Campbell, B., 2006. Adrenarche and the evolution of human life history. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, *18*(5), pp.569-589.

Cawthon, R.M., 2002. Telomere measurement by quantitative PCR. *Nucleic acids research*, 30(10), pp.e47-e47.

Centers for Disease Control and Prevention. (2020). What is epigenetics?. Retrieved from https://www.cdc.gov/genomics/disease/epigenetics.htm

Chaleckis, R., Murakami, I., Takada, J., Kondoh, H. and Yanagida, M., 2016. Individual variability in human blood metabolites identifies age-related differences. *Proceedings of the National Academy of Sciences,* 113(16), pp.4252-4259.

Chrisler, J.C., Barney, A. and Palatino, B., 2016. Ageism can be hazardous to women's health: Ageism, sexism, and stereotypes of older women in the healthcare system. *Journal of Social Issues*, *72*(1), pp.86-104.

Christensen, B.C., Houseman, E.A., Marsit, C.J., Zheng, S., Wrensch, M.R., Wiemels, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R. and Sugarbaker, D.J., 2009. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS genetics, 5(8),* p.e1000602.

Caudill, V.R., Qin, S., Winstead, R., Kaur, J., Tisthammer, K., Pineda, E.G., Solis, C., Cobey, S., Bedford, T., Carja, O. and Eggo, R.M., 2020. CpG-creating mutations are costly in many human viruses. *Evolutionary ecology, 34,* pp.339-359.

Coppé, J.P., Desprez, P.Y., Krtolica, A. and Campisi, J., 2010. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annual review of pathology: mechanisms of disease*, *5*, pp.99-118.

Cursino, M.L., 2021 (unpublished). Are the assumptions eugenics makes about race consistent with what race actually is?. *Submitted as part of the BA(Hons) degree in Philosophy at the University of Manchester,* pp. 1-64.

Dammering, F., Martins, J., Dittrich, K., Czamara, D., Rex-Haffner, M., Overfeld, J., de Punder, K., Buss, C., Entringer, S., Winter, S.M. and Binder, E.B., 2021. The pediatric buccal epigenetic clock identifies significant ageing acceleration in children with internalizing disorder and maltreatment exposure. *Neurobiology of stress*, *15*, p.100394.

Dedeurwaerder, S., Defrance, M., Bizet, M., Calonne, E., Bontempi, G. and Fuks, F., 2014. A comprehensive overview of Infinium HumanMethylation450 data processing. *Briefings in bioinformatics*, *15*(6), pp.929-941.

Dupont, C., Armant, D.R. and Brenner, C.A., 2009, September. Epigenetics: definition, mechanisms and clinical perspective. In *Seminars in reproductive medicine* (Vol. 27, No. 05, pp. 351-357). Thieme Medical Publishers.

Dupras, C., Beck, S., Rothstein, M.A., Berner, A., Saulnier, K.M., Pinkesz, M., Prince, A.E., Liosi, S., Song, L. and Joly, Y., 2019. Potential (mis) use of epigenetic age estimators by private companies and public agencies: human rights law should provide ethical guidance. *Environmental Epigenetics*, *5*(3), p.dvz018.

Essex, R., Riaz, A., Casalotti, S., Worthing, K., Issa, R., Skinner, J. and Yule, A., 2022. A decade of the hostile environment and its impact on health. *Journal of the Royal Society of Medicine*, *115*(3), pp.87-90.

Field, A.E., Robertson, N.A., Wang, T., Havas, A., Ideker, T. and Adams, P.D., 2018. DNA methylation clocks in aging: categories, causes, and consequences. *Molecular cell, 71*(6), pp.882-895.

Föhr, T., Waller, K., Viljanen, A., Sanchez, R., Ollikainen, M., Rantanen, T., Kaprio, J. and Sillanpää, E., 2021. Does the epigenetic clock GrimAge predict mortality independent of genetic influences: an 18 year follow-up study in older female twin pairs. *Clinical epigenetics*, *13*(1), pp.1-9.

Franceschi, C., Bonafè, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E. and De Benedictis, G., 2000. Inflamm-aging: an evolutionary perspective on immunosenescence. *Annals of the new York Academy of Sciences*, *908*(1), pp.244-254.

Fransquet, P.D., Wrigglesworth, J. and Ryan, J., 2023. DNA Methylation Clocks in Age-related Disease. In *Handbook of Epigenetics* (pp. 479-495). Academic Press.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*, *89*(5), pp.1827-1831.

Galanter, J.M., Gignoux, C.R., Oh, S.S., Torgerson, D., Pino-Yanes, M., Thakur, N., Eng, C., Hu, D., Huntsman, S., Farber, H.J. and Avila, P.C., 2017. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *elife, 6*, p.e20532.

Gardner, M., Bann, D., Wiley, L., Cooper, R., Hardy, R., Nitsch, D., Martin-Ruiz, C., Shiels, P., Sayer, A.A., Barbieri, M. and Bekaert, S., 2014. Gender and telomere length: systematic review and meta-analysis. *Experimental gerontology*, *51*, pp.15-27.

Gardner, K. and Shukur, A., 1994. *" I'm Bengali, I'm Asian, and I'm living here": the changing identity of British Bengalis*. Na.

Glantz, S.A., Slinker, B.K. and Neilands, T.B., 2001. *Primer of applied regression & analysis of variance, ed* (Vol. 654). McGraw-Hill, Inc., New York.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C. and Zhang, Y., 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, *466*(7310), pp.1129-1133.

Hannum, G., Guinney, J., Zhao, L., Zhang, L.I., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y. and Deconde, R., 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell*, *49*(2), pp.359-367.

Hayflick, L. and Moorhead, P.S., 1961. The serial cultivation of human diploid cell strains. *Experimental cell research*, *25*(3), pp.585-621.

Héberlé, É. and Bardet, A.F., 2019. Sensitivity of transcription factors to DNA methylation. *Essays in biochemistry, 63*(6), pp.727-741.

Hervouet, E., Vallette, F.M. and Cartron, P.F., 2010. Dnmt1/transcription factor interactions: an alternative mechanism of DNA methylation inheritance. *Genes & cancer*, *1*(5), pp.434-443.

Houghton, L.C., 2013. Juvenility. *Puberty, and Adolescence Among Bangladeshi and British Youth: Durham University*.

Houghton, L.C., Cooper, G.D., Bentley, G.R., Booth, M., Chowdhury, O.A., Troisi, R., Ziegler, R.G., Hoover, R.N., & Katki, H.A. **2014b**. A migrant study of pubertal timing and tempo in British-Bangladeshi girls at varying risk for breast cancer. *Breast Cancer Research*, 16(6), 1-10.

Houghton, L.C., Cooper, G.D., Booth, M., Chowdhury, O.A., Troisi, R., Ziegler, R.G., Katki, H.A., Hoover, R.N. and Bentley, G.R., **2014a**. Childhood environment influences adrenarcheal timing among first-generation Bangladeshi migrant girls to the UK. *PloS one*, *9*(10), p.e109200.

Horvath, S., 2013. DNA methylation age of human tissues and cell types. *Genome biology*, *14*(10), pp.1-20.

Horvath, S., Pirazzini, C., Bacalini, M.G., Gentilini, D., Di Blasio, A.M., Delledonne, M., Mari, D., Arosio, B., Monti, D., Passarino, G. and De Rango, F., **2015b**. Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. *Aging (Albany NY)*, *7*(12), p.1159.

Horvath, S., Gurven, M., Levine, M.E., Trumble, B.C., Kaplan, H., Allayee, H., Ritz, B.R., Chen, B., Lu, A.T., Rickabaugh, T.M. and Jamieson, B.D., 2016. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome biology*, *17*(1), pp.1-23.

Horvath, S. and Levine, A.J., 2015. HIV-1 infection accelerates age according to the epigenetic clock. *The Journal of infectious diseases*, *212*(10), pp.1563-1573.

Horvath, S., Mah, V., Lu, A.T., Woo, J.S., Choi, O.W., Jasinska, A.J., Riancho, J.A., Tung, S., Coles, N.S., Braun, J. and Vinters, H.V., **2015a**. The cerebellum ages slowly according to the epigenetic clock. *Aging (Albany NY)*, *7*(5), p.294.

Horvath, S., Oshima, J., Martin, G.M., Lu, A.T., Quach, A., Cohen, H., Felton, S., Matsuyama, M., Lowe, D., Kabacik, S. and Wilson, J.G., 2018. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany NY)*, *10*(7), p.1758.

Horvath, S. and Raj, K., 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, *19*(6), pp.371-384.

Hunter, R.G. and McEwen, B.S., 2013. Stress and anxiety across the lifespan: structural plasticity and epigenetic regulation. *Epigenomics*, *5*(2), pp.177-194.

Jylhävä, J., Pedersen, N.L. and Hägg, S., 2017. Biological age predictors. *EBioMedicine*, *21*, pp.29-36.

Johnson, A.A., Torosin, N.S., Shokhirev, M.N. and Cuellar, T.L., 2022. A set of common buccal CpGs that predict epigenetic age and associate with lifespan-regulating genes. *Iscience, 25*(11).

Johnson, T.E., 2006. Recent results: biomarkers of aging. *Experimental gerontology*, *41*(12), pp.1243-1246.

Johnson, C.H., Ivanisevic, J. and Siuzdak, G., 2016. Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews Molecular cell biology*, *17*(7), pp.451-459.

Kennedy, B.K., Berger, S.L., Brunet, A., Campisi, J., Cuervo, A.M., Epel, E.S., Franceschi, C., Lithgow, G.J., Morimoto, R.I., Pessin, J.E. and Rando, T.A., 2014. Geroscience: linking aging to chronic disease. *Cell*, *159*(4), pp.709-713.

Kriaucionis, S. and Bird, A., 2003. DNA methylation and Rett syndrome. *Human molecular genetics*, *12*(suppl_2), pp.R221-R227.

Lang, J., McKie, J., Smith, H., McLaughlin, A., Gillberg, C., Shiels, P.G. and Minnis, H., 2020. Adverse childhood experiences, epigenetics and telomere length variation in childhood and beyond: a systematic review of the literature. *European child & adolescent psychiatry*, *29*, pp.1329-1338.

Lapham, K., Kvale, M.N., Lin, J., Connell, S., Croen, L.A., Dispensa, B.P., Fang, L., Hesselson, S., Hoffmann, T.J., Iribarren, C. and Jorgenson, E., 2015. Automated assay of telomere length measurement and informatics for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics*, *200*(4), pp.1061-1072.

Levine, M.E., Lu, A.T., Chen, B.H., Hernandez, D.G., Singleton, A.B., Ferrucci, L., Bandinelli, S., Salfati, E., Manson, J.E., Quach, A. and Kusters, C.D., 2016. Menopause accelerates biological aging. *Proceedings of the National Academy of Sciences*, *113*(33), pp.9327-9332.

Levine, M.E., Lu, A.T., Quach, A., Chen, B.H., Assimes, T.L., Bandinelli, S., Hou, L., Baccarelli, A.A., Stewart, J.D., Li, Y. and Whitsel, E.A., 2018. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (albany NY)*, *10*(4), p.573.

Li, A., Koch, Z. and Ideker, T., 2022. Epigenetic aging: Biological age prediction and informing a mechanistic theory of aging. *Journal of Internal Medicine*, *292*(5), pp.733-744.

Lin, Q., Weidner, C.I., Costa, I.G., Marioni, R.E., Ferreira, M.R., Deary, I.J. and Wagner, W., 2016. DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. *Aging (Albany NY)*, *8*(2), p.394.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M. and Edsall, L., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *nature*, *462*(7271), pp.315-322.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M. and Kroemer, G., 2013. The hallmarks of aging. *Cell*, *153*(6), pp.1194-1217.

Lowe, D., Horvath, S. and Raj, K., 2016. Epigenetic clock analyses of cellular senescence and ageing. *Oncotarget*, *7*(8), p.8524.

Lu, A.T., Quach, A., Wilson, J.G., Reiner, A.P., Aviv, A., Raj, K., Hou, L., Baccarelli, A.A., Li, Y., Stewart, J.D. and Whitsel, E.A., 2019. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (albany NY)*, *11*(2), p.303.

Maddar, H., Ratzkovsky, N. and Krauskopf, A., 2001. Role for telomere cap structure in meiosis. *Molecular biology of the cell*, *12*(10), pp.3191-3203.

Mc Auley, M.T., 2023. An evolutionary perspective of lifespan and epigenetic inheritance. *Experimental Gerontology*, *179*, p.112256.

McEwen, L.M., O'Donnell, K.J., McGill, M.G., Edgar, R.D., Jones, M.J., MacIsaac, J.L., Lin, D.T.S., Ramadori, K., Morin, A., Gladish, N. and Garg, E., 2020. The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proceedings of the National Academy of Sciences*, *117*(38), pp.23329-23335.

McGill, M.G., Pokhvisneva, I., Clappison, A.S., McEwen, L.M., Beijers, R., Tollenaar, M.S., Pham, H., Kee, M.Z., Garg, E., de Mendonça Filho, E.J. and Karnani, N., 2022. Maternal prenatal anxiety and the fetal origins of epigenetic aging. *Biological psychiatry*, *91*(3), pp.303-312.

Moore, L.D., Le, T. and Fan, G., 2013. DNA methylation and its basic function. *Neuropsychopharmacology, 38(1)*, pp.23-38.

Mourtzi, N., Sertedaki, A. and Charmandari, E., 2021. Glucocorticoid signaling and epigenetic alterations in stress-related disorders. *International journal of molecular sciences*, *22*(11), p.5964.

Murphy, S.A., Bentley, G.R. and O'Hanesian, M.A., 1995. An analysis for menstrual data with time-varying covariates. *Statistics in Medicine*, *14*(17), pp.1843-1857.

Newman, A.B., 2015. Is the onset of obesity the same as aging?. *Proceedings of the National Academy of Sciences*, *112*(52), pp.E7163-E7163.

Noroozi, R., Ghafouri-Fard, S., Pisarek, A., Rudnicka, J., Spolnicka, M., Branicki, W., Taheri, M. and Pośpiech, E., 2021. DNA methylation-based age clocks: From age prediction to age reversion. *Ageing Research Reviews*, *68*, p.101314.

Okano, M., Bell, D.W., Haber, D.A. and Li, E., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), pp.247-257.

Peixoto, P., Cartron, P.F., Serandour, A.A. and Hervouet, E., 2020. From 1957 to nowadays: a brief history of epigenetics. *International journal of molecular sciences*, *21*(20), p.7571.

Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K. and Wilson, Y.A., 2015. The transcriptional landscape of age in human peripheral blood. *Nature communications*, *6*(1), pp.1-14.

Perroud, N., Rutembesa, E., Paoloni-Giacobino, A., et al. (2011). The Tutsi genocide and transgenerational transmission of maternal stress: epigenetics and biology of the HPA axis. *World Journal of Biological Psychiatry, 12(4),* 342-352.

Pipek, O.A. and Csabai, I., 2023. A revised multi-tissue, multi-platform epigenetic clock model for methylation array data. *Journal of Mathematical Chemistry, 61*(2), pp.376-388.

Quach, A., Levine, M.E., Tanaka, T., Lu, A.T., Chen, B.H., Ferrucci, L., Ritz, B., Bandinelli, S., Neuhouser, M.L., Beasley, J.M. and Snetselaar, L., 2017. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (albany NY)*, *9*(2), p.419.

Raghavan, S., 2013. *(1971) A Global History of the Creation of Bangladesh*. Harvard University Press.

Rakyan, V.K., Down, T.A., Maslau, S., Andrew, T., Yang, T.P., Beyan, H., Whittaker, P., McCann, O.T., Finer, S., Valdes, A.M. and Leslie, R.D., 2010. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome research*, *20*(4), pp.434-439.

Resendiz, M., Watkins, D.S., Öztürk, N.C. and Zhou, F.C., 2023. Environmental Influence on Epigenetics. *Handbook of Epigenetics,* pp.639-668.

Ryan, C.P., 2021. "Epigenetic clocks": Theory and applications in human biology. *American Journal of Human Biology*, *33*(3), p.e23488.

Saldaña-Tejeda, A., 2018. Mitochondrial mothers of a fat nation: Race, gender and epigenetics in obesity research on Mexican mestizos. *BioSocieties*, *13*(2), pp.434-452.

Saulnier, K.M. and Dupras, C., 2017. Race in the postgenomic era: Social epigenetics calling for interdisciplinary ethical safeguards. *The American Journal of Bioethics*, *17*(9), pp.58-60.

Scarano, E., Iaccarino, M., Grippo, P. and Winckelmans, D., 1965. On methylation of DNA during development of the sea urchin embryo. *Journal of molecular biology*, *14*(2), pp.603-607.

Sehl, M.E., Henry, J.E., Storniolo, A.M., Ganz, P.A. and Horvath, S., 2017. DNA methylation age is elevated in breast tissue of healthy women. *Breast cancer research and treatment*, *164*, pp.209-219.

Singh, G., Morrison, J. and Hoy, W., 2019. DOHaD in Indigenous populations: DOHaD, epigenetics, equity and race. *Journal of Developmental Origins of Health and Disease*, *10*(1), pp.63-64.

Smith, S.M. and Vale, W.W., 2022. The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress. *Dialogues in clinical neuroscience*.

Song, M.A., Brasky, T.M., Marian, C., Weng, D.Y., Taslim, C., Dumitrescu, R.G., Llanos, A.A., Freudenheim, J.L. and Shields, P.G., 2015. Racial differences in genome-wide methylation profiling and gene expression in breast tissues from healthy women. *Epigenetics*, *10*(12), pp.1177-1187.

Stöger, R., Choi, M., Begum, K., Leeman, G., Emes, R.D., Melamed, P. and Bentley, G.R., 2023. Childhood environment influences epigenetic age and methylation concordance of a CpG clock locus in British-Bangladeshi migrants. *Epigenetics*, *18*(1), p.2153511.

Suor, J.H., Sturge-Apple, M.L., Davies, P.T., Cicchetti, D. and Manning, L.G., 2015. Tracing differential pathways of risk: Associations among family adversity, cortisol, and cognitive functioning in childhood. *Child development*, *86*(4), pp.1142-1158.

Tanaka, T., Biancotto, A., Moaddel, R., Moore, A.Z., Gonzalez-Freire, M., Aon, M.A., Candia, J., Zhang, P., Cheung, F., Fantoni, G. and CHI consortium, 2018. Plasma proteomic signature of age in healthy humans. *Aging cell*, *17*(5), p.e12799.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), pp.267-288.

Turecki, G. and Meaney, M.J., 2016. Effects of the social environment and stress on glucocorticoid receptor gene methylation: a systematic review. *Biological psychiatry*, *79*(2), pp.87-96.

UK Home Office. Skilled Worker Visa webpage. *Retrieved in August 2023 from: https://www.gov.uk/skilled-worker-visa.*

Van der Kolk, B., 2014. The body keeps the score: Brain, mind, and body in the healing of trauma. *New York*, *3*.

Vick, A.D. and Burris, H.H., 2017. Epigenetics and health disparities. *Current epidemiology reports*, *4*, pp.31-37.

Waddington, C.H., 2012. The epigenotype. *International journal of epidemiology*, *41*(1), pp.10-13.

Ward-Caviness, C.K., Nwanaji-Enwerem, J.C., Wolf, K., Wahl, S., Colicino, E., Trevisi, L., Kloog, I., Just, A.C., Vokonas, P., Cyrys, J. and Gieger, C., 2016. Long-term exposure to air pollution is associated with biological aging. *Oncotarget*, *7*(46), p.74510.

Weidner, C.I., Lin, Q., Koch, C.M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D.O., Jöckel, K.H., Erbel, R., Mühleisen, T.W. and Zenke, M., 2014. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome biology*, *15*, pp.1-12.

Yehuda, R., Daskalakis, N.P., Bierer, L.M., Bader, H.N., Klengel, T., Holsboer, F. and Binder, E.B., 2016. Holocaust exposure induced intergenerational effects on FKBP5 methylation. *Biological psychiatry*, *80*(5), pp.372-380.

Ying, Z. and Chen, T., 2023. Mechanisms of DNA Methylation and Demethylation During Mammalian Development. In *Handbook of Epigenetics* (pp. 11-26). Academic Press.

Zheng, S.C., Widschwendter, M. and Teschendorff, A.E., 2016. Epigenetic drift, epigenetic clocks and cancer risk. *Epigenomics*, *8*(5), pp.705-719.

Zhou, X., Sen, I., Lin, X.X. and Riedel, C.G., 2018. Regulation of age-related decline by transcription factors and their crosstalk with the epigenome. *Current genomics, 19*(6), pp.464-482.

Appendix

Appendix A. ANOVA statistical results

DNA methylation-based epigenetic clock	**P-value from Brown-Forsythe ANOVA test**	**P-value from Welch ANOVA test**
Horvath's multi-tissue clock normalised data entry	0.0018	0.0015
Horvath's multi-tissue clock raw data entry	0.7039	0.5847
Horvath's multi-tissue clock	0.5474	0.5479
PedBE clock	0.6547	0.6500
Skin and blood clock	0.1893	0.1692
PhenoAge clock	0.0437	0.0348
LinAge clock	0.2564	0.2560

Table A. ANOVA statistical results for 6 epigenetic clocks, with three entries on Horvath's multi-tissue clock. The table shows the P-values from the Brown-Forsythe ANOVA test and the P-value from the Welch ANOVA test.

Appendix B. Composite graphs that indicate the gradient of age acceleration in the white British cohort across 8 distinct sets of results. 3 from Horvath's clock and the other 5 from distinct clocks.

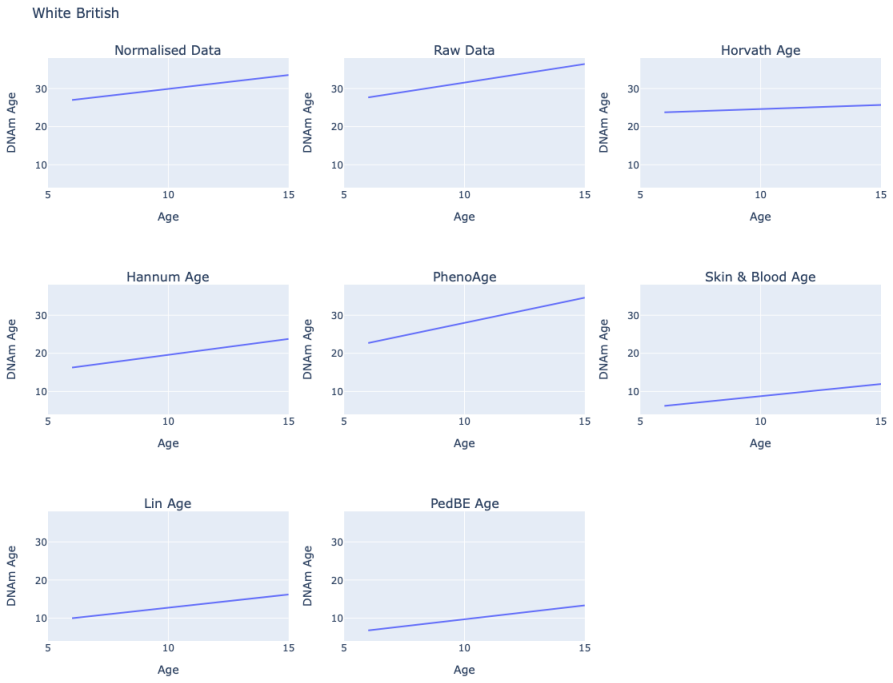


Figure B. Linear regression of DNA methylation-based age estimates plotted against the white British population's chronological ages. The data was obtained from different epigenetic clocks to illustrate how well calibrated the clocks are to assess DNA methylation levels at age informing CpG sites from paediatric buccal samples. From top left across (A) normalised data results from Horvath's clock (B) raw data results from Horvath's clock (C) data results from Horvath's clock (D) data results from Hannum's clock (E) data results from PhenoAge clock (F) data results from Skin and Blood clock (G) Data results from LinAge clock (H) data results from PedBE clock. As illustrated above, the PedBE clock was better calibrated than other clocks to accurately predict the epigenetic age estimates from this study's cohort.

Appendix C. Composite graphs that indicate the gradient of age acceleration in the Bangladeshi cohort across 8 distinct sets of results. 3 from Horvath's clock and the other 5 from distinct clocks.

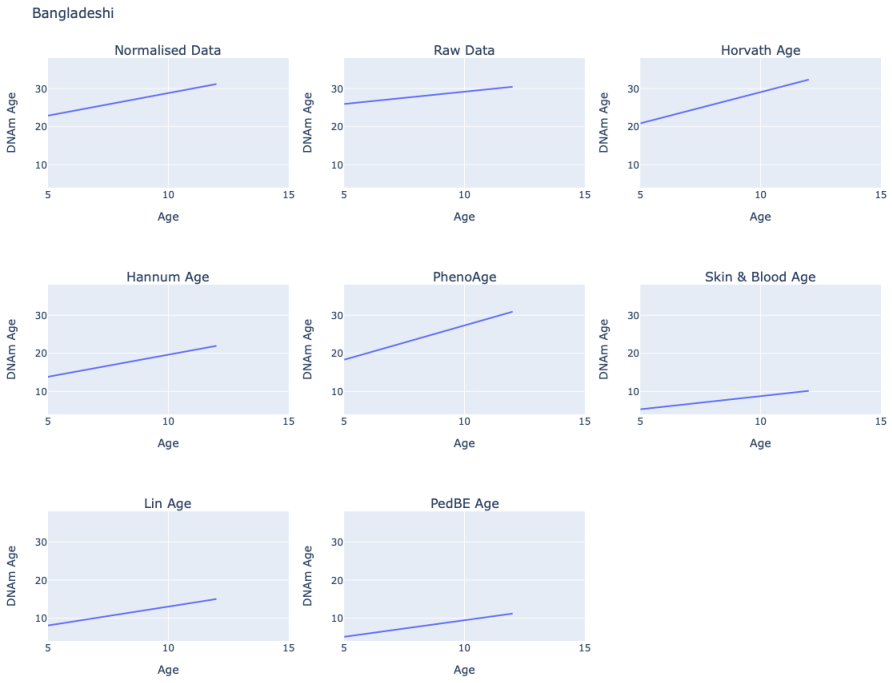


Figure C. Linear regression of DNA methylation-based age estimates plotted against the Bangladeshi population's chronological ages. The data was obtained from different epigenetic clocks to illustrate how well calibrated the clocks are to assess DNA methylation levels at age informing CpG sites from paediatric buccal samples. From top left across (A) normalised data results from Horvath's clock (B) raw data results from Horvath's clock (C) data results from Horvath's clock (D) data results from Hannum's clock (E) data results from PhenoAge clock (F) data results from Skin and Blood clock (G) Data results from LinAge clock (H) data results from PedBE clock. As illustrated above, the PedBE clock was better calibrated than other clocks to accurately predict the epigenetic age estimates from this study's cohort.

Appendix D. Composite graphs that indicate the gradient of age acceleration in the second-generation cohort across 8 distinct sets of results. 3 from Horvath's clock and the other 5 from distinct clocks.

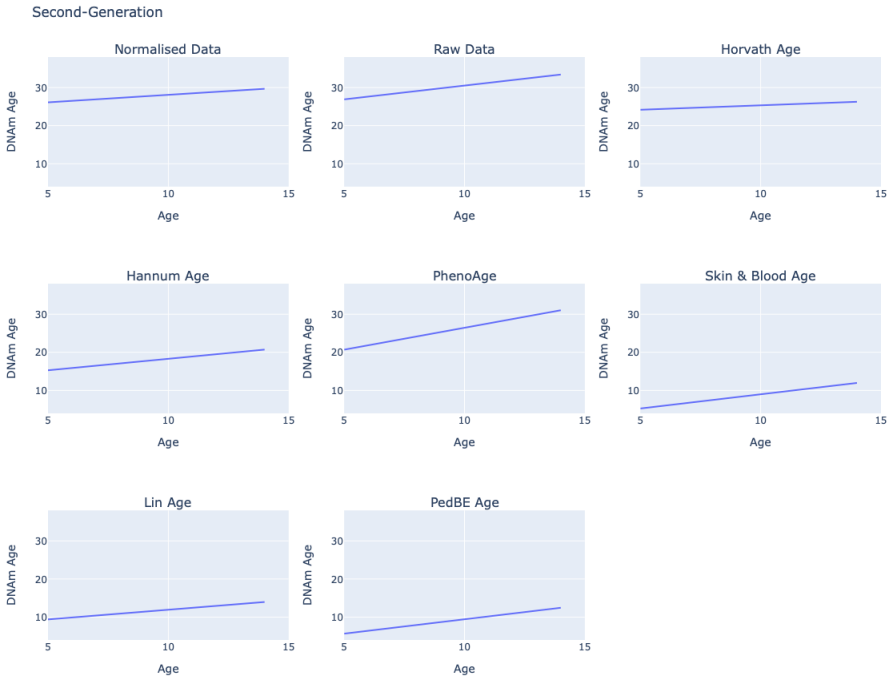


Figure D. Linear regression of DNA methylation-based age estimates plotted against the second-generation population's chronological ages. The data was obtained from different epigenetic clocks to illustrate how well calibrated the clocks are to assess DNA methylation levels at age informing CpG sites from paediatric buccal samples. From top left across (A) normalised data results from Horvath's clock (B) raw data results from Horvath's clock (C) data results from Horvath's clock (D) data results from Hannum's clock (E) data results from PhenoAge clock (F) data results from Skin and Blood clock (G) Data results from LinAge clock (H) data results from PedBE clock. As illustrated above, the PedBE clock was better calibrated than other clocks to accurately predict the epigenetic age estimates from this study's cohort.