How does allostatic load relate to cognitive ability, depressive symptoms and other measures of physiological weathering in older age?

Zander Crook

PhD Psychology
The University of Edinburgh
2019

Abstract

Allostasis is the process through which the human body aims to achieve homeostasis - long-term physiological stability in each of its systems - through short-term changes. If someone experiences chronic stress or repeated bouts of stress, their allostatic systems can experience wear and tear and enter a state of allostatic load.

The thesis first considered how allostatic load is operationalised. It outlined problems with modelling allostatic load as a latent factor, including concerns based on psychometric and substantive studies. This chapter also posited desirable qualities of allostatic load operationalisations.

The subsequent substantive chapters were empirical studies that explored the role of allostatic load in successful ageing. In the first three of four studies, we fitted latent growth curve models to data from Waves 1-4 of the Lothian Birth Cohort 1936 (*n* [*M* age] at Waves 1-4 = 1,091 [69.5 years], 866 [72.5 years], 697 [76.3 years], 550 [79.3 years]). Additional models tested the role of non-random dropout.

The first of these studies investigated how allostatic load relates to four other measures of physiological weathering: [Klemera-Doubal] biological age; extrinsic epigenetic age; intrinsic epigenetic age; and telomere length. Allostatic load was most strongly related to biological age, which is an age-linked measure calculated using some of the same biomarkers as allostatic load. At Wave 1, higher allostatic load was related to older biological age, with a moderate effect size (*r* [*SE*] =

.28 [.05], $p < .001$). Greater increase over time in allostatic load was strongly related to accelerated biological ageing (r [SE] = .53 [.15], $p < .001$). However, most correlations between physiological weathering measures were small, which suggested that they index different aspects of physiological weathering.

The next study found that at Wave 1, higher allostatic load was related to lower general cognitive ability, with a small effect size (r [SE] = −.13 [.04], $p = .004$). However, after controlling for the variance related to participant dropout, the relationship was attenuated by around a third (r [SE] = −.08 [.04], $p = .08$). The third Lothian Birth Cohort 1936 study found that at Wave 1, higher allostatic load was related to having more depressive symptoms, with a small-to-moderate effect size (r [SE] = .19 [.05], $p < .001$). These studies also went beyond previous research by testing relationships of associated changes over time. These tests suggested changes in allostatic load are not strongly related to changes in cognitive ability and depressive symptoms.

The last empirical study in this thesis used data from the Scotland-based 36-Day-Sample (n range for models = 280-332; M age at cognitive testing = 78.2 years). In this study, producing more cortisol during the first 45 minutes after waking was related to higher general cognitive ability (β [SE] = 0.14 [0.07], *Bayes factor*$_{10}$ [BF_{10}] = 5.23), but the total cortisol produced during the waking day was not related to general cognitive ability ($BF_{10} < 1/3$). We also found that in carriers of the e4 allele of the apolipoprotein E gene, some measures of cortisol dynamics were more negatively associated with residualised immediate recall scores and some measures of total cortisol production were more negatively associated with residualised Raven's matrices scores. Additionally, this study found that depressive symptoms were related to lower general cognitive ability (β [SE] = −0.15 [0.08], BF_{10} = 3.47), but the relationship was not more negative in e4 allele carriers (β [SE]

= 0.08 [0.07], BF_{10} = 0.21).

The final chapter considered the results in a wider context, as well as possible future directions. We discussed the limitations of our allostatic load measure, which did not include any hypothalamic–pituitary–adrenal axis biomarkers. We also considered the constraints on generalisability of all allostatic load findings given the heterogeneity of the operationalisations used. We questioned the clinical utility of allostatic load measurement for cognitive decline and depressive illness, given the small effect sizes we obtained for the relevant relationships. Finally, we concluded that future research will benefit from using larger samples and teasing apart how these relationships vary based on the allostatic load biomarkers and operationalisations used.

Lay summary

If you experience too much stress, your body's stress systems will work too hard and become damaged. This includes psychological stress, for example, work stress or anxieties about a family member's health, and physiological stress, such as running a marathon without proper training. Lack of activity, for example, lying stationary in bed for a month, can also negatively affect the body. Negative impacts on your stress systems then have knock-on effects on your other bodily systems. This includes your cardiovascular system, which works to keep your heart healthy and your body fit, your metabolic system, which processes the food and drink you consume, and your immune system, which fights viruses and works to heal damaged muscles and tissue. The resulting stress-induced state, in which various bodily systems are not working well, is called allostatic load. It can be measured by measuring and statistically combining markers from different bodily systems, such as blood pressure for the cardiovascular system, body mass index for the metabolic system, and C-reactive protein levels in blood for the immune system. This thesis looked at allostatic load, how it relates to cognitive ability, depressive symptoms and other measures of poor bodily functioning.

We found that in the Lothian Birth Cohort 1936, a sample of 1,091 older adults, changes over time in allostatic load were closely related to changes over time in biological age, which is a chronological age-linked measure that is calculated using some of the same biomarkers as

allostatic load. In the same sample, we found allostatic load was related to general cognitive ability and depressive symptoms at around age 70. We did not find that changes in allostatic load were strongly related to changes in general cognitive ability or depressive symptoms. Finally, in the 36-Day-Sample, a sample of over 300 older adults, we tested how cognitive ability is associated with depressive symptoms, neuroticism, and the levels and within-day changes of cortisol, a stress-related allostatic load biomarker. We also tested whether these relationships were more harmful in participants who had a particular genetic variant, the e4 allele of the apolipoprotein E gene. We found that participants who produced more cortisol in the first 45 minutes after waking tended to have higher general cognitive ability. Participants with more depressive symptoms tended to have lower general cognitive ability, and there was evidence that this association was not stronger in those with the e4 allele. Furthermore, our results suggested that there was not an association between the total cortisol produced while awake and general cognitive ability.

It must be emphasised that while we found associations between the functioning of the body and the mind, these associations were small-to-moderate in size. Knowing an older individual's allostatic load score would help you to guess their cognitive ability or how many depressive symptoms they experience (or vice versa), but only slightly.

Declarations

I, Zander Crook, certify that I composed this thesis.

Where I used other work in composing this thesis, I have referenced this in text and in the bibliography.

For the studies in this thesis, I led the study design, conducted the analyses and led the paper writing. My primary supervisor, Tom Booth, assisted with the writing of the chapters and with the design of all studies included. Other collaborators also made contributions to some of the work presented in this thesis. Ian J. Deary provided comments and edits on the background and preregistration documents for Chapters 4, 5, and 6. These documents formed the basis of the introduction and methods sections for these chapters. He also provided comments and edits on Chapter 7. Riccardo E. Marioni contributed to the study design of Chapter 4 and commented on a draft of this chapter. Mathew A. Harris contributed to the study design of Chapter 7 as well as to the preparation of the journal manuscript that formed the basis of this chapter. Further co-authors also helped with comments and suggested edits to that manuscript, namely Simon R. Cox, Catharine R. Gale, Caroline E. Brett, and John M. Starr.

This thesis is submitted for the PhD Psychology programme at the University of Edinburgh and has not been submitted for any other degree or professional qualification.

The published version of Chapter 2 is included as an Appendix. It is my own work and was written in collaboration with my primary supervisor, Tom Booth.

Signed,

Crook

Zander Crook

4th August 2019

Funding

This thesis was funded by an Advanced Quantitative Methods PhD award from the Economic and Social Research Council (grant number C106891X).

Chapters 4, 5 and 6 use data from the Lothian Birth Cohort 1936 study, which has received funding from Research into Ageing (programme grant 251; Wave 1), Age UK (Disconnected Mind programme grant), and the Medical Research Council (grant numbers for Waves 2-4: G0701120, G1001245, and MR/M013111/1).

Chapter 4 uses DNA methylation data from the Lothian Birth Cohort 1936, which was collected with funding from Wellcome, Age UK, and the Universities of Edinburgh and Queensland.

Chapter 7 uses data from the 36-Day-Sample follow-up study, which has received funding from the cross-council Lifelong Health and Wellbeing Initiative (grant number MRC G1001401/1).

Work on the Lothian Birth Cohort 1936 and 36-Day-Sample studies was undertaken in the Cognitive Ageing and Cognitive Epidemiology, which is funded by the Medical Research Council and the Biotechnology and Biological Sciences Research Council (grant number MR/K026992/1).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The allostatic load model

The healthy human body is capable of dealing with a wide range of mental and physical challenges. It is like a machine in that regular use helps to keep its components in good condition. For example, if a person takes part in regular physical exercise, they will generally become more fit and more able to meet certain physical challenges. However, like with many machines, very rare use, very frequent use, or extreme use can cause damage. For example, in the case of post-traumatic stress disorder, extreme psychological stress can lead to adverse mental (Koenen et al., 2003) and physiological (Pacella, Hruska, & Delahanty, 2013) health outcomes. The allostatic load (AL) model provides a framework for understanding and studying the wide-ranging effects of stress on physiological and psychological health. This model, and key relationships with other physiological and psychological measures to be studied in this thesis, will now be described.

Homeostatic systems (e.g., body temperature) have to be kept within small ranges, with severe health consequences if they are not (McEwen, 1998). Allostatic systems (e.g., the hypothalamic-pituitary-adrenal

[HPA] axis), on the other hand, can function within wider ranges, with dysregulation typically leading to damage being incrementally inflicted on the body (McEwen, 1998).

Allostasis is the ability of the body's physiological systems to maintain long-term stability by successfully adapting in the short term to situations and events (McEwen & Wingfield, 2007). Repeated, prolonged or insufficient exposure to stress can lead to allostatic systems becoming dysregulated (McEwen & Wingfield, 2007). Allostatic load, or allostatic overload, is the the accumulation of damage to the body's physiological systems that is caused by allostatic states, that is, the long-term overactivity or underactivity of allostatic systems (McEwen & Wingfield, 2007).

McEwen (1998) outlined four origins of allostatic load:

- repeated stress responses, where the healthy stress response occurs too often;
- a lack of adaptation to stress, where the body does not habituate to a particular stressor and so it repeats the initial heightened response to stress during subsequent exposures;
- prolonged response to stress, where the heightened stress response continues for an extended period;
- and an inadequate response to stress, where a fully heightened stress response is not achieved by one allostatic system, so another allostatic system overresponds to compensate.

McEwen & Seeman (1999) added to the allostatic load model by presenting the following stages of allostasis and allostatic load, each of which plays a different role in the cascade of reactions and effects involved in allostasis:

- primary mediators (e.g., cortisol), chemical messengers, such as catecholamines, cytokines, and hormones, that are released as part of allostasis (McEwen, 2005);

- primary effects (e.g., enzymes), cellular events that primary mediators affect as part of allostasis;
- secondary outcomes (e.g., triglycerides), organ- or tissue-specific outcomes of the long-term levels of (and fluctuations in levels of) the primary effects, which were themselves brought about by the primary mediators;
- and tertiary outcomes (e.g., cardiovascular disease), diseases or disorders caused by allostatic load, which itself was brought about by high and/or low values of the primary mediators and secondary outcomes.

An example cascade starts with psychological stress leading to increased secretion of the primary mediator cortisol (Chida & Steptoe, 2009). The higher level of cortisol could relate to the primary effect of DNA methylation (Lam et al., 2012). This increased cortisol would then lead to increased levels of the secondary outcome glucose, directly and/or through effects on diet (Epel, Lapidus, McEwen, & Brownell, 2001; McEwen & Seeman, 1999). This cascade would be part of a broader move towards a state of allostatic load, which could ultimately lead towards the tertiary outcome of Type 2 diabetes (Rosmond, 2003).

Allostatic load, like in the example cascade, involves dysregulation in multiple physiological systems. Primary mediators are mainly markers from the neuroendocrine system (e.g., glucocorticoids released by the HPA axis) and the immune/inflammatory system (e.g., interleukin-6). The most commonly measured secondary outcomes are markers from the metabolic (e.g., total cholesterol) and cardiovascular systems (e.g., blood pressure). Summary measures of allostatic load generally include markers of different types from different physiological systems.

Studies using such summary measures have shown the importance of studying allostatic load, as they have linked allostatic load scores to important health outcomes, including mortality. For example, in the MacArthur Study of Successful Aging, higher allostatic load was linked to greater all-cause mortality risk (odds ratio = 1.23; Seeman, McEwen, Rowe, & Singer, 2001). This association was not only due to commonly administered metabolic system markers; in fact, neuroendocrine summary scores seemed to be more strongly predictive of mortality than metabolic syndrome summary scores (odds ratios = 1.36 and 1.18; Seeman et al., 2001). A similar effect size was observed in the Taiwan-based Social Environment and Biomarkers of Aging study (sample aged 54 and over at baseline; *n* = 1,023), where an additional allostatic biomarker in the top or bottom sample decile was associated with a 20% increase in 10-year all-cause mortality risk (hazard ratio = 1.20; Hwang et al., 2014). The effect size was smaller in the Scottish Health Survey (at baseline *mean* [*standard devation*] age = 51.0 [18.0]; *n* = 4,488), where an additional biomarker indicating risk increased 10-year all-cause mortality risk by 8% (Robertson, Beveridge, & Bromley, 2017). This study used a younger sample, did not have neuroendocrine markers and scored allostatic load with the number of biomarkers in the highest risk sample quartile, rather than the top and bottom sample deciles (Robertson et al., 2017). Finally, allostatic load was also related to mortality in the large sample of the National Health and Nutrition Examination Survey III (sample aged 30 and over; *n* = 9,942). For the most predictive allostatic load measure, those in the highest quintile of allostatic load had almost seven times the mortality risk of those in the lowest quintile (hazard ratio for continuous allostatic load = 6.97; Levine & Crimmins, 2014). Notably, though, the strength of the association differed based on the allostatic load operationalisation used (Levine & Crimmins, 2014). Furthermore, another mortality study used data from the same cohort but operationalised allostatic load using a different biomarker set with no neuroendocrine markers (Howard & Sparks, 2016). This study found that an additional biomarker beyond a clinical cutoff increased all-cause mortality risk by 7%, with a follow-up period of up to 18 years (hazard ratio = 1.07; sample aged 25 and over; *n* = 11,733; Howard & Sparks, 2016).

1.2 Allostatic load operationalisation

When conducting allostatic load research, it is important to carefully consider the operationalisation used as this can affect the strength and patterns of results. There is notable heterogeneity in the biomarkers and the statistical methods used to calculate allostatic load scores. The biomarker sets used generally contain markers from multiple physiological systems, such as the neuroendocrine, metabolic, immune and cardiovascular systems, but operationalisations vary in the systems represented and the specific biomarkers used (Juster, McEwen, & Lupien, 2010, Figure 4).

Scores are also affected by the calculation method used. There are many methods in use. A recent study comparing allostatic load operationalisations in the same dataset found that in 21 studies, scores were calculated in 18 ways (Duong, Bingham, Aldana, Chung, & Sumner, 2017). Methods used to calculate allostatic load scores include: sums of markers dichotomised into 0s (not indicating risk) and 1s (indicating risk) based on sample quantiles (Seeman, Singer, Rowe, Horwitz, & McEwen, 1997) or clinical thresholds (Seeman et al., 2008); sums of *z* score-transformed markers (Seeman et al., 1997; Seplaki, Goldman, Glei, & Weinstein, 2005); and correlation-based measures, such as factor analysis (Seeman et al., 2010), principal component analysis (Wallace et al., 2013), and latent class analysis (Fried et al., 2009).

Recently, some researchers have argued for the use of a bifactor confirmatory factor analysis model in which allostatic load biomarker levels are causally influenced by a general latent factor of allostatic load and system-specific latent factors (Wiley, Gruenewald, Karlamangla, & Seeman, 2016). However, there are practical and theoretical problems with this allostatic load operationalisation. In Chapter 2, we will consider this operationalisation and what properties are desirable for allostatic load operationalisations. Then, we will outline the allostatic load operationalisation we chose to use in Chapters 4, 5 and 6, with reference to its measurement properties and previous method comparisons.

1.3 Allostatic load and other physiological weathering measures

Allostatic load theory posits particular mechanisms underlying observed allostatic load. However, summary measures of allostatic load are sometimes used more as general measures of overall physiological dysregulation or physiological weathering (e.g., Goldman et al., 2006). There are other broad measures of physiological weathering with different theoretical and methodological foundations, although these have not frequently been studied together with allostatic load. Three such measures are closely linked to the study of biological ageing: biomarker panel-based measures of biological age; DNA methylation-based measures of epigenetic age; and leukocyte telomere length.

Attempts to measure biological age have been made for decades (see Jia, Zhang, & Chen, 2017 for a review). The Klemera-Doubal method (Klemera & Doubal, 2006) has performed well in studies comparing calculation methods (Cho, Park, & Lim, 2010; Jee & Park, 2017). This method uses chronological age along with ageing biomarkers, some of which have also been used in allostatic load measurement, to estimate biological age (Klemera & Doubal, 2006). Estimates have been linked to a variety of mental and physical health outcomes, including mortality (Belsky et al., 2015; Levine, 2013).

Estimates of epigenetic age use DNA methylation rather than any of the biomarkers used to operationalise allostatic load. DNA methylation occurs when a methyl group is added to cytosine nucleotides on the genome at cytosine-phosphate-guanine (CpG) sites (Beck & Rakyan, 2008). The proportion of methylation at CpG sites can be measured (Beck & Rakyan, 2008). Recently, through working with models that predict chronological age from methylation at particular CpG sites, researchers have developed methods to estimate epigenetic age from DNA methylation in blood and/or other tissues at a selection of CpG sites (Hannum et al., 2013; Horvath, 2013). Epigenetic age estimates from these methods are strongly related to chronological age and, when combined with chronological age, can be used to generate estimates of epigenetic age acceleration/deceleration, the pace of the epigenetic clock (Hannum et al., 2013; Horvath, 2013). Like biological age, epigenetic age has been linked to mortality (Marioni et al., 2016).

Lastly, telomere length has been put forward as a biomarker for ageing (Mather, Jorm, Parslow, & Christensen, 2011). Telomeres are repetitive DNA sequences that sit at the end of each chromosome, protecting them from damage (Riethman, 2008). Telomeres are shortened each time a cell replicates DNA and by oxidative stress, until they are shorter than a critical length, at which point DNA replication stops and the cell enters a state of senescence (Simons, 2015). Telomeres shorten with age and telomere length has sometimes been linked to age-related measures, so it is a contested marker of biological age (Mather et al., 2011). It has been incorporated in biological age estimation (Zhang et al., 2014), although it may rather be “a biomarker of somatic redundancy, the body's capacity to absorb damage" (Boonekamp, Simons, Hemerik, & Verhulst, 2013, p. 330). Telomere length has also been linked to mortality (Boonekamp et al., 2013) and its relationship with mortality is independent of that of epigenetic age (Marioni et al., 2016).

Researchers have begun to explore the relationships between these different measures. Allostatic load was not related to telomere length in a sample of reproductive-aged women (Ahrens, Rossen, & Simon, 2016), but three allostatic load measures were strongly correlated with biological age in the National Health and Nutrition Examination Survey III, a large representative sample in the United States (*n* = 9,942; *r* range = .49 to .65; Levine & Crimmins, 2014). At age 38 in the New Zealand-representative Dunedin Study cohort, biological age was correlated weakly with epigenetic age (*r* range = .07 to .15), but telomere length was not statistically significantly correlated with biological age or epigenetic age (*r* range = -.02 to -.05; Belsky et al., 2018). Chapter 4 will attempt to clarify the relationships between these measures, including hitherto unexplored longitudinal relationships.

1.4 Allostatic load and cognitive ability

It is important to research the contemporaneous and longitudinal relationships between allostatic load and cognitive ability, particularly in older age. Cognitive ageing is a major public health issue, especially in countries with ageing populations (Deary et al., 2009) and researchers have suggested that allostatic load could improve the assessment of risk for age-associated cognitive decline (Seeman et al., 1997).

The brain is central to both cognitive functioning and the stress response involved in allostasis, so effects on the brain can alter both the stress response and cognitive ability (McEwen & Gianaros, 2011).

There are also potential causal pathways in both directions between the two measures. For example, the allostatic load marker cortisol may decrease brain area volumes, which in turn can affect cognitive functioning (Echouffo-Tcheugui et al., 2018). Cognitive ability is related to unhealthy behaviours (Gottfredson & Deary, 2004), which have been associated with allostatic load (Forrester, Leoutsakos, Gallo, Thorpe, & Seeman, 2019). Cohort studies, such as those discussed here and conducted in this thesis, may inform future mechanistic research but their focus is on estimating associations.

Allostatic load has previously been related to contemporaneously measured cognitive ability (Booth et al., 2015; Crook et al., 2018; Karlamangla et al., 2014; Seplaki et al., 2005). Plus, in the MacArthur Studies of Successful Aging (based in the United States of America) and the Lothian Birth Cohort 1936 (based in Scotland), but not the Social Environment and Biomarkers of Aging Study (based in Taiwan), allostatic load has been linked to subsequent cognitive decline (Crook et al., 2018; Karlamangla, Singer, McEwen, Rowe, & Seeman, 2002; Seeman et al., 2001, 1997; cf. Goldman et al., 2006). A longitudinal link in the other direction has also been reported: in the West of Scotland Twenty-07 Study, slower processing speed at age 16 was related to higher allostatic load at age 36 (Gale et al., 2015).

Aspects of the relationship between allostatic load and cognitive ability remain unclear, including whether each variable predicts subsequent change in the other when longitudinal measures of both are modelled together, whether longitudinal change in both variables is correlated, and how allostatic load relates over time to general and specific domains of cognitive ability. Chapter 5 will attempt to clarify the longitudinal relationship between allostatic load and cognitive ability.

1.5 Allostatic load and depressive symptoms

Understanding how allostatic load relates to depression is important, as it could aid in the diagnosis and treatment of depression (Lopresti, Maker, Hood, & Drummond, 2014) and physiological dysregulation, enabling more personalised and effective treatment. It is well established that certain allostatic load biomarkers tend to be dysregulated in people with depression. This includes endocrine (Murri et al., 2014), inflammatory (Valkanova, Ebmeier, & Allan, 2013), and metabolic (Marazziti, Rutigliano, Baroni, Landi, & Dell'Osso, 2014) biomarkers. Physiological dysregulation in people with depression is not uniform, as profiles of dysregulation differ between subtypes of depression (Lamers et al., 2013).

As mentioned previously, the brain is the central mediator of the stress response (McEwen & Gianaros, 2011). It is also central to depressive symptoms. Hence, the brain can be a key mediator of the relationship between allostatic load and depressive symptoms. For example, the hippocampus is involved in downregulation of the stress response and is reduced in size in depression as well as in states of chronic stress, chronic inflammation, and after prolonged glucocorticoid exposure (McEwen & Rasgon, 2018). Observational cohort studies sometimes help to suggest future mechanistic studies of allostatic load and depressive symptoms, but their focus is on estimating the contemporaneous and longitudinal associations between the measures.

To date, few studies have explored the relationship between allostatic load summary measures and depressive symptoms. Allostatic load summary measures have been related to contemporaneously measured depressive symptoms (Juster et al., 2011; Kobrosly, Seplaki, Cory-Slechta, Moynihan, & van Wijngaarden, 2013; Kobrosly, van Wijngaarden, Seplaki, Cory-Slechta, & Moynihan, 2014) and depressive symptoms measured years later (Gale et al., 2015; Goldman et al., 2006; Juster et al., 2011). Research has not yet fully explored how allostatic load and depressive symptoms relate over time. Notably, the relationship between change in allostatic load and change in depressive symptoms has not been tested. This change-change relationship can be tested by correlating the slopes in a latent growth curve model. Chapter 6 will fit latent growth curve models to investigate the contemporaneous and longitudinal relationships between allostatic load and depressive symptoms.

1.6 Genetic modification of allostatic load biomarker influences

As mentioned previously, relationships between allostatic load and cognitive ability have been reported. What is less clear is whether the relationship between any allostatic load marker and cognitive ability is modified by the possession of a particular genetic variant. Understanding whether these relationships are moderated by genes is important because it could improve our understanding of the mechanisms of allostatic load and potentially have implications for genetically informed clinical practice.

One previously reported but unclear gene-by-allostatic load biomarker interaction effect involves the interaction between two variables independently linked to cognitive ability: the e4 allele of the apolipoprotein E (*APOE*) gene and cortisol levels. *APOE* e4 has been related to lower cognitive ability and greater cognitive decline (Davies et al., 2015, 2014). Also, various measures of cortisol levels and dynamics have been related to cognitive ability (Beluche, Carrière, Ritchie, & Ancelin, 2010; Franz et al., 2011; Lee et al., 2007) and cognitive decline (Franz et al., 2011; Li et al., 2006). Some measures of cortisol tend to be positively associated with cognitive outcomes (e.g., the cortisol awakening response; Stawski et al., 2011) while others are negatively associated with cognitive outcomes (e.g., evening cortisol levels; Stawski et al., 2011). Some studies have suggested that *APOE* e4 status may moderate the relationship between cortisol and cognitive ability such that it is more negative in e4 carriers than e4 non-carriers, however, not all studies on this topic have formally tested the interaction, and the results have varied across measures of cortisol and cognitive ability (Gerritsen, Comijs, Deeg, Penninx, & Geerlings, 2011; Lee et al., 2008; Singh-Manoux et al., 2014).

Chapter 7 will use a range of cortisol measures and cognitive tests to attempt to find out whether the level and dynamics of the secretion of cortisol, a key primary mediator of allostatic load, are related to different measures of cognitive ability and whether these relationships are moderated by the possession of the *APOE* e4 allele.

1.7 Summary

In summary, much has been learned about allostatic load over the last 20 years, but questions remain regarding the relationships between allostatic load biomarkers, the best way to operationalise allostatic load, and how allostatic load relates to other broad physiological dysregulation measures. Questions also remain about how allostatic load relates to cognitive ability and depressive symptoms over time, and how genetic variants may modify relationships between allostatic load biomarkers and cognitive ability. This thesis will attempt to make a unique contribution to the allostatic load literature by answering some of these questions.

Chapter 2

Considering the factor analytic operationalisation of allostatic load

This chapter is based around a letter that I (Z.C.) and my primary supervisor (T.B.) wrote to the editor of *Psychosomatic Medicine* in response to a study by Wiley et al. (2016). Their study argued for a bifactor confirmatory factor analysis operationalisation of allostatic load. In this letter, we raised concerns with factor analytic operationalisations of allostatic load. The primary positive contribution of this letter is that we put forward desirable properties of allostatic load scores. These properties will help both researchers choosing an allostatic load operationalisation and researchers conducting future allostatic load measurement studies. Thanks to funding from the Economic and Social Research Council, the letter was published open access with a Creative Commons Attribution 4.0 International (CC BY 4.0) licence (Crook & Booth, 2017b). Here, the letter is presented with the only edits made to match the letter to the format of this thesis. After the edited reproduction of the letter, the chapter concludes with a coda that presents this thesis's allostatic load operationalisation and explains why it was chosen.

2.1 Letter to the editor introduction

In a recent issue of *Psychosomatic Medicine*, Wiley et al. (2016) made a valuable contribution to the discussion of the optimal measurement of allostatic load. In the most comprehensive factor analytic investigation of allostatic load to date, they found that a bifactor model with a general AL factor and seven physiological system factors fits better than a higher-order model in which the seven system factors loaded on the general AL factor. Similar models have been applied by the author (T.B.) and others to operationalise allostatic load (Booth et al., 2013; Howard & Sparks, 2016; Seeman et al., 2010). Here, we consider the primary theoretical assumptions underlying latent variable modelling, argue that the construct of allostatic load is inconsistent with these assumptions, and propose alternate operationalisations of allostatic load.

2.2 Underlying construct (common cause)

A latent variable model is estimated based on the patterns of covariance in a set of variables. By including an allostatic load general factor in a latent variable model, researchers are positing that an underlying construct is the common cause of the observed covariation in all of the modelled biological measures. Although the theoretical relation of the common cause or construct to the original variables differs in bifactor versus higher-order models, in either case, we must ask: What could this common factor be? Wiley et al. stated that the AL factor "[captures] the notion that there is an underlying process influencing multiple physiological systems” (Wiley et al., 2016, p. 4). However, the observation of a general factor estimated from interindividual summary statistics (i.e., covariances) says little about what this process may actually be.

2.3 Independence conditional on the latent trait

In a typical latent variable model, the latent variable is modelled as causing the values on the indicator variables. A set of indicator variables are correlated, but this correlation is solely a result of the shared causal relation with the latent variable. As a result, once the latent variable has been modelled based on the correlations between indicator variables, the indicator variables should be uncorrelated, or independent. For example, a personality psychologist might model the latent variable of extraversion as causing people to report both feeling “comfortable around people” and being “the life of the party”, and the two items would be assumed to be independent after conditioning on extraversion. If, instead, feeling comfortable around people caused people to be the life of the party, and this dependence remained after conditioning on the latent variable, the assumption of conditional independence would be violated.

The assumption of conditional independence is unlikely to hold in an AL latent variable model. Levels of different biomarkers are linked causally to each other, rather than only through the common cause latent variable(s). For example, body mass index (BMI) has previously been used as a metabolic system allostatic load biological measure (e.g., Booth et al., 2013; Crimmins, Johnston, Hayward, & Seeman, 2003). However, Mendelian randomisation studies have found that increased BMI has a causal effect on levels of other metabolic biological measures as well as levels of allostatic load biomarkers used to represent other physiological systems, such as blood pressure and inflammation (e.g., Millard et al., 2015). Thus, it is most likely that the biomarkers are not conditionally independent but are instead dynamically related in complex networks. Such networks can produce observed correlations between variables that have no common cause (van der Maas et al., 2006).

2.4 Interchangeability of indicators

A further assumption of the latent variable model is that the definition of the latent variable does not change when different sets of indicators are used (Bollen & Bauldry, 2011). This holds because the indicators are affected by, but do not affect, the latent variable. Another key finding of Wiley et al. was that fitting models in which the biological measures from each of the 7 physiological systems were excluded caused no large changes in allostatic load factor loadings (Wiley et al., 2016). This method provides only a weak test of interchangeability. The stability of general intelligence factor loadings has long been a research focus for intelligence researchers, so allostatic load researchers may benefit from applying their approaches to this issue (e.g., Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Thorndike, 1987). For example, researchers could compute and correlate allostatic load scores from different nonoverlapping multisystem sets of biological measures (Bollen & Bauldry, 2011). The existence of diverse causal links between allostatic load biological measures from different physiological systems suggests to us that the nature of what relates the biomarkers may change depending on which measures are included in the model. We predict that more thorough, more powerful tests of the stability of AL factor loadings will find that it does not hold.

2.5 Formative versus reflective indicators

In the common factor model, the biological measures are reflective indicators, that is, they are manifested by a common cause latent variable. However, to the extent that the model assumptions are violated (previously discussed), the factor model is not appropriate. Thus, it may instead be profitable to consider the biological measures as formative indicators, that is, as variables that define the construct (Bollen & Bauldry, 2011). This way of thinking about how the biological measures relate to allostatic load is consistent with any number of weighted or sum scores. It is also consistent with allostatic load theory, in that more severe, more widespread physiological dysregulation will relate to higher AL scores.

Alternatively, the associations between AL biological measures could be modelled using each measure individually, without the need for any single latent or observed summary. This could be done with network analysis, which has been used beneficially by researchers studying symptom networks in mental disorders (Borsboom & Cramer, 2013). It should be noted that for any given network model, there is a statistically equivalent factor analysis model (Kruis & Maris, 2016). This does not mean, though, that both models are reasonable, nor that they are theoretically equivalent (Borsboom, 2017; Epskamp & Fried, 2018). Network models can be more insightful for some research problems because they switch the focus away from the common variance and onto the individual indicators and their interrelationships (Epskamp & Fried, 2018). This approach could be fruitful for allostatic load research.

However, large-sample longitudinal measurements of allostatic load biomarkers are typically collected years apart, so longitudinal network analyses using currently available data may not reveal much about the causal connections between allostatic load biomarkers

(Fried et al., 2017). To investigate the strength and directions of the links between biomarkers, future research should aim to collect longitudinal data on allostatic load biomarkers with shorter intervals between measurements, such as hours, days or weeks rather than years (Eatough, Shockley, & Yu, 2016; van Ockenburg, Booij, Riese, Rosmalen, & Janssens, 2015). This can be combined with data collection on activities, perceived stress, and psychological symptoms (Beal, 2015; Treadway & Leonard, 2016).

Allostatic load indicators can also be modelled separately without consideration of their associations. Consistent with this approach, *Psychosomatic Medicine* typically provides data of separate biological measures when articles report about complex phenomena such as AL and metabolic syndrome.

Aside from any issues regarding model assumptions, two further points warrant comment about the models presented by Wiley et al.

2.6 Improved model fit for bifactor approach

Mediators of the stress response from different physiological systems are causally linked in nonlinear networks that include reciprocal associations (McEwen, 2006). The complex structure of the associations between allostatic load markers helps to explain why the bifactor allostatic load model fits better than the hierarchical allostatic load model. The hierarchical model imposes "proportionality constraints" (Yung, Thissen, & McLeod, 1999, p. 115) the ratio of the allostatic load general factor loadings to the system factor loadings is constrained to equality within the biological measures of each physiological system. Considering the diverse causal links between different allostatic load biomarkers, both within and across systems, these proportionality constraints are likely to be violated. Furthermore, it has been shown that when the true model contains “unmodelled complexity” (Murray & Johnson, 2013, p. 407) in the form of small correlated residuals and cross-loadings, or even modelled complexity in the form of correlated residuals across factors, fit indices and criteria may be biased in favour of the bifactor model. Consequently, the better fit of the bifactor model follows from allostatic load theory and research, as well as from methodological findings, for reasons other than those Wiley et al. (2016) focused on.

2.7 Variance explained by physiological dysregulation factors

Statistically, a desirable property of a general factor is that it accounts for most variance in the constituent indicator variables. In the study by Wiley et al., the allostatic load factor explained only approximately 11% of the variance in the AL biological measures. Some of the physiological system-specific factors were also weak. For example, the hypothalamic-pituitary-adrenal axis and inflammation factors explained only approximately 9% and 16% of the variance in their respective biological measures. Note that weak factor saturation of physiological dysregulation factors has also been an issue in other samples (Booth et al., 2013; Seeman et al., 2010).

2.8 Properties of optimal scores for allostatic load

Ideally, allostatic load scores should be: 1) calculated using biological measures from various physiological systems; 2) consistently calculated across samples; and (3) closely related to criterion variables. Those who desire scores that are rooted in allostatic load theory would prefer the allostatic load scoring method that produces the scores most closely related to chronic/repeated perceived stress. For a pragmatist, the focus may not be on investigating how different physiological dysregulation scores relate to prior perceived stress but rather on finding the scores that most strongly predict important health outcomes such as cardiovascular disease and death. It may also be advantageous to have scores that explicitly represent the accumulation of the effects of repeated environmental challenges.

Our theoretical and methodological concerns with the factor analytic operationalisation of allostatic load suggest to us that factor scores will not prove to be the optimal allostatic load scoring method. We therefore believe that further research is required to determine the optimal operationalisation(s) of allostatic load.

2.9 Coda

It is worth emphasising that the critiques in this letter do not only apply to allostatic load research. Similar issues exist in other subject areas, including psychology. For example, depression is frequently modelled as a latent variable, but network models have recently suggested that depressive symptoms (depression items) are not independent conditional on the latent trait because there are causal links between some symptoms (McNally, 2016). Furthermore, personality factors may violate the assumption interchangeability of indicators (Konstabel et al., 2017) and intelligence factor models have been shown to be biased towards showing better model fit for the bifactor versus higher-order factor models (Gignac, 2016; Morgan, Hodge, Wells, & Watkins, 2015; Murray & Johnson, 2013).

Outside of psychology, socioeconomic status has been modelled as a latent variable (e.g., Scharoun-Lee, Adair, Kaufman, & Gordon-Larsen, 2009), but the factor indicators used do not have a clearly defined common cause. Like allostatic load, it may be more appropriately modelled as a formative rather than a reflective latent variable, or not as a latent variable at all.

In this thesis, we chose to use to calculate allostatic load scores using the mean absolute *z* score method. Teasing apart how scores are calculated using this method elucidates its favourable properties. Like with the vast majority of allostatic load operationalisations, scores are calculated from a panel of biomarkers from different physiological systems. Firstly, the distribution of each marker is checked and if a marker's distribution is not approximately Gaussian, then it will be transformed to be so. Then, the values of each biomarker are transformed into *z* scores where 0 is the mean and 1 the standard deviation. Next, for all biomarkers where risk can be indicated at both ends of the distributions - that is, high or low values - the *z* scores are changed into absolute values. This means that, for example, whether an individual has systolic blood pressure 2 standard deviations above or below the mean, the contribution to the allostatic load score is the same. Finally, the intraindividual mean of all available absolute *z* scores is calculated to give their allostatic load score.

The *z* score approach has been supported by studies that have compared how scores from different allostatic load operationalisations relate to criterion variables. Using data from the Social Environment and Biomarkers of Aging Study, Seplaki et al. (2005) compared nine allostatic load operationalisations by calculating scores and comparing their associations with five criterion variables: self-rated health; mobility limitations; the presence of any limitations in activities of daily living; depressive symptoms; and mistakes in identifying the current day, date, month and year. The only two operationalisations associated with all five criterion variables were a *z* score measure and a method summing all markers in the top or bottom sample deciles (Seplaki et al., 2005). The authors concluded that their results support the use of allostatic operationalisations that preserve the continuous nature of the constituent biomarker data and incorporate risk at both tails of biomarker distributions where appropriate (Seplaki et al., 2005). Although the top and bottom decile sum performed well, it is unlikely to be the optimal operationalisation as it requires dichotimising each allostatic load biomarker, which can introduce bias (MacCallum, Zhang, Preacher, & Rucker, 2002).

A recent study provided another real data example of criterion variable effect sizes being greater for *z* score allostatic load scores compared to allostatic load scores computed with dichotimisation. The study used data from the National Health and Nutrition Examination Survey III to compare the predictive power of allostatic load scores for all-cause mortality, cardiovascular disease mortality, and cancer mortality (*n* = 9,942; Levine & Crimmins, 2014). As noted in Chapter 1, the effect sizes varied based on the operationalisation of allostatic load. For all three criterion variables, allostatic load scores based on absolute *z* scores outpredicted allostatic load scores based on clinical or other previously defined thresholds (Levine & Crimmins, 2014).

We chose not to test operationalisations other than the z-score method because other widely-tested operationalisations involve dichotomising indicator variables, which can bias estimates. Dichotimising variables removes meaningful information, typically reduces effect sizes and statistical power, and can produce potentially spurious effects, including interaction effects (MacCallum et al., 2002; Maxwell & Delaney, 1993; Thoresen, 2019).

In our allostatic load studies, we used secondary data from the Lothian

Birth Cohort 1936. The following chapter will outline the methods common to all studies in thesis using an allostatic load summary measure, including the biomarkers used to calculate scores.

Chapter 3

Lothian Birth Cohort 1936 methods

This chapter contains the methods shared by the first three of four empirical studies in this thesis. These studies used the same sample, the same allostatic load operationalisation, and the same statistical modelling approach to testing hypotheses.

The three studies will test how allostatic load relates to: 1) other measures of physiological weathering; 2) cognitive ability; and (3) depressive symptoms. We were interested not only in contemporaneous relationships, but also in associated change relationships, that is, whether change over time in allostatic load is related to change over time in the other variable(s). Where previous research on these relationships exists, it has tended to test contemporaneous relationships and has rarely, if ever, investigated associated change relationships, in part due to the lack of suitable data. The Lothian Birth Cohort 1936 (LBC1936) study offered the rare opportunity to test these associated change relationships. In these three studies, we tested contemporaneous and associated change relationships by fitting latent growth curve models to data from Waves 1-4 of the LBC1936.

To avoid unnecessary repetition and ensure the unique methodological aspects of each study are highlighted, this chapter contains details of the shared methods and the respective study chapters contain details of the measures and other methods specific to each study.

3.1 Participants

The LBC1936 study is a healthy ageing study that has followed up Scottish Mental Survey 1947 participants in older age. On 4th June 1947, the Scottish Mental Survey 1947 tested the intelligence of almost 1936-born children attending school in Scotland. In all, 70,805 children participated. The LBC1936 study followed up participants based in the Lothian area, in and around the city of Edinburgh. The study aimed to conduct in-depth cognitive testing to study cognitive ageing since childhood and in older age. It also aimed to collect data on a diverse range of variables related to cognitive ageing, including demographic, social, genetic, medical, and physiological variables.

Lothian-based 1936-born residents were identified using the Lothian Community Health Index, a record of patients registered at GP practices in the area. Recruitment letters were sent to 3,686 of these residents between June 2004 and November 2006. Initially, there were 1,703 responses, 1,351 of whom were interested in hearing more about the study. A second letter was sent to 1,741 residents, from which there were 615 responses, 216 of which were from people interested in participating. Participants were also recruited via newspaper advertising, although not many participants were recruited from this due to the comprehensive nature of the Community Health Index records. In all, 1,226 people were interested and eligible. Of these people, 85 withdrew before the start of the study and a further 50 were not able to be contacted or were not able to attend an appointment before the end of Wave 1 testing in May 2007. This left a total of 1,091 participants at Wave 1. At each wave, around a fifth of participants have dropped out, with the reasons for dropout including withdrawal with no further reason, death, ill health, and loss of contact (Taylor, Pattie, & Deary, 2018). Data was collected for Wave 1 in 2004-2007 (*n* = 1,091 [50.2% male]; *M* age = 69.6), for Wave 2 in 2007-2010 (*n* = 866 [51.7% male]; *M* age = 72.5), Wave 3 in 2011-2013 (*n* = 697 [51.6% male]; *M* age = 76.3), and for Wave 4 in 2014-2017 (*n* = 550 [50.0% male]; *M* age = 79.3).

Each wave comprises a participant interview and a study questionnaire. The participant interview includes questions on social and medical variables, cognitive tests, physical tests, and the taking of blood samples. The study questionnaire was completed after the participant interview and included social, demographic, and activity questions, personality questionnaires, a quality of life survey, and a food questionnaire.

The LBC1936 study provides a rich dataset that is well-suited to answering the studies' research questions. A data search conducted prior to this thesis found that the LBC1936 possessed various advantages over other studies with allostatic load biomarkers, particularly for longitudinal and psychological studies. The LBC1936 study has four waves of data on the same allostatic load biomarkers, which enables the relationships between longitudinal change in allostatic load and longitudinal change in other variables to be investigated. The range of biological variables, including DNA methylation and telomere length, enables longitudinal relationships between allostatic load and other physiological weathering measures to be tested. A further strength is that cognitive ability has been tested using a full battery of gold standard intelligence tests. In cohort studies without a focus on cognitive ageing, it is more common to test cognitive ability using only a few cognitive tests, or even only a cognitive impairment screener.

The sample also has some notable limitations. No primary mediators of the stress response have been measured in the full sample, which means the allostatic load measure was computed using data from secondary outcomes of the stress response only. This also means that no neuroendocrine biomarkers were included in allostatic load calculation. Further, the generalisability of findings in the sample is limited because the sample has a limited age range, is ethnically white, and has higher premorbid intelligence than the population it was sampled from (Čukić, Brett, Calvin, Batty, & Deary, 2017; Gow et al., 2011). Dropout may also have reduced the representativeness of the sample. Compared to returners, participants who subsequently dropped out tend to have had lower socioeconomic position, lower older age cognitive ability, and lower physical fitness (Taylor et al., 2018).

For further information on participant recruitment and the procedures of the study, see the LBC1936 cohort profiles (Deary, Gow, Pattie, & Starr, 2012; Deary et al., 2007; Taylor et al., 2018).

3.2 Allostatic load

3.2.1 Allostatic load biomarkers

Values from nine allostatic load biomarkers were used to calculate individuals' allostatic load scores. Body mass index (BMI) was computed by dividing participants' weight (in kilograms) by their squared height (in metres). Systolic blood pressure (SBP) and diastolic blood pressure (DBP) input variables were the mean of three seated readings. The following biomarkers were measured using blood samples: albumin; C-reactive protein (CRP); fibrinogen; glycated haemoglobin (HbA1c); total cholesterol to high-density lipoprotein ratio (HDLR); and triglycerides. Blood samples were taken by a research nurse at the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh. Albumin was measured using Vitros ALB slides, with colourimetric tests conducted using the Vitros Fusion 5.1 FS and Vitros 4600 Chemistry Systems. C-reactive protein (CRP) was measured with the OrthoFusion 5.1 F.S. analyser using dry immuno-rate slides (Vitros Chemistry Products CRP slides, Ortho Clinical Diagnostics, Buckinghamshire, UK). Fibrinogen was measured with an automated Clauss assay (TOPS coagulator, Instrumentation Laboratory, Warrington, UK). Non-fasting glycated haemoglobin (HbA1c) was measured with an Adams HA-8160 HbA1c analyser, which uses a high performance liquid chromatography method. High-density lipoprotein and total cholesterol were measured using the Abbott Architect c16000 and the high-density lipoprotein ratio was calculated by dividing total cholesterol by high-density lipoprotein. Triglycerides were also measured using the Abbott Architect c16000.

The allostatic load indicators were chosen based on their availability in the sample. Regarding the physiological systems represented by the allostatic load indicators, there were two markers from the cardiovascular system (SBP and DBP), three markers from the the immune system (albumin, CRP, and fibrinogen), and four markers from the metabolic system (body mass index, HbA1c, HDLR, and triglycerides). A limitation of the panel of allostatic load biomarkers is that it does include any neuroendocrine system markers.

Our allostatic load measure has some overlap with the markers used for other physiological summary measures. The Framingham Risk Score also uses SBP and uses HDL and TC as separate markers, rather than the ratio between them (D’Agostino et al., 2008). The metabolic syndrome is diagnosed using three of the markers we used (SBP, DBP, and triglycerides), one marker we used in creating a variable (HDL), a different obesity marker (waist circumference rather than BMI), and one other marker (fasting glucose; Alberti et al., 2009). Hence, our measure has a focus on metabolic and related markers, with the addition of immune system markers.

Regarding distributions of the markers, we preregistered that we would not remove any outlier values (see section 3.5; Crook et al., 2017a, 2017b, 2017c). We did transform biomarker values with skewed distributions (see section 3.2.2).

3.2.2 Allostatic load calculation

An individual's allostatic load score was the mean of the absolute *z* scores for their allostatic load biomarkers. For one of these biomarkers, the total cholesterol to high-density lipoprotein ratio, risk is indicated by only high values, so *z* scores below zero on this biomarker were changed to zero before overall allostatic load scores are calculated.

For each of the nine allostatic load indicator variables, skewness was calculated at Wave 1. We planned that if a variable had Wave 1 skewness $> |1.5|$, then the optimal univariate Box-Cox transformation for the Wave 1 data would be found and applied to all waves of data of that variable. This was the case for three allostatic load indicator variables at Wave 1. The skew of triglycerides was 5.38, the skew of CRP was 4.86, and the skew of HbA1c was 2.99. Hence, all waves of these variables were transformed. The *z* scores of the resulting nine variables were summed to create allostatic load scores for each participant at each wave.

We also planned that if the Wave 1 allostatic load score had skewness $>$ $|1.5|$, then the same Box-Cox transformation process would be repeated for all waves of this variable, but the skewness was below this threshold and so the allostatic load variables were not transformed.

3.3 Covariates

Age in days at each day of assessment was recorded.

Sex, as observed at birth, was recorded at the time of the Scottish Mental Survey 1947.

3.4 Statistical analysis

3.4.1 Latent growth curve models

Latent growth curve models (also known as latent growth models or latent curve models) are fitted to repeated measures data, typically to answer research questions about change over time. In a latent growth curve model, there is an intercept (typically level at time zero) and a slope (change over time) for every individual in the sample. This means that individual differences in trajectories are included in the model, so the relationships between trajectories and other variables can be tested. For example, researchers can test the relationships between the trajectory of a variable and trajectories of other variables, and/or time-invariant covariates (e.g., sex), and/or time-varying covariates (e.g., marital status).

Latent growth curve models are typically fitted within either a structural equation modelling framework or a multilevel modelling framework. Often, but not always, numerically identical models can be fitted within either framework (Curran, Obeidat, & Losardo, 2010). In this thesis, the structural equation modelling framework will be used, because it facilitates the inclusion of latent variables (e.g., general cognitive ability) and the assessment of absolute model fit (Curran et al., 2010). In the structural equation modelling framework, the level and change in the growth curve variable is represented by intercept and slope latent variables. The observed variable values load on the intercept and slope latent variables (Bollen & Curran, 2006).

The studies in the following three chapters attempt to answer novel research questions about the relationships between trajectories in allostatic load and trajectories in other variables. In some cases, we are also interested in the average levels and trajectories and the variation around them, and in how the levels and trajectories relate to covariates. Hence, latent growth curve modelling is an ideal method for these studies.

The studies fitted parallel process latent growth curve models, in which the levels and trajectories of multiple variables are modelled together and can be related to each other. Figure 3.1 shows the general specification of the key parameters in the parallel process latent growth curve models. Note that some parameters are omitted from the figure for ease of presentation.

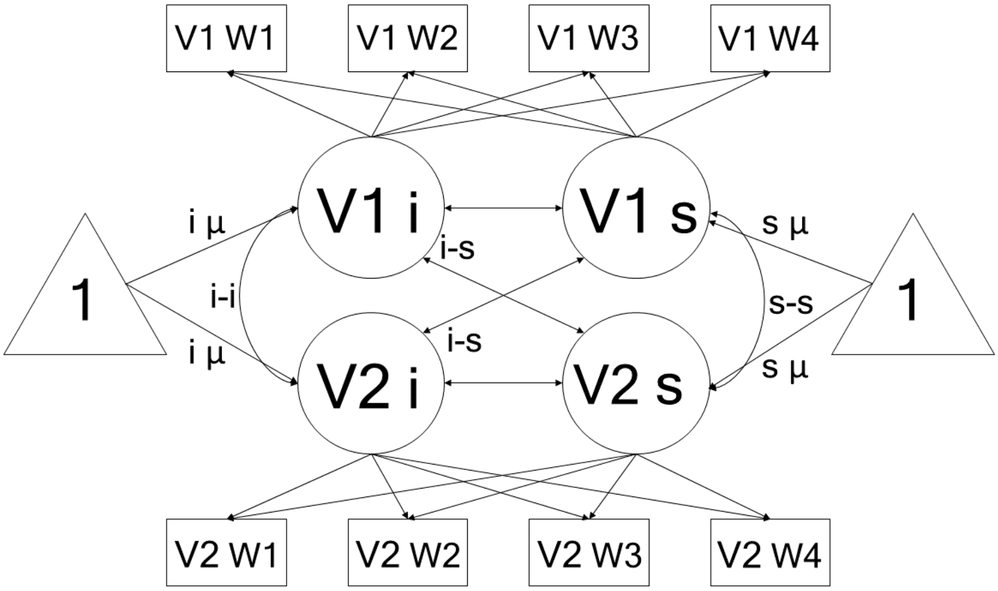

Figure 3.1: General parallel process latent growth curve model specification. i = intercept. s = slope. V = variable. W = wave.

Here, V1 and V2 are two variables being modelled longitudinally based on data from Waves 1-4. In this kind of path diagram, latent variables are drawn as circles, observed variables inside rectangles, and constants inside triangles. Each intercept growth factor (V1 i and V2 i) is estimated by fixing the loadings of the growth indicator variables (V1 W1-4 and V2 W1-4) on the intercept growth factor to 1. The arrows from the "1" constants to the growth factors show that means are estimated for each of the growth factors. Growth factor variances have been omitted from the diagram for clarity but will also be estimated in our models. The mean of an intercept factor (i μ) represents the average level of the construct at baseline and the variance of the intercept factor represents the individual variation around this average. An intercept-intercept or i-i correlation between two measures represents their association at baseline. The slope factors (V1 s and V2 s) are estimated by fixing the loadings of Wave 1 indicators to 0, and the loadings for Waves 2-4 at the mean number of years since Wave 1 data collection (in our studies: 2.98, 6.74, and 9.81). With this scaling, the mean of a slope factor (s μ) provides an estimate of the average linear change per year and the variance of a slope factor provides an estimate of the individual variation around this average. An intercept-slope or i-s correlation represents the association between the baseline level in one measure and the subsequent change in another. The slope-slope or s-s correlation between two variables represents their dual change association, that is, the association between longitudinal changes in the measures.

3.4.1.1 Covariates

Age differences within wave were controlled for by including age at each wave as a predictor of all growth trajectory indicators at its respective wave (e.g., Wave 3 age was included as a predictor of Wave 3 allostatic load). In our models, age was what is typically referred to in the methodological literature as a time-varying covariate, as within-wave differences varied across waves. Age was rescaled so it was in years. This was done to avoid model-fitting issues that can arise due to differences in the sizes of variances. To aid the interpretability of intercepts, the age of the youngest participant at Wave 1 cognitive testing in whole years was be subtracted from all age values (e.g., if the youngest participant was 60.2 years, 60 would be subtracted from all ages). To statistically control for sex, it was included as a predictor of all growth factors. In our models, sex was a time-invariant covariate as it did not change over the course of the study. It was a numeric binary variable with females coded as 0 and males coded as 1[1].

---

[1] We preregistered a plan to code females as -0.5 and males as 0.5. We used a different coding system to the one in our preregistrations because we decided that using a more commonly used coding system would facilitate model interpretation. Note that while the change in coding affected the covariates' parameter estimates, it did not affect the hypothesis test results.

3.4.1.2 Longitudinal dropout

All models were fitted using full information maximum likelihood estimation, which means that participants with some data missing were included in the models. At Waves 2-4, the sample size was 20-21% lower than at the previous wave. To investigate the potential influence of study dropout on the substantive conclusions of the study, we fitted our parallel process latent growth curve models using two approaches, each of which makes different assumptions about the missing data.

The first set of models assumed missing data were missing at random (MAR), that is, the models assumed the missingness of the data was random once the observed data in the model was taken into consideration (Enders, 2010). The second set of models assumed missing data were not missing at random (NMAR; such data can also be referred to as missing not at random [MNAR]), that is, the missingness of the data may be related to the unobserved missing values even after taking into consideration the observed data in the model (Enders, 2010). For example, high allostatic load values may be more likely to be missing: a participant who experiences a particularly large increase in allostatic load after Wave 1 may not survive long enough or be healthy enough to participate at Wave 2 (see Hwang et al., 2014 on allostatic load and mortality).

The NMAR models we fitted were pattern-mixture models (Muthén, Asparouhov, Hunter, & Leuchter, 2011). In these models, three dummy variables representing dropout time were added (d2-d4) as predictors of all the growth factors. These dummy variables were coded 1 if a participant dropped out at the corresponding wave and 0 otherwise. For example, for a participant who provided data at Waves 1 and 2 but not Waves 3 and 4, the dummy variables would be coded as follows: d2 = 0, because the participant provided data at Wave 2; d3 = 1, because the participant dropped out at Wave 3; and d4 = 0, because the participant had previously dropped out. The parameter estimate for the regression of a growth factor on a dummy variable represented the difference in the growth factor mean for those who dropped out at the corresponding wave (Muthén et al., 2011). For example, if the regression of the allostatic load slope on d4 produced a strong positive estimate, this would indicate that those who dropped out at Wave 4 had a more positive allostatic load slope than those who did not drop out at Wave 4. This would mean that Wave 4 dropouts were modelled as having a steeper increase in allostatic load between Waves 1 and 4. The regressions of each slope growth factor on d2 and d3 were constrained to equality to identify the model.

We used pattern-mixture modelling because it associates dropout with the level and change of the variables being modelled, rather than the values of specific measurements (as in some NMAR selection models, e.g., Diggle & Kenward, 1994), which makes more sense given our variables, and because it enabled us to estimate and easily interpret the parameters testing our hypotheses, unlike latent class NMAR models (Dantan, Proust-Lima, Letenneur, & Jacqmin-Gadda, 2008).

For all models discussed above, we compared the results of the MAR and NMAR models. We focused our interpretation on the MAR models and noted where there were differences between the MAR and NMAR models.

3.4.1.3 One process models

The focus of our hypotheses was generally on the parallel process latent growth curve models, but in each study we also fitted and report results from one process models, where only one set of growth factors was included (e.g., the allostatic load intercept and slope). Unconditional models, which had no covariates included, helped us to assess whether the linear latent growth curve model fit the data well. Conditional MAR models, which added sex and age as covariates, were used to interpret sex differences where they were of interest. Finally, we also interpreted the regressions in conditional NMAR models, which added the dummy variables for dropout as predictors of the growth factors. Note that for convenience, some allostatic load one process models are reported in multiple chapters.

3.4.2 Model testing

We used the R programming language (version 3.5.1; R Core Team, 2018b) within the RStudio integrated development environment (version 1.1.456; RStudio Team, 2016) for data wrangling and statistical analyses. R packages helped to facilitate this. We used devtools (version 1.13.6; Wickham, Hester, & Chang, 2018) for package acquisition, foreign (version 0.8.71; R Core Team, 2018a) for data importing, psych (version 1.8.4; Revelle, 2018), taRifx (version 1.0.6; Friedman, 2014), and mice (version 3.3.0; van Buuren & Groothuis-Oudshoorn, 2011) for data description and data wrangling, and ggplot2 (version 2.2.1; Wickham, 2009) with ggthemr (version 1.1.0; Tobin, 2018) for plotting.

Presentation of the results was facilitated by rmarkdown (version 1.9; Allaire et al., 2018), knitr (version 1.20; Xie, 2018b), kableExtra (version 0.9.0; Zhu, 2018) and bookdown (version 0.7; Xie, 2018a).

The lavaan package (version 0.5.23.1097; Rosseel, 2012) was used for model fitting. To correct for possible multivariate non-normality, robust standard errors for model parameters were computed using a sandwich estimator.

3.5 Preregistration

Each of these studies was preregistered on the Open Science Framework (Crook et al., 2017a, 2017b, 2017c). The preregistrations are early versions of the Open Science Framework projects, frozen after the study was planned but before data analysis was conducted. Each contains a background document about the rationale for the study and a structured preregistration document with the design of the study. Note that the introduction and method sections of the LBC1936 chapters in this thesis were based on the documents from the preregistrations.

Chapter 4

Allostatic load, biological age, epigenetic age and telomere length: how do they relate over time in the Lothian Birth Cohort 1936?

4.1 Introduction

Stress and ageing are globally recognised public health concerns (World Health Organization, 1998, 2015). It is crucially important that we seek to understand the mechanisms underlying the adverse health effects of stress and ageing so we can improve the prevention and treatment of stress- and ageing-related morbidity.

Both stress and ageing can cumulatively damage the body, causing physiological weathering that can be measured in various ways. Stress is a core aspect of the theory underlying biomarker panel-based summary measures of allostatic load. Ageing, meanwhile, has motivated research into leukocyte telomere length as well as the development of biomarker panel-based measures of biological age and DNA methylation-based measures of epigenetic age. Allostatic load, biological age, epigenetic age, and telomere length have each been related to important health outcomes, including mortality (Hwang et al., 2014; Levine, 2013; Marioni et al., 2015; Rode, Nordestgaard, & Bojesen, 2015).

Although the foundations of these physiological weathering measures overlap, few studies to date have investigated the relationships between them, and no studies to date have related all four measures. In particular, more research is needed into how these physiological weathering measures relate in older age, including the longitudinal associations between changes in these measures.

The first part of this chapter will introduce biological age, epigenetic age and telomere length (for an introduction to allostatic load, see Section 1.1). Next, we will discuss overlap in the underpinnings of these physiological weathering measures and review previous studies relating them. We will then introduce the present study, which will investigate the contemporaneous and longitudinal relationships between these physiological weathering measures in older age.

4.1.1 Biological age

For over 50 years, researchers have been estimating biological age by using chronological age together with ageing-related variables (Hollingsworth, Hashizume, & Jablon, 1965). The Klemera-Doubal method (Klemera & Doubal, 2006), which uses parameter estimates from regressions of chronological age on biological age biomarkers, has been preferred by studies comparing biological age methods (Cho et al., 2010; Jee & Park, 2017; for a review of methods, see Jia et al., 2017). Biological age estimation has incorporated markers from the immune, metabolic and cardiovascular systems, as well as markers of kidney, liver and lung function (Levine, 2013). A recent study of 38-year-olds found Klemera-Doubal biological age estimates were normally distributed, with a mean of 38 years and a standard deviation of 3.2 years (Belsky et al., 2015). The mean biological and chronological ages matching shows that the two are close aligned, but the notable deviation around the mean shows there is substantial individual variation in humans' biological ages estimated using this method.

4.1.2 Epigenetic age

Epigenetic age (sometimes referred to as biological age) is estimated using DNA methylation data. DNA methylation occurs when a methyl group is added to cytosine nucleotides on the genome at cytosine-phosphate-guanine (CpG) sites (Beck & Rakyan, 2008). The proportion of methylation at CpG sites can be measured (Beck & Rakyan, 2008). Recently, through working with models that predict chronological age from methylation at particular CpG sites, researchers have developed methods to estimate epigenetic age from DNA methylation in blood and/or other tissues (Hannum et al., 2013; Horvath, 2013). Epigenetic age estimates from these methods are strongly related to chronological age and, when combined with chronological age, can be used to generate estimates of epigenetic age acceleration/deceleration, the pace of the epigenetic clock (Hannum et al., 2013; Horvath, 2013). Extrinsic epigenetic age acceleration estimates can also be calculated. These estimates adjust epigenetic age acceleration estimates for blood cell composition using the Klemera-Doubal method, the same method often used to calculate the previously mentioned estimates of biological age (Chen et al., 2016).

4.1.3 Telomere length

Lastly, telomere length has been used a measure of physiological weathering. Telomeres are repetitive DNA sequences that sit at the end of each chromosome, protecting them from damage (Riethman, 2008). Telomeres are shortened each time a cell replicates DNA and by oxidative stress, until they are shorter than a critical length, at which point DNA replication stops and the cell enters a state of senescence (Simons, 2015).

Telomeres shorten with age (Müezzinler, Zaineddin, & Brenner, 2013). Thus, many researchers have investigated telomere length as a physiological age measure, and some have previously incorporated telomere length in biomarker-panel based estimates of biological age (Zhang et al., 2014). However, the status of telomere length as a biomarker of ageing is contested. Some believe the evidence is inconclusive (Mather et al., 2011; Sanders & Newman, 2013), while others have considered it one of the most promising ageing biomarkers (Blackburn, Epel, & Lin, 2015; Jylhävä, Pedersen, & Hägg, 2017; López-Otín, Blasco, Partridge, Serrano, & Kroemer, 2013).

4.1.4 Overlap in the underpinnings of physiological weathering measures

Because allostatic load theory focuses on stress, and the other measures discussed here aim to track physiological ageing, it may seem that these measures relate only to distinct causes of physiological weathering. However, to some unknown extent, all these measures may index aspects of physiological weathering that can be caused by stress and/or ageing.

Although stress, rather than age, is key to the foundation of allostatic load theory, allostatic load summary measures have been linked to age (Crimmins et al., 2003). And while epigenetic age and telomere length are considered biomarkers of physiological age, they have both been linked to stress (Shalev et al., 2013; Zannas et al., 2015). Plus, recent research has suggested that physiological stress responses and ageing may interact to affect epigenetic ageing (see Gassen, Chrousos, Binder, & Zannas, 2017 for a review).

Another reason that these measures should be related is that there is overlap in the mechanisms underlying them. The mitochondrial allostatic load theory may help to explain part of the relationships between the measures (Picard et al., 2014). Mitochondria are organelles – cell parts with a particular purpose – but unlike other organelles, they have their own DNA. Mitochondria's main role is to generate energy and they generate most of the energy that cells require. Mitochondria are both involved in and affected by the hormones released when the body responds to stress, which means their function is linked to allostasis and levels of allostatic load (Picard et al., 2014). Chronic exposure to stress hormones can lead to mitochondrial allostatic load, where mitochondria and their function are negatively affected (Picard & McEwen, 2018). In a state of mitochondrial allostatic load, mitochondrial function can be impaired, mitochondrial DNA can be damaged, and mitochondria size and number can be increased or decreased (Picard & McEwen, 2018).

The mitochondrial allostatic load theory links allostatic load, telomere length and epigenetic age, because mitochondrial allostatic load can reduce telomere length and affect epigenetic functioning (Picard et al., 2014). Specifically, mitochondrial DNA damage that affects energy production can shorten telomeres, possibly through reactive oxygen species and/or oxidative stress (Passos, Saretzki, & von Zglinicki, 2007). There can also be causal effects in the reverse direction, as telomere dysfunction can cause mitochondrial dysfunction (Sahin et al., 2011).

Mitochondria can also be related to epigenetic age because they affect the methylation of genes (Picard & McEwen, 2018). In an in vitro study, the expression of most human genes was regulated by mitochondria (Picard, Zhang, et al., 2014).

Regarding a different overlap, C-reactive protein, a primary mediator of allostatic load that is also used in biological age estimation, was found to cause a reduction in telomere length in a recent Mendelian randomisation study (Rode, Nordestgaard, Weischer, & Bojesen, 2014).

In addition, although allostatic load and biological age use distinct algorithms, they typically incorporate several of the same markers in estimation. This means that the relationship between allostatic load summary measures and biological age estimates will be influenced not only by any overlap in the underlying mechanisms involved, but also by overlap in the indicators used. This may mean that only a study with very many allostatic load and biological age markers could successfully disentangle the part of the allostatic load-biological age association that relates to underlying mechanisms and the part that relates to overlap in the indicators used. This will be outside the scope of the present study.

4.1.5 Mechanisms of ageing

A 2013 review paper posited nine hallmarks of ageing (López-Otín et al., 2013). These were separated into: primary hallmarks, which cause cellular damage; antagonistic hallmarks, which work to reduce the damage but can become harmful themselves; and integrative hallmarks, which result from the other hallmarks and relate closely to ageing-related functional decline (López-Otín et al., 2013). The hallmarks are briefly introduced below with notes based on the review paper (López-Otín et al., 2013).

Primary hallmarks:

- **Genomic instability**: DNA (including mitochondrial DNA) can be damaged by exogenous (e.g., UV radiation) and endogenous (e.g., reactive oxygen species) variables. Examples of the resulting genomic instabilities are new mutations and chromosomes being moved or deleted.
- **Telomere attrition**: as mentioned above, telomeres shorten as cells replicate DNA. Mouse studies have established causal links from telomere shortening to cellular senescence and accelerated ageing, and have found that telomerase activation can decelerate ageing.
- **Epigenetic alterations**: these include DNA methylation, histone modification, and chromatin remodelling, which lead to epigenetic changes that accelerate ageing (e.g., worsened DNA repair, chromosomal instability).
- **Loss of proteostasis (protein homeostasis)**: this loss occurs when there are nonnative protein aggregates in tissues. It is caused by stress (e.g., heat shock or oxidative stress) and in turn it has deleterious effects on stress responses.

Antagonistic hallmarks:

- **Deregulated nutrient sensing**: increased nutrient signalling affects ageing through various paths, including through insulin-like growth factor 1 (IGF-1). It can be controlled through dietary restriction.
- **Mitochondrial dysfunction**: this dysfunction is caused by various defects, including mitochondrial DNA mutations, and it leads to further mitochondrial and cellular damage. It is related to ageing through higher reactive oxygen species production and higher inflammation, as well as through other pathways.

- **Cellular senescence**: in ageing, cells experience more damage and are repaired more slowly, which leads to cellular senescence, where cells no longer divide. It is commonly caused by telomere shortening. Possible consequences include increased inflammation and stem cell exhaustion. Cellular senescence is an adaptive response that only becomes harmful when tissues can no longer regenerate cells, so it may not meet all the criteria to be a hallmark of ageing.

Integrative hallmarks:

- **Stem cell exhaustion**: reduced stem cell activity can lead to muscle loss, reduced bone density, and increased inflammation. Stem cells are an important target for treating age-related disease and loss of function.
- **Altered intercellular communication**: inflammation negatively affects intercellular communication, which itself can affect inflammatory processes as well as neuroendocrine processes and communication between different tissue types.

Not all of these have been evidenced by clinical studies in humans (Shiels, Stenvinkel, Kooman, & McGuinness, 2017). Telomere attrition and epigenetic alterations are closely related to telomere length and epigenetic age respectively. Allostatic load and biological age estimates are not as directly related to the hallmarks, but may well be affected to some extent by all of them.

4.1.6 Relationships between measures of physiological weathering

Researchers have begun to explore the relationships between these physiological weathering measures.

4.1.6.1 Allostatic load, telomere length and biological age

In the National Health and Nutrition Examination Survey (NHANES) III, a representative sample in the United States, three allostatic load measures were strongly positively correlated with biological age in a sample of participants aged 30 or over (data collected 1988-1994; *n* = 9,942; *r* range = .49 to .65; Levine & Crimmins, 2014).

In a study analysing later NHANES data from women aged 20-44, scores from most allostatic load operationalisations were related to shorter telomere length in unadjusted analyses (data collected 1999-2002; *n* = 1,503; Ahrens et al., 2016). For the allostatic load operationalisation most strongly related to telomere length, 1 additional biomarker beyond the clinical cutoff was associated with a reduction in telomere length of 1.3% (Ahrens et al., 2016). However, the association was no longer statistically significant after adjustment for age and other covariates (Ahrens et al., 2016).

Papers studying allostatic load markers without calculating allostatic load summary scores have suggested allostatic load and telomere length may be correlated. In an early small-sample study, telomere length was strongly correlated with 12-hour urinary cortisol and two of four salivary cortisol measures (*r* range = -.64 to -.40), and these associations were also statistically significant in adjusted regression models (*N* = 23; age range = 51-79 y; Tomiyama et al., 2012).

In the Netherlands Study of Depression and Anxiety (*N* = 2,981 [1,737 with current depressive and/or anxiety disorder; 592 with remitted depressive and/or anxiety disorder; 652 with no current or remitted psychiatric disorder]; age range = 18-65 y), shorter telomere length was related to "inflammation, high awakening cortisol response, and increased heart rate [...], especially when they [were] dysregulated cumulatively" (Révész et al., 2014, p. 1422) in participants without any psychiatric disorder history as well as those with current or remitted depressive and/or anxiety disorders. In an adjusted model, it was found that compared to having 0 out of 4 biomarkers (the following specific biomarkers were chosen due to being independently related to telomere length: C-reactive protein, interleukin-6, cortisol awakening response area under the curve with respect to the increase, and heart rate) in the highest risk tertile, having 3 or 4 biomarkers in the highest risk tertile was associated with a reduction in telomere length of 157 base pairs, a reduction roughly equivalent to 8-11 years of ageing (Révész et al., 2014).

Also, in the Heart Scan Study, a subsample of the United Kingdom-based Whitehall II cohort, adjusted models found that men with short telomeres and high telomerase activity had allostatic responses to psychological stress tasks (n = 333; age range = 54-76 y; Zalli et al., 2014). For example, 40-45 minutes after a mental stress task, a man having above-median telomere length predicted a reduction in systolic blood pressure of ~25mmHg, while a man having below-median telomere length and above-median telomerase activity predicted a reduction of only ~17mmHg (Zalli et al., 2014). This result suggested that men with short telomeres and high telomerase activity "exhibited impaired poststress recovery" (Zalli et al., 2014, p. 4520).

To summarise, some research has suggested that allostatic load and telomere length may be related, but the study designs and samples used to explore this relationship have been too diverse for clear conclusions to be drawn. In particular, it is unclear how allostatic load summary measures and telomere length relate in older adults, as no study to date has used a sample comprised entirely of older adults to test this relationship.

4.1.6.2 Biological age, epigenetic age and telomere length

At age 38 in the New Zealand-representative Dunedin Study cohort, biological age was correlated weakly with epigenetic age (*r* range = .07 to .15), and telomere length was not statistically significantly correlated with biological age or epigenetic age (*r* range = -.02 to -.05; *N* = 1,037; Belsky et al., 2018).

In subsamples of the Germany-based ESTHER cohort study, epigenetic age was related to relative telomere length in sex-adjusted regression models (effect sizes not reported), but epigenetic age *acceleration* did not predict relative telomere length in any regression models (subsample *n*s = 969 and 851; subsample mean ages = 62 y and 63 y; Breitling et al., 2016).

4.1.7 The present study

It is important that we clarify how strongly these physiological weathering measures relate in older age so that we can determine to what extent they index the same aspects of stress and ageing processes. However, the strength of these relationships in older age is currently unclear. This is because previous physiological weathering studies have typically used data from only one time point, have tended to used mixed-age, younger and/or relatively small samples, and have never (to our knowledge) incorporated all four of the physiological weathering measures discussed here.

The present study will use data from Waves 1-4 of the Lothian Birth Cohort 1936 (LBC1936), at which participants were aged approximately 70, 73, 76 and 79. A previous LBC1936 study found small and non-statistically significant cross-sectional correlations between telomere length and epigenetic age at ages 70, 73 and 76 (Marioni et al., 2016). The present study will go beyond that and all other previous studies by modelling all four of the physiological weathering measures discussed here using four waves of data. This will enable novel tests of the associations between longitudinal changes in physiological weathering measures.

This study will attempt to answer the following research questions in relation to the four physiological weathering measures of interest (allostatic load, biological age, epigenetic age, and telomere length):

1. How strongly related are contemporaneous measures of physiological weathering at age 70?
2. What is the rate of longitudinal change in measures of physiological weathering between ages 70 and 79?
3. How strongly related are allostatic load, biological age, epigenetic age and telomere length at age 70 to changes in each of those measures between ages 70 and 79?
4. How strongly related are longitudinal changes in different measures of physiological weathering between ages 70 and 79?

Based on previous research (Belsky et al., 2018; Levine & Crimmins, 2014), it is plausible that the correlation between allostatic load and biological age will be large, because these two measures are calculated using some of the same biomarkers. The other correlation estimates could be expected to be very small-to-small (Belsky et al., 2018).

4.2 Methods

4.2.1 Measures

4.2.1.1 Biological age

Values of the following seven biomarkers were used: HbA1c; the forced expiratory volume in one second (FEV1); SBP; total cholesterol; CRP; creatinine; and albumin. The measurement of five of these biomarkers was described in relation to allostatic load measurement. Creatinine, an indicator of kidney function, was measured using blood samples. The FEV1, a measure of lung function, was tested three times per participant using a Micro Medical Spirometer. The highest value of the three was used in biological age calculation.

4.2.1.1.1 Biological age estimation. Following Belsky et al. (2018), we calculated biological age using the Klemera-Doubal method (Klemera & Doubal, 2006) and previously established weights (Levine, 2013). An R script for calculating biological age estimates using these weights was graciously provided by Morgan Levine.

4.2.1.2 Epigenetic age

4.2.1.2.1 DNA methylation. DNA extraction from whole blood samples and methylation typing was conducted at the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh. The methylation typing was of 485,512 probes. DNA samples were bisulphite converted and then hybridised to the Infinium HumanMethylation450 array using the Infinium HD Methylation protocol with Tecan robotics (Illumina). Internal controls were used to background-correct and normalise the raw intensity data. The R package minfi (Aryee et al., 2014) was used to generate methylation beta values. The array control probe signals were manually inspected for quality. Low quality samples were removed, including those with staining signal and/or inadequate hybridisation, bisulphite conversion, or nucleotide extension. Following the low call threshold recommended by Illumina, samples were removed if under 450,000 probes were detected at $p < .01$. Data from participants whose self-reported sex did not match the sex predicted by XY probes were removed. Then, X and Y and chromosome probes were removed. After quality control, there were 450,726 probes.

4.2.1.2.2 Epigenetic age estimation. We considered two measures of epigenetic age, as defined in Chen et al. (2016): intrinsic epigenetic age and extrinsic epigenetic age. Each of these epigenetic age measures are estimated by applying an established formula to data on DNA methylation from age-linked CpG markers. The resulting epigenetic age estimates are, like biological age estimates, strongly but not perfectly correlated with chronological age. The two epigenetic age measures used differ in the formula applied, the CpGs used, and the adjustments made based on blood cell data (for full descriptions, see Chen et al., 2016).

Intrinsic epigenetic age is based on Horvath's estimate of epigenetic age (calculated using 353 CpGs; Horvath, 2013). Horvath (2013) obtained epigenetic age calculation weights by fitting elastic net regression models with CpG data as independent variables and a function of age as the dependent variable. Note that because all participants in our sample were aged over 20, the function of age reduced to (age − 20) / 21 (Horvath, 2013, additional file 2). To calculate epigenetic age in new samples, regression prediction is used. Horvath's intercept and beta values are applied to participants' CpG data and the resulting values are multiplied by the inverse of the age function (Horvath, 2013, additional file 2). The intrinsic epigenetic age measure has additional adjustments for blood immune cell counts (plasmablasts, naive and exhausted CD8+ T Cells, CD4+ T cells, natural killer cells, monocytes, and granulocytes). This means it is cell-intrinsic and independent of much of the variation in blood cell composition. Intrinsic epigenetic age was calculated by adding the residuals from the regression of Horvath epigenetic age on the immune cell counts to the mean of Horvath epigenetic age.

Extrinsic epigenetic age takes a weighted average of Hannum's epigenetic age (calculated using 71 CpGs; Hannum et al., 2013), and three immune blood cell types. Note that the weighted average is computed using the Klemera-Doubal method (Klemera & Doubal, 2006), which we will use to calculate biomarker panel-based estimates of biological age. Similar to Horvath (2013), Hannum et al. (2013) used elastic net regression to find beta weights for CpGs related to chronological age. However, there are several differences between their approaches. Notably, Hannum et al. (2013) used covariates in the selection regressions, had a different CpG selection procedure, and did not use the function of age above as the dependent variable. Notably, Hannum et al. (2013) used covariates in the selection regressions, had a different CpG selection procedure, and did not use the function of age above as the dependent variable. Extrinsic epigenetic age was calculated using the Horvath online calculator (Horvath, n.d.).

4.2.1.3 Telomere length

Telomere length was measured with quantitative real-time polymerase chain reaction (Martin-Ruiz et al., 2004). Polymerase chain reactions were conducted using an Applied Biosystems (Pleasonton, CA, USA) 7900HT Fast Real Time PCR System. To correct for variation between plates, four internal control samples were run within each plate.

4.2.2 Statistical analysis

4.2.2.1 Correlations

Firstly, we computed a correlation matrix of the estimates for all physiological weathering measures at all waves. This enabled us to assess the lagged correlations, or the autocorrelations, between repeated measures of the same physiological weathering variable. The lag 1 correlation refers to the correlation between measurements one wave apart, the lag 2 correlation to measurements two waves apart, and the lag 3 correlation to measurements three waves apart. For example, because we used data from four waves, the lag 2 correlations for a physiological weathering measure are the Wave 1-Wave 3 correlation and the Wave 2-Wave 4 correlation. The correlation matrix also enabled us to assess the average correlation strength between different physiological weathering measures measured at the same wave.

4.2.2.2 Latent growth curve models

In this study, we modelled intercept and slope growth factors for allostatic load, biological age, epigenetic age, and telomere length. We fitted our MAR and NMAR parallel process latent growth curve models twice, once with extrinsic epigenetic age as the epigenetic age measure and once with intrinsic epigenetic age in its place.

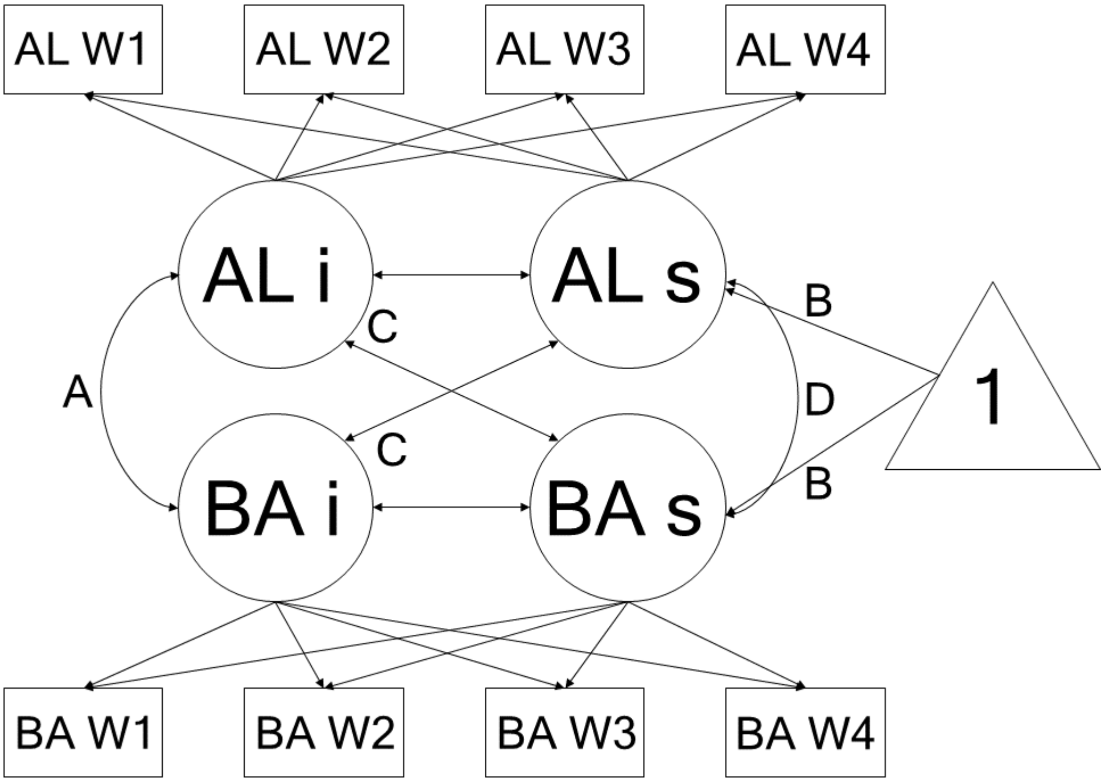

Figure 4.1: Physiological weathering latent growth curve model. A-D indicate labelled paths described in the text. AL = allostatic load. BA = biological age. i = intercept. s = slope. W = wave.

Figure 4.1 shows the core structure of the dual change latent growth curve models for two of the physiological weathering variables: allostatic load and biological age. Only two variables are displayed for ease of presentation, but our parallel process latent growth curve models in this study all included four pairs of physiological weathering growth factors. This diagram has a similar layout to the general specification in 3.1.

Research question 1, on the contemporaneous relationships at age 70, was tested by the correlations between the intercept growth factors; for example, path A in Figure 4.1 tests the age 70 relationship between allostatic load and biological age. Research question 2, on the longitudinal change in measures of physiological weathering between ages 70 and 79, was tested by the means and variances of the slope growth factors (see the paths labelled B for the slope means). Research question 3, on the relationships between age 70 levels and age 70-79 changes, was tested by the intercept-slope correlations (see the paths labelled C). Research question 4, on the relationships between age 70-79 changes in different measures, was tested by the slope-slope correlations (see path D).

We preregistered that if the correlations not involving epigenetic age were similar across these model structures, we would focus our interpretation on the results from the models with extrinsic epigenetic age. However, there were model fitting problems with extrinsic epigenetic age and the models with intrinsic epigenetic age showed markedly better fit. Hence, we focused our interpretation on the models with intrinsic epigenetic age as the epigenetic age measure.

4.2.2.2.1 Additional covariates. DNA methylation assessment for Waves 2 and 3 was conducted in two batches, with approximately half of the samples from each wave assessed in the first batch and the remaining samples assessed in the second batch. Hence, to statistically control for DNA methylation testing batch variation, Wave 2 and Wave 3 batch dummy variables were included in our latent growth curve models as predictors of Wave 2 and Wave 3 epigenetic age respectively, for both extrinsic and intrinsic epigenetic age. The batch dummy variables were coded as -0.5 for participants whose samples were assessed in the first batch and 0.5 for participants whose samples were assessed in the second batch. We chose to use the -0.5/0.5 coding rather than our preregistered 0/1 coding so that the 0 point was in between the two batch codes, meaning that the intercepts in the model were not biased by batch effects.

4.2.2.2.2 Sensitivity analysis. All models that included epigenetic age estimates were repeated after estimates were adjusted for three DNA methylation assessment variables: the plate used, the array used, and the position of samples on the array. The adjusted epigenetic age estimates were residuals from linear regressions where the independent variables were the plate used, the array used, and the position of samples on the array. These residuals were re-centred on the previously observed mean estimate before inclusion in the sensitivity analysis models.

4.3 Results

Table 4.1: Descriptive statistics for age and physiological weathering measures

	n	Mean	SD	Skew	Kurtosis
Age W1	1091	69.53	0.83	-0.05	-0.88
Age W2	866	72.49	0.71	-0.02	-0.84
Age W3	697	76.25	0.68	-0.05	-0.83
Age W4	550	79.32	0.62	0.06	-0.76
Allostatic load W1	1090	0.69	0.24	1.07	2.12
Allostatic load W2	866	0.71	0.24	1.17	3.17
Allostatic load W3	696	0.85	0.25	0.71	1.16
Allostatic load W4	549	0.87	0.26	0.62	0.48
Biological age W1	1041	68.83	3.50	0.22	-0.06
Biological age W2	815	71.02	3.57	0.23	0.22
Biological age W3	609	75.53	3.55	0.14	-0.04
Biological age W4	485	77.32	3.51	0.47	0.89
Extrinsic epigenetic age W1	906	76.00	7.15	-0.08	1.08
Extrinsic epigenetic age W2	801	77.52	7.00	0.16	1.45
Extrinsic epigenetic age W3	619	83.18	6.88	0.64	2.84

Table 4.1: Descriptive statistics for age and physiological weathering measures *(continued)*

	n	Mean	SD	Skew	Kurtosis
Extrinsic epigenetic age W4	507	88.51	6.40	1.03	2.60
Intrinsic epigenetic age W1	906	68.57	8.17	0.24	2.37
Intrinsic epigenetic age W2	801	72.72	8.14	1.13	7.95
Intrinsic epigenetic age W3	619	77.47	7.97	1.76	13.11
Intrinsic epigenetic age W4	507	81.09	7.17	0.61	0.71
Telomere length W1	1070	4200.51	559.67	1.00	2.65
Telomere length W2	844	3966.30	737.84	1.55	7.33
Telomere length W3	691	3738.69	686.06	0.05	2.81

Note.

Telomere length is in base pairs.

Table 4.1 contains descriptive statistics for the model variables. Allostatic load showed slight mean increases between Waves 1 and 2 as well as between Waves 3 and 4, but a much sharper increase was observed between Waves 2 and 3. Similarly, the increase in biological age means between Waves 2 and 3 was over double the Wave 1-2 and Wave 3-4 increases. The changes in the other physiological weathering means were more consistent across waves, with the exception of the small mean increase in extrinsic epigenetic age between Waves 1 and 2. The three age-linked physiological weathering display different patterns of relations to chronological age. Mean biological age was slightly lower than the mean chronological age at all waves. Mean extrinsic epigenetic age was notably higher than mean chronological age at all waves. Mean intrinsic epigenetic age was slightly lower than mean chronological age at Wave 1, but slightly higher than mean chronological age at Waves 2-4.

Across all waves, the mean age acceleration was -1.14 years for biological age, 0.32 years for intrinsic epigenetic age and 6.63 years for extrinsic epigenetic age.

There was considerably more spread in epigenetic age estimates than biological age estimates. The epigenetic age standard deviations were typically around twice the biological age standard deviation, with intrinsic epigenetic age consistently being the measure with the widest spread.

Table 4.2: Physiological weathering measure correlations

	AL W1	AL W2	AL W3	AL W4	BA W1	BA W2	BA W3	BA W4	EEA W1	EEA W2	EEA W3	EEA W4	IEA W1	IEA W2	IEA W3	IEA W4	TL W1	TL W2	TL W3
AL W1		.03	.04	.04	.03	.03	.04	.05	.03	.04	.04	.04	.03	.04	.04	.04	.03	.03	.04
AL W2	.45		.03	.04	.03	.03	.04	.05	.04	.04	.04	.04	.04	.04	.04	.04	.03	.03	.04
AL W3	.31	.46		.04	.04	.04	.04	.05	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
AL W4	.27	.40	.47		.04	.04	.05	.05	.05	.04	.05	.04	.05	.04	.05	.04	.04	.04	.04
BA W1	.19	.10	.07	.03		.03	.03	.04	.03	.04	.04	.05	.03	.04	.04	.05	.03	.04	.04
BA W2	.10	.13	.09	.03	.70		.03	.04	.04	.04	.04	.05	.04	.04	.04	.05	.04	.04	.04
BA W3	.04	.05	.14	.01	.58	.66		.04	.04	.04	.04	.05	.04	.04	.04	.05	.04	.04	.04
BA W4	−.03	.05	.07	.05	.56	.59	.66		.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05
EEA W1	.08	.00	−.02	−.03	.03	.07	−.01	.11		.03	.04	.04	.03	.04	.04	.05	.03	.04	.04
EEA W2	.04	.02	.02	−.02	.01	.05	.04	.07	.52		.03	.04	.04	.03	.04	.04	.04	.04	.04
EEA W3	−.01	−.01	−.02	−.06	.04	.00	.06	.08	.56	.72		.03	.04	.04	.04	.05	.04	.04	.04
EEA W4	−.01	.01	−.06	.00	.02	.06	.09	.08	.62	.62	.73		.05	.04	.04	.04	.04	.04	.05
IEA W1	.05	−.02	.00	−.02	.04	.08	.00	.04	.38	.22	.22	.24		.03	.04	.04	.03	.04	.04
IEA W2	.05	.01	−.03	.00	.02	.03	−.03	.01	.21	.38	.34	.31	.59		.03	.03	.04	.04	.04
IEA W3	.02	.00	−.03	−.05	.00	−.04	.00	.01	.18	.31	.41	.32	.61	.76		.03	.04	.04	.04
IEA W4	−.02	.01	−.03	−.05	.05	.06	.11	.06	.21	.27	.28	.38	.57	.68	.77		.04	.04	.05
TL W1	−.04	.07	.05	.06	−.05	−.03	.01	.04	.05	.04	−.05	.01	.00	−.01	−.08	.04		.03	.03

Table 4.2: Physiological weathering measure correlations *(continued)*

	AL W1	AL W2	AL W3	AL W4	BA W1	BA W2	BA W3	BA W4	EEA W1	EEA W2	EEA W3	EEA W4	IEA W1	IEA W2	IEA W3	IEA W4	TL W1	TL W2	TL W3
TL W2	-.02	.06	.07	.04	-.04	-.02	-.03	.05	.03	-.02	-.07	.02	.02	-.03	-.03	.09	.52		.02
TL W3	.06	.09	.10	.05	.00	.00	-.03	.04	-.03	-.08	-.13	-.05	-.03	-.03	-.02	.03	.48	.79	

Note.

Lower triangle contains pairwise Pearson's correlations (n range = 286-1070). Upper triangle contains standard errors for the correlations. AL = allostatic load. BA = biological age. EEA = extrinsic epigenetic age. IEA = intrinsic epigenetic age. TL = telomere length. W = Wave.

Table 4.2 contains correlations for the physiological weathering measures. The mean lag 1 correlations for each measure were: .46 for allostatic load; .68 for biological age; .65 for extrinsic epigenetic age; .71 for intrinsic epigenetic age; and .65 for telomere length. The mean lag 2 correlations for each measure were: .36 for allostatic load; .58 for biological age; .59 for extrinsic epigenetic age; .65 for intrinsic epigenetic age; and .48 for telomere length, for which only one lag 2 correlation was available. The lag 3 (Wave 1-Wave 4) correlations for each measure were: .27 for allostatic load; .56 for biological age; .62 for extrinsic epigenetic age; and .57 for intrinsic epigenetic age. In summary, there were strong longitudinal correlations for all physiological weathering measures except allostatic load, which had the weakest lag 1 correlation and showed the most marked decreases when the correlation lag was increased. The only measure that did not show consistent correlation decreases as the lag increased was extrinsic epigenetic age, which had the second-weakest lag 1 correlation but the strongest lag 3 correlation.

The mean absolute correlation between different contemporaneously measured physiological weathering measures, excluding the correlations between the two epigenetic age measures, was .06 at Wave 1, .04 at Wave 2, .06 at Wave 3, and .05 at Wave 4. After further excluding the correlations between allostatic load and biological age, the mean absolute correlation was .04 at Wave 1, .03 at Wave 2, .05 at Wave 3, and .05 at Wave 4.

4.3.1 One process latent growth curve models

Table 4.3: Unconditional one process MAR model results: Growth factor means and variances

	Estimate	SE	p-value
AL i mean	0.682	0.007	< 0.001
AL i variance	0.029	0.004	< 0.001
AL s mean	0.021	0.001	< 0.001
AL s variance	0.0003	0.0001	< 0.001
BA i mean	68.76	0.11	< 0.001
BA i variance	9.29	0.56	< 0.001
BA s mean	0.95	0.01	< 0.001
BA s variance	0.05	0.01	< 0.001
EEA i mean	74.80	0.24	< 0.001
EEA i variance	29.60	2.33	< 0.001
EEA s mean	1.37	0.03	< 0.001
IEA i mean	68.82	0.24	< 0.001
IEA i variance	41.04	5.46	< 0.001
IEA s mean	1.27	0.03	< 0.001
IEA s variance	0.05	0.04	0.274
TL i mean	4.192	0.017	< 0.001
TL i variance	0.185	0.026	< 0.001
TL s mean	-0.068	0.003	< 0.001
TL s variance	0.005	0.001	< 0.001

Note.

AL = allostatic load. BA = biological age. EEA = extrinsic epigenetic age. i = intercept [growth factor]. IEA = intrinsic epigenetic age. s = slope [growth factor]. TL = telomere length.

Table 4.3 contains growth factor mean and variance estimates from latent growth curve models that did not include covariates. One such model was fitted for each physiological weathering variable. Here, the intercept growth factor means represent the average levels at baseline and the slope growth factor means represent the average changes per year. The intercept and slope variances represent the variability in baseline levels and trajectories respectively.

Table 4.9 contains model fit statistics for all models fitted. The fit of the unconditional models helps to tell us whether a linear trajectory describes the progression of these physiological weathering measures well. Good fit was observed for all allostatic load and intrinsic epigenetic age models as well as for the biological age and telomere length models with covariates. Poor fit was observed for all extrinsic epigenetic age models. As noted in the table, the extrinsic epigenetic age slope and Wave 3 telomere length variances were initially estimated as negative and so they (and covariances involving the extrinsic epigenetic age slope) were fixed at zero in these and subsequent models.

The slope means and variances in the one process models address research question 2, which concerns the rates of change over time in these physiological weathering measures. The allostatic load slope mean is positive, which means that the average participant's biomarker levels became more extreme over the course of the study. The increase was relatively modest: the average participant's average biomarker moved 0.021 *SD* further away from the Wave 1 mean of the biomarker each year, which accumulates to an average 0.21 *SD* deviation over a 10-year period. There was, though, notable variation in individual slopes.

Due to model fitting problems, telomere length was rescaled into kilobase pairs. The telomere length slope mean tells us that the average loss was estimated at 68 base pairs per year, and the slope variance suggests there was substantial variability in the rates of loss.

Biological age tracked closely with chronological age; each chronological year, the average participant's biological age increased around 1 year, and there was substantial variation in the biological age trajectories. However, according to intrinsic and extrinsic epigenetic age, participants were ageing much faster. According to intrinsic epigenetic age, the average participant aged an additional 99 days each year, while according to extrinsic epigenetic age, ageing was even faster, with the average participant ageing an additional 135 days each year. The models did not confirm that there was substantial variability in epigenetic age trajectories: the extrinsic epigenetic age slope variance had to be fixed to zero to successfully fit the model, and the intrinsic epigenetic age variance had a large standard error relative to its size.

Table 4.4: Conditional one process MAR model results: Growth factor estimates and regressions on sex

	Estimate	SE	p-value
AL i mean	0.683	0.010	< 0.001
AL i variance	0.028	0.003	< 0.001
AL i on male regression	-0.003	0.014	0.852
AL s mean	0.0214	0.0016	< 0.001
AL s variance	0.0003	0.0001	< 0.001
AL s on male regression	-0.0005	0.0022	0.807
BA i mean	69.02	0.15	< 0.001
BA i variance	8.57	0.55	< 0.001
BA i on male regression	-0.56	0.20	0.006
BA s mean	0.85	0.02	< 0.001
BA s variance	0.03	0.01	0.002
BA s on male regression	0.21	0.03	< 0.001
EEA i mean	73.01	0.31	< 0.001
EEA i variance	26.22	2.23	< 0.001
EEA i on male regression	3.45	0.4	< 0.001
EEA s mean	1.32	0.03	< 0.001
EEA s on male regression	0.09	0.05	0.055
IEA i mean	68.37	0.36	< 0.001
IEA i variance	39.42	5.37	< 0.001
IEA i on male regression	0.88	0.49	0.070
IEA s mean	1.24	0.04	< 0.001
IEA s variance	0.04	0.04	0.355

Table 4.4: Conditional one process MAR model results: Growth factor estimates and regressions on sex *(continued)*

	Estimate	SE	p-value
IEA s on male regression	0.04	0.05	0.386
TL i mean	4.072	0.022	< 0.001
TL i variance	0.170	0.024	< 0.001
TL i on male regression	0.237	0.033	< 0.001
TL s mean	-0.052	0.004	< 0.001
TL s variance	0.005	0.001	< 0.001
TL s on male regression	-0.031	0.006	< 0.001

Note.
AL = allostatic load. BA = biological age. EEA = extrinsic epigenetic age. i = intercept [growth factor]. IEA = intrinsic epigenetic age. s = slope [growth factor]. TL = telomere length.

Table 4.4 contains results from conditional latent growth curve models which included statistical control for within-wave age differences and sex effects on the growth factors. One such model was fitted for each physiological weathering variable. The extrinsic and intrinsic epigenetic age models also added statistical control for DNA methylation batch effects at Waves 2 and 3. Here, the growth factor means are the estimates for a female who was the mean age of assessment at each wave. To obtain the growth factor mean estimates for males, add the estimate for the relevant regression on male.

Biological age and extrinsic epigenetic age gave different answers as to which sex had aged faster at baseline. On average, males were almost 3.5 years older than females in extrinsic epigenetic age, while for biological age, females were slightly over 7 months older than males. However, the biological age slope indicates that males aged faster than females during the course of the study. Every 12 months, the average female increased only around 10 months in biological age, while the average male experienced biological age acceleration as they increased around 13 months in biological age.

The pattern of level and change for telomere length was similar to the pattern for biological age. Males had telomeres 237 base pairs longer at baseline, but they lost 83 base pairs each year, compared to 52 base pairs for females.

For the other growth factors, there were no notable sex differences.

Table 4.5: Conditional one process NMAR model results: Growth factor estimates and regressions on sex

	Estimate	SE	p-value
AL i mean	0.655	0.012	< 0.001
AL i variance	0.028	0.003	< 0.001
AL i on male regression	-0.001	0.014	0.927
AL s mean	0.0235	0.0017	< 0.001
AL s variance	0.0003	0.0001	< 0.001
AL s on male regression	-0.0008	0.0022	0.721
BA i mean	68.36	0.18	< 0.001
BA i variance	8.13	0.52	< 0.001
BA i on male regression	-0.55	0.20	0.006
BA s mean	0.87	0.02	< 0.001
BA s variance	0.03	0.01	0.002
BA s on male regression	0.20	0.03	< 0.001
EEA i mean	72.19	0.35	< 0.001
EEA i variance	26.07	2.2	< 0.001
EEA i on male regression	3.52	0.40	< 0.001
EEA s mean	1.39	0.03	< 0.001
EEA s on male regression	0.08	0.05	0.106
IEA i mean	68.31	0.39	< 0.001
IEA i variance	39.39	5.31	< 0.001
IEA i on male regression	0.86	0.49	0.083
IEA s mean	1.23	0.04	< 0.001

Table 4.5: Conditional one process NMAR model results: Growth factor estimates and regressions on sex *(continued)*

	Estimate	SE	p-value
IEA s variance	0.04	0.04	0.359
IEA s on male regression	0.04	0.05	0.464
TL i mean	4.069	0.026	< 0.001
TL i variance	0.170	0.024	< 0.001
TL i on male regression	0.238	0.033	< 0.001
TL s mean	-0.051	0.004	< 0.001
TL s variance	0.005	0.001	< 0.001
TL s on male regression	-0.031	0.006	< 0.001

Note.
AL = allostatic load. BA = biological age. EEA = extrinsic epigenetic age. i = intercept [growth factor]. IEA = intrinsic epigenetic age. s = slope [growth factor]. TL = telomere length.

Table 4.6: Conditional one process NMAR model results: Pattern-mixture model results

	Beta	SE	p-value
AL i on d2 regression	0.064	0.020	0.001
AL i on d3 regression	0.086	0.022	< 0.001
AL i on d4 regression	0.028	0.020	0.176
AL s on d2/d3 regression	-0.0137	0.0067	0.040
AL s on d4 regression	-0.0009	0.0037	0.800
BA i on d2 regression	1.48	0.27	< 0.001
BA i on d3 regression	1.43	0.31	< 0.001
BA i on d4 regression	1.11	0.30	< 0.001
BA s on d2/d3 regression	-0.17	0.07	0.018
BA s on d4 regression	0.06	0.04	0.140
EEA i on d2 regression	2.48	0.56	< 0.001
EEA i on d3 regression	1.8	0.59	0.002
EEA i on d4 regression	1.58	0.66	0.016
EEA s on d2/d3 regression	-0.69	0.16	< 0.001
EEA s on d4 regression	-0.16	0.09	0.065
IEA i on d2 regression	-0.33	0.61	0.588
IEA i on d3 regression	0.48	0.73	0.508
IEA i on d4 regression	0.22	0.88	0.801
IEA s on d2/d3 regression	-0.14	0.17	0.405
IEA s on d4 regression	0.19	0.09	0.039
TL i on d2 regression	0.005	0.042	0.901
TL i on d3 regression	0.028	0.047	0.549
TL s on d2/d3 regression	-0.018	0.018	0.326

Note.

AL = allostatic load. BA = biological age. d2-d4 = binary dummy variables coded as 1 if a participant dropped out at the corresponding wave (2-4) and 0 otherwise. EEA = extrinsic epigenetic age. i = intercept [growth factor]. IEA = intrinsic epigenetic age. s = slope [growth factor]. TL = telomere length.

Tables 4.5 and 4.6 contain the results from conditional models that added dummy variables for dropout (d2-d4 for Waves 2-4), which were coded as 1 if the participant dropped out at that wave and 0 if they did not. One such model was fitted for each physiological weathering variable. Table 4.5 contains estimates for parameters that were also in the previous set of models, while Table 4.6 contains estimates for the dummy variable regressions that were added to the models. In these models, the growth factor means are the estimates for a female who did not drop out. To obtain estimates for an average male and/or a participant who dropped out at a given wave, add the relevant regression estimate(s).

Those who dropped out at Wave 2 or 3 had higher baseline allostatic load, biological age and extrinsic epigenetic age. However, they had flatter slopes (i.e., slower ageing) for biological age and extrinsic epigenetic age.

Those who dropped out at Wave 4 also had higher baseline biological age and extrinsic epigenetic age, as well as a flatter extrinsic epigenetic age slope. However, they also had a steeper slope for intrinsic epigenetic age.

4.3.2 Parallel process latent growth curve models

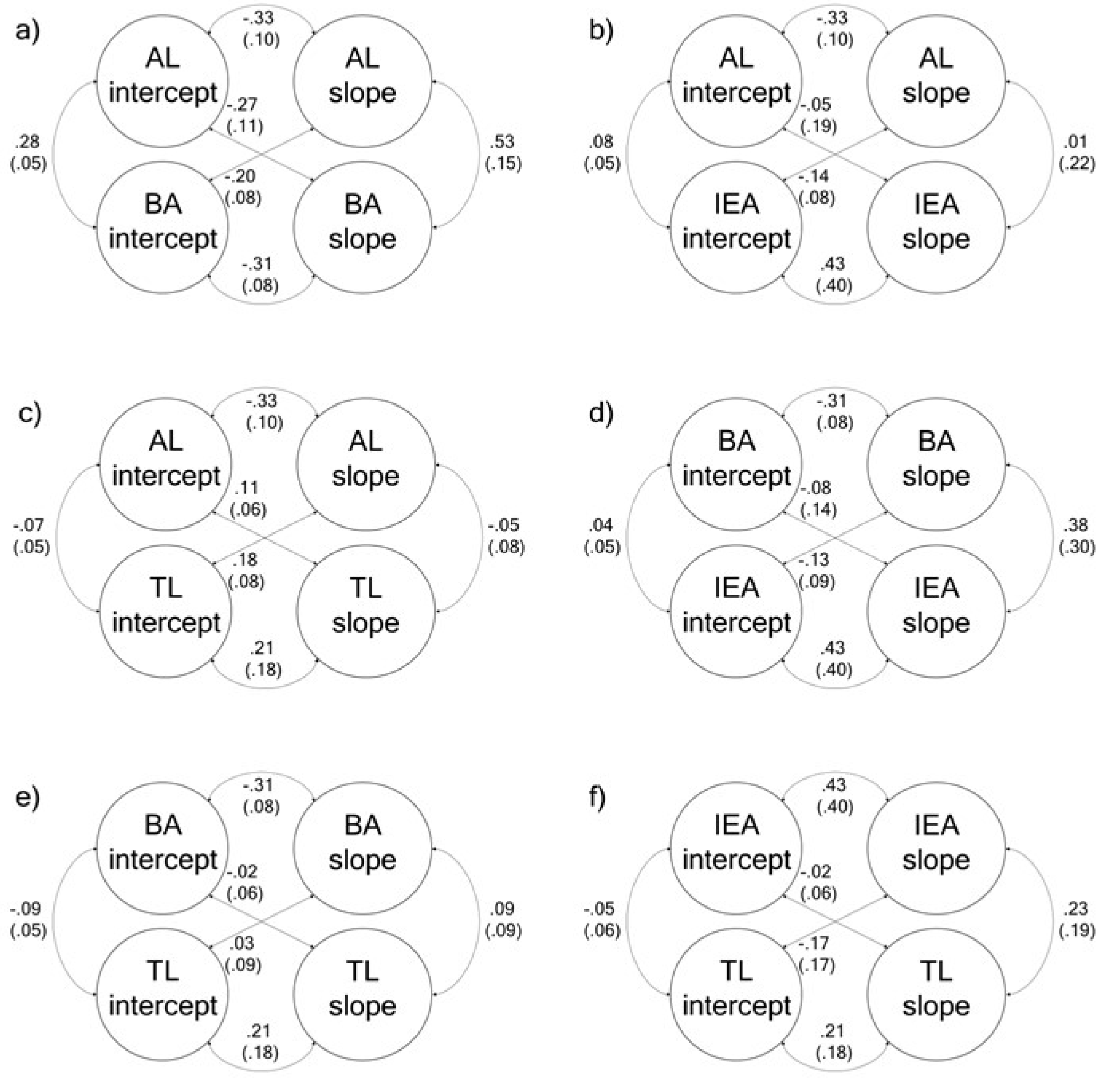

Figure 4.2: Parallel process model results: Growth factor correlations from the MAR model with IEA. Results are from one model; pairings are displayed separately for clarity. Correlation estimates between the growth factors are shown, with the associated standard errors in parentheses. AL = allostatic load. BA = biological age. IEA = intrinsic epigenetic age. TL = telomere length.

We fitted parallel process latent growth curve models to test the associations between levels and trajectories in allostatic load, biological age, epigenetic age, and telomere length. We fitted separate models using the two epigenetic age measures. The MAR and NMAR models with intrinsic epigenetic age fit better than the models using extrinsic epigenetic age (see Table 4.9), so we chose to focus our interpretation on the models with intrinsic epigenetic age as the epigenetic age measure. Figure 4.2 contains the growth factor correlations from the MAR model with intrinsic epigenetic age. In these models, the intercept-intercept, intercept-slope and slope-slope correlations addressed research questions 1, 3 and 4 respectively.

The two most closely related physiological weathering measures were allostatic load and biological age. Their intercept-intercept correlation indicated that at baseline, higher allostatic load was moderately associated with older biological age (*r* [*SE*] = .28 [.05]). The slope-slope correlation between the measures showed they tracked closely over time: increases in allostatic load were strongly related to accelerated biological ageing (*r* [*SE*] = .53 [.15]).

Higher allostatic load and older biological age at baseline were moderately associated with smaller increases in allostatic load and slower biological ageing, as indicated by the four intercept-slope correlations.

At baseline, longer telomere length was associated with older biological age, with a small effect size (*r* [*SE*] = -.09 [.05]). Longer telomere length at baseline also had a small-to-moderate association with a greater subsequent increase in allostatic load.

There were no correlations detected involving intrinsic epigenetic age. Some estimates for the correlations with the intrinsic epigenetic age slope were moderate in size, but standard errors for these correlations were large, so more data would be required to powerfully test these associations.

Table 4.7: Parallel process model results: Growth factor correlations from the MAR model with EEA

	r	SE	p-value
AL i-AL s correlation	-0.33	0.10	0.001
BA i-BA s correlation	-0.31	0.08	< 0.001
TL i-TL s correlation	0.21	0.18	0.230
AL i-BA i correlation	0.28	0.05	< 0.001
AL i-EEA i correlation	0.10	0.05	0.036
AL i-TL i correlation	-0.07	0.05	0.202
BA i-EEA i correlation	0.07	0.04	0.107
BA i-TL i correlation	-0.09	0.05	0.050
EEA i-TL i correlation	-0.08	0.07	0.226
AL s-BA s correlation	0.53	0.15	< 0.001
AL s-TL s correlation	-0.05	0.08	0.541
BA s-TL s correlation	0.09	0.09	0.321
AL i-BA s correlation	-0.27	0.11	0.018
AL i-TL s correlation	0.11	0.06	0.092
BA i-AL s correlation	-0.20	0.08	0.011
BA i-TL s correlation	-0.03	0.06	0.645
EEA i-AL s correlation	-0.11	0.07	0.135
EEA i-BA s correlation	-0.10	0.08	0.215
EEA i-TL s correlation	-0.08	0.07	0.252
TL i-AL s correlation	0.18	0.08	0.037
TL i-BA s correlation	0.03	0.09	0.739

Note.
AL = allostatic load. BA = biological age. i = intercept [growth factor]. EEA = extrinsic epigenetic age. s = slope [growth factor]. TL = telomere length.

Table 4.7 contains results from the MAR model with extrinsic epigenetic age, rather than intrinsic epigenetic age, as the epigenetic age measure. Note that because the extrinsic epigenetic age slope had to be fixed at zero, all covariances involving the extrinsic epigenetic age slope were also fixed at zero. At baseline, older extrinsic epigenetic age was associated with higher allostatic load, with a small effect size (*r* [*SE*] = .10 [.05]).

4.3.2.1 Adding dummy variables for dropout

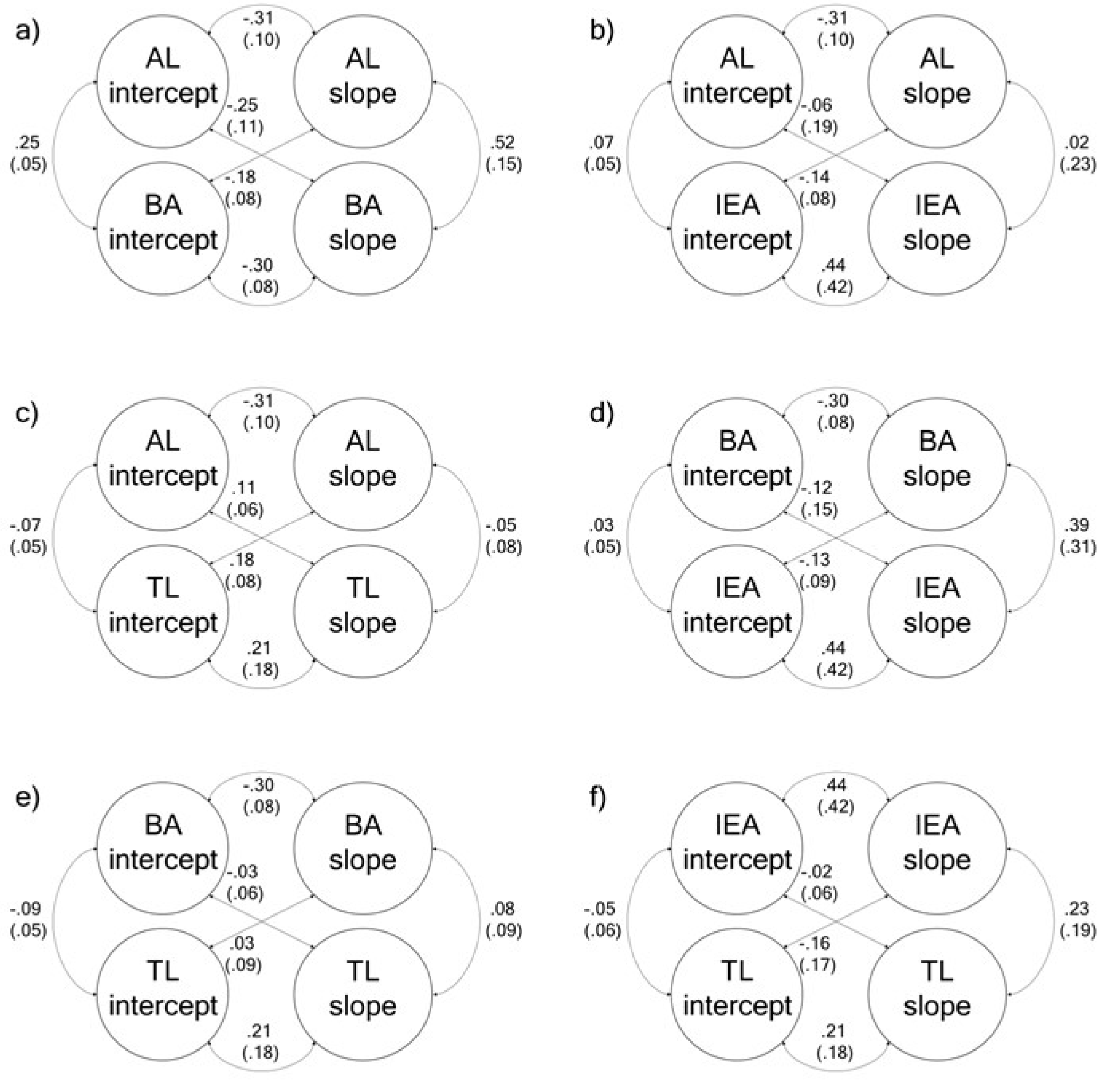

Figure 4.3: Parallel process model results: Growth factor correlations from the NMAR model with IEA. Results are from one model; pairings are displayed separately for clarity. Correlation estimates between the growth factors are shown, with the associated standard errors in parentheses. AL = allostatic load. BA = biological age. IEA = intrinsic epigenetic age. TL = telomere length.

Table 4.8: Parallel process model results: Growth factor correlations from the NMAR model with EEA

	r	SE	p-value
AL i-AL s correlation	-0.31	0.10	0.002
BA i-BA s correlation	-0.30	0.09	< 0.001
TL i-TL s correlation	0.21	0.18	0.230
AL i-BA i correlation	0.25	0.05	< 0.001
AL i-EEA i correlation	0.08	0.05	0.078
AL i-TL i correlation	-0.07	0.05	0.198
BA i-EEA i correlation	0.04	0.04	0.307
BA i-TL i correlation	-0.09	0.05	0.053
EEA i-TL i correlation	-0.08	0.07	0.235
AL s-BA s correlation	0.52	0.15	0.001
AL s-TL s correlation	-0.05	0.08	0.530
BA s-TL s correlation	0.08	0.09	0.357
AL i-BA s correlation	-0.25	0.11	0.028
AL i-TL s correlation	0.11	0.06	0.088
BA i-AL s correlation	-0.18	0.08	0.025
BA i-TL s correlation	-0.03	0.06	0.553
EEA i-AL s correlation	-0.10	0.07	0.183
EEA i-BA s correlation	-0.09	0.08	0.274
EEA i-TL s correlation	-0.08	0.07	0.230
TL i-AL s correlation	0.18	0.08	0.035
TL i-BA s correlation	0.03	0.09	0.716

Note.
AL = allostatic load. BA = biological age. i = intercept [growth factor]. EEA = extrinsic epigenetic age. s = slope [growth factor]. TL = telomere length.

Growth factor correlations from the parallel process models that assumed missing data were not missing at random are in Figure 4.3 (for the model with IEA) and Table 4.8 (for the model with EEA). Results in these models were similar to the results in the missing at random models. Two baseline (intercept-intercept) correlations that were marginally statistically significant in the missing at random models were no longer statistically significant using an alpha of .05. The first was the biological age-telomere length correlation (*r* [*SE*] = -.09 [.05]; *p* = .052). The interpretation of this parameter did not change based on the not missing at random models, as the correlation was near-identical in size and had similar *p*-values in all models. The second was the allostatic load-extrinsic epigenetic age correlation, which was attenuated by around 20% (*r* [*SE*] = .08 [.04]; *p* = .078). Some other correlations were also slightly attenuated, but not to the extent that the changes affected broad conclusions about the relationships. On the whole, the results of these models suggested that the contemporaneous longitudinal relationships between physiological weathering measures were not substantially biased by the measures' relationships with dropout.

We also fitted sensitivity analysis models using epigenetic age estimates pre-adjusted for three DNA methylation assessment variables. Results in the sensitivity analysis models were near-identical, which showed that these methodological variables had no notable impact on our results.

4.3.3 Model fit

Table 4.9: Model fit statistics for physiological weathering models

Model	CFI	RMSEA	SRMR	Notes
AL MAR unconditional	0.904	0.098	0.054	
AL MAR conditional	0.905	0.051	0.028	
AL NMAR conditional	0.901	0.045	0.037	
BA MAR unconditional	0.914	0.141	0.065	Despite poor fit, this model produced reasonable estimates.
BA MAR conditional	0.917	0.074	0.038	
BA NMAR conditional	0.927	0.060	0.042	
EEA MAR unconditional	0.848	0.148	0.096	EEA slope variance was fixed to zero as it was initially estimated as negative. Covariances involving the EEA slope were also fixed to zero. These constraints was retained in subsequent models.
EEA MAR conditional	0.854	0.091	0.054	
EEA NMAR conditional	0.873	0.074	0.065	
IEA MAR unconditional	0.998	0.020	0.029	
IEA MAR conditional	1.000	0.000	0.017	
IEA NMAR conditional	0.998	0.010	0.028	
TL MAR unconditional	0.948	0.147	0.078	Wave 3 telomere length variance was fixed at zero as it was initially estimated as negative. This constraint was retained in subsequent models.
TL MAR conditional	0.945	0.074	0.042	

Table 4.9: Model fit statistics for physiological weathering models *(continued)*

Model	CFI	RMSEA	SRMR	Notes
TL NMAR conditional	0.945	0.064	0.036	
Parallel processes MAR with IEA	0.947	0.041	0.032	BA and IEA values were divided by 20 to avoid model fitting issues caused by variances being of very different sizes. This transformation was also used in subsequent parallel process models for BA, IEA and EEA.
Parallel processes NMAR with IEA	0.949	0.037	0.039	
Parallel processes MAR with EEA	0.900	0.053	0.041	
Parallel processes NMAR with EEA	0.909	0.047	0.049	

Note.

For CFI and RMSEA, robust variants were used. AL = allostatic load. BA = biological age. CFI = comparative fit index. EEA = extrinsic epigenetic age. IEA = intrinsic epigenetic age. RMSEA = root mean square error of approximation. SRMR = square root mean residual. TL = telomere length.

Table 4.9 contains model fit statistics for all models fitted. Across the one process models, *RMSEA* and *SRMR* tended to change more than *CFI* when covariates were added. All intrinsic epigenetic age models had excellent fit, while all extrinsic epigenetic age models had poor fit. Unconditional models for both biological age and telomere length had poor fit according to *RMSEA*, but after covariates were added to these models they had good fit. The unconditional allostatic load model had borderline satisfactory fit, while conditional allostatic load models had good fit according to *RMSEA* and *SRMR* but still had borderline satisfactory fit according to the *CFI*.

The better fitting parallel process models were those with intrinsic epigenetic age. The parallel process models with intrinsic epigenetic age had good fit, while the parallel process models with extrinsic epigenetic age had good fit according to *RMSEA* and *SRMR* but only borderline satisfactory fit according to the *CFI*.

4.4 Discussion

Allostatic load, biological age, telomere length and epigenetic age measures are all considered to index physiological weathering caused by stress and/or ageing. All have been linked to important health outcomes, including mortality. It is important, then, to understand whether these measures are closely related and offer the same information about physiological weathering, or only weakly linked, in which case they mostly provide us with different information. Our study tested the relationships between these measures. It went beyond previous related research by having four waves of physiological weathering data from an age-homogeneous older age sample. This enabled us to model associated change in the measures during a particular period in later life: between ages 70 and 79. We found that these measures differ in their distributions, their relationships with sex, and their relationships with dropout. The two most closely linked measures, both contemporaneously and longitudinally, were allostatic load and biological age. Typically, however, estimates of the contemporaneous and longitudinal associations between these measures were strikingly small, which suggests that they tap different aspects of stress- and age-related physiological weathering.

4.4.1 Physiological weathering distributions and longitudinal changes

All physiological weathering measures tended to change in the expected directions as the cohort aged. Changes were not homogeneous, though; there was notable variability in individual trajectories for all measures except intrinsic epigenetic age.

Interestingly, the three measures designed to track with age - biological age, extrinsic epigenetic age, and intrinsic epigenetic age - related differently to chronological age. Biological age tracked chronological age the most closely. The average female experienced slightly slowed ageing during the study, while the average male experienced slight age acceleration. Unlike biological age, both epigenetic age measures suggested that both the average male and the average female experienced notable age acceleration over the course of the study. Extrinsic epigenetic age further suggested that before baseline, the average female had experienced considerable age acceleration and the average male had experienced around that much age acceleration again.

At each wave, biological age was approximately normally distributed, like in the Dunedin Study (Belsky et al., 2015), while the epigenetic age distributions had long tails, as some participants had epigenetic age estimates far above or below their chronological age. Both allostatic load and telomere length were sometimes positively skewed, meaning there were individuals with particularly high allostatic load and particularly long telomere length.

4.4.2 How do the weathering measures relate to each other?

Prior to this study, it was not clear how allostatic load and biological age related over time. We found that changes in each were strongly correlated: individuals who increased sharply in allostatic load were more likely to experience accelerated biological ageing. We also found correlations between higher baseline allostatic load and biological age and less steep increases in allostatic load and biological age, although these were likely partly the result of regression to the mean along with both measures being positively skewed (see Barnett, van der Pols, & Dobson, 2005 for a primer on regression to the mean).

At baseline, higher allostatic load was moderately correlated with older biological age. The only previous sample to relate contemporaneously measured allostatic load and biological age found that they had a strong positive correlation rather than a moderate strength correlation, but this study used a mixed-age sample with a minimum age of 30 and a mean age of 49.6 (continuous allostatic load-biological age r = .571; Levine & Crimmins, 2014). Further research using different samples, and simulations based on different general population values, will help to clarify how these measures' relationships differ in age-homogeneous versus mixed-age samples.

It is important to emphasise that because allostatic load scores from different operationalisations are imperfectly correlated, the relationships between allostatic load and other measures will vary based on the operationalisation used. Plus, because allostatic load and biological age operationalisations share some biomarkers, it is possible for allostatic load and biological age to correlate more strongly than different operationalisations of allostatic load. For example, Levine & Crimmins (2014) found that continuous allostatic load was more strongly correlated with biological age than it was with two cutoff-based measures of allostatic load. Interestingly, there are patterns of inter- or intra-individual physiological differences that similarly affect biological age estimates and continuous allostatic load scores, but are unlikely to affect allostatic load scores based on sample distributions. For example, consider a hypothetical individual in our study whose HbA1c increased from 41 to 42 mmol/mol from Waves 1 to 2, moving them into the high risk range for diabetes (National Institute for Health and Care Excellence, 2017). Their continuous allostatic load score and their estimated biological age would increase, but in this sample, all commonly used sample-based allostatic load scores would not be affected. In summary, there are methodological reasons why some operationalisations of allostatic load, including ours, might be expected to be more closely related to biological age than others.

The present study also tested the relationship between allostatic load and extrinsic epigenetic age, an epigenetic age measure that incorporates additional blood cell composition data. We found that at baseline, higher allostatic load was related to older extrinsic epigenetic age, with a small effect size. This relationship seemed to be weaker in participants who did not drop out of the study, as after dropout was added to the model, this correlation was attenuated by about 20% and was no longer statistically significant with an alpha of .05.

An unexpected finding was that longer telomere length at baseline was consistently associated with a steeper subsequent increase in allostatic load, with a small-to-moderate effect size. Previous studies had found that if there was any relationship, it was between shorter telomere length and higher allostatic load (Ahrens et al., 2016; Révész et al., 2014). Research has also associated shorter telomere length with HPA axis dysregulation (e.g., Tomiyama et al., 2012; Gotlib et al., 2015) and our allostatic load operationalisation did not use any HPA axis biomarkers. It may be, then, that telomere length relates differently to different components of allostatic load.

Strikingly small correlations were observed between measures that were not calculated using some of the same data (i.e., all correlations except allostatic load-biological age and extrinsic epigenetic age-intrinsic epigenetic age). This suggests that they index different aspects of stress and ageing and/or that they are noisy measures. It is notable that the measures seemed to mostly index different aspects of stress and ageing, considering that potential causal paths exist between some of the measures. Consider telomere length and markers of biological age. Telomere length has causal effects on biological age markers of heart and lung function (Demanelis, Tong, & Pierce, 2019), while inflammatory markers of biological age may shorten telomeres through generating oxygen free radicals (Kirchner et al., 2017). Other causal effects linking the two have also been posited (Blackburn et al., 2015). Yet, in our latent growth curve models, estimates of the baseline correlation between older biological age and shorter telomere length were small (all *r*s = -0.09). The raw contemporaneous correlations we observed between biological age and telomere length (-0.02 to -0.05) were smaller than the baseline estimate from our models, and very similar in size to the correlation observed by Belsky et al. (2018) at age 38 (r = -.05). We also observed weak correlations between biological age and intrinsic epigenetic age (r range = -0.03 to 0.06), all of which were smaller than the modest correlation observed by Belsky et al. (2018) at age 38 (r = .08). Our study replicates the weak correlations reported by Belsky et al. (2018) in multiple waves of an older age sample, and extends it: correlations between allostatic load and extrinsic epigenetic age were also small.

It may be helpful to take a psychometric approach to considering the strengths of these correlations. Method variance, such as wording similarities in questionnaires, can contribute to the correlations between psychometric items and so latent method factors are sometimes included in factor analyses (Tomas & Oliver, 1999). In the present study, method effects help to explain the strong correlations between measures that share measurements (e.g., the two epigenetic age measures) but they do not explain why most of the correlations are small.

Constructs from item response theory may also help us to understand the observed correlations. In item response theory, an item is *difficult* if the participant must be high on the latent trait to have a positive answer (Reise, Ainsworth, & Haviland, 2005). Also, items have information functions that show how well they discriminate between individuals across different levels of the latent trait (Reise et al., 2005). Here, it could be that the different measures are not strongly correlated because they have different difficulty levels for physiological weathering; for example, epigenetic age could be more sensitive to weathering than allostatic load. However, given the number of physiological processes involved, it seems more likely that the correlations are low because the measures mostly provide information about different physiological processes.

In summary, physiological weathering measures calculated using some of the same data tend to have moderate-to-strong associations with each other, while those calculated using different data tend to be weakly associated if they are associated at all. This suggests that they carry mostly different information about the effects of stress and ageing.

4.4.3 Strengths and limitations

Crucially, our study had four waves of data, collected over around 10 years, for allostatic load biomarkers, biological age biomarkers, and DNA methylation, along with three waves of telomere length data. This rich and rare data enabled us to conduct novel tests of the relationships between associated changes in these measures. We used an older age, age-homogeneous sample, which differentiated our study from previous studies correlating these measures. In addition, our study used pattern mixture modelling to check an assumption that is often not tested, and these models showed our results were not notably biased by participant dropout.

Our study also had limitations. Some of the standard errors for associated change correlations were large, and so we were not able to provide precise estimates of the relationships, or strong evidence whether these relationships do or do not exist. Our allostatic load measure lacked HPA axis biomarkers, and our biological age estimates were calculated using fewer biomarkers than those in previous studies (Belsky et al., 2015; Levine & Crimmins, 2014). In addition, while we had four waves of data for almost all of the weathering measures, we only had three waves of data for telomere length. Finally, it is worth noting that while the use of an older age sample differentiated our study from previous research, it also means that our results cannot be generalised to other age groups.

4.4.4 Future directions

Considering that the measures studied here seem to provide mostly different information about stress- and ageing-related physiological weathering, a focus of future research should be clarifying the pathways to different measures of physiological weathering as well as the pathways from the measures of physiological weathering to adverse health outcomes, including mortality.

Future investigations of allostatic load and biological age measurement should also consider using biomarkers measured more frequently than once in a number of years. While some biomarkers used for these measures are relatively stable throughout the day (e.g., CRP; Meier-Ewert et al., 2001), others are less so (e.g., SBP; Pickering, 1990). This could lead to the undesirable situation where an individual has a rather different allostatic load level and/or biological age at different times in the same day. Therefore, research should aim to understand and incorporate the stress- and age-linked intraindividual variability in indicators of allostatic load and biological age.

Epigenetic age measures will also benefit from further calibration using large samples.

Future research comparing weathering measures should, if possible, include mitochondrial health (Picard, Prather, et al., 2018). Mitochondrial allostatic load may cause both broader physiological allostatic load and cellular ageing (Picard et al., 2014), but it has not yet been confirmed whether it is more strongly related to allostatic load summary scores or other measures of physiological weathering.

4.4.5 Conclusion

We tested the longitudinal older age associations between five physiological weathering measures: allostatic load, biological age, extrinsic epigenetic age, intrinsic epigenetic age and telomere length. Aside from the two epigenetic age measures, the most closely related measures were allostatic load and biological age, which were operationalised using most of the same biomarkers. Higher allostatic load was moderately correlated with older biological age at baseline, and increases in allostatic load over time were strongly related to increased biological age acceleration. Strikingly, most correlations between these measures were small. This suggests that these measures provide different information about the physiological weathering caused by stress and ageing.

4.5 Coda

If allostatic load was very strongly related to other physiological weathering measures (e.g., rs ≈ |.90|), this would suggest that allostatic load and the other measures mostly index the same processes of age-related physiological weathering. This would have meant that associations with criterion variables would be similar. Additionally, it would have suggested that allostatic load operationalisations may need a calibration check to see if they more closely index stress- or age-related physiological dysregulation. However, this study found that allostatic load scores are not strongly related to epigenetic age or telomere length and are mostly independent of biological age estimates. Hence, allostatic load seems to mostly index different aspects of physiological weathering to other measures. This finding sets the foundation for the following two chapters, which will conduct novel tests relating longitudinal changes in allostatic load to longitudinal changes in two psychological variables: cognitive ability and depressive symptoms.

Chapter 5

How do allostatic load and cognitive ability relate over time in the Lothian Birth Cohort 1936?

5.1 Introduction

Non-pathological cognitive ageing has a tremendous health cost (Deary et al., 2009). Public health bodies are increasingly focusing efforts on identifying and measuring risk factors for deleterious cognitive ageing (e.g., Cambridge Institute of Public Health, 2016; Center for Disease Control and Prevention, 2017). It is crucial that we improve our understanding of how risk factors relate to cognitive ability and decline in older age, because this has the potential to help researchers to identify the mechanisms involved and help clinicians to assess and treat their patients. A large body of literature has identified the physiological effects of stress as being detrimental to health, including cognitive functioning. However, there remains limited research investigating whether the cumulative effects of stress impact changes in cognitive ability over time.

5.1.1 Possible mechanisms linking allostatic load and cognitive ability

The association between allostatic load and cognitive ability likely results from causal pathways in both directions and the effects of common causes.

The brain is central to both the stress response and cognitive functioning, so stress can affect both allostatic load and cognitive ability. The primary mediators released during the stress response can lead to allostatic load and can also affect cognitive ability through their effects on the brain. For example, prolonged higher levels of cortisol can affect levels of secondary outcome allostatic load markers and can cause decreased brain area volumes (Echouffo-Tcheugui et al., 2018; McEwen & Gianaros, 2011).

Researchers have tended to focus more on how allostatic load may affect cognitive ability, but there are also potential causal pathways from cognitive ability to allostatic load. For example, higher cognitive ability is associated with a greater stress response (Ginty, Phillips, Der, Deary, & Carroll, 2011; Ginty, Phillips, Roseboom, Carroll, & deRooij, 2012). There are also potential long-term effects: lower cognitive ability is associated with harmful health behaviours (Gottfredson & Deary, 2004), which are related to allostatic load (Forrester et al., 2019). In sum, the potential causal paths involving allostatic load and cognitive ability are myriad. The present study focuses on whether allostatic load can be useful as an indicator of risk for deleterious cognitive ageing, but its exploratory analyses may help to suggest directions for future research focused on mechanisms.

5.1.2 Allostatic load biomarkers and cognitive ability

Allostatic load's putative biomarkers from various physiological systems have been related to cognitive change in older age, although not consistently for all biomarkers studied. Examples of biomarkers linked to subsequent cognitive decline are higher midlife body mass index, a metabolic marker (e.g., Dahl et al., 2013), higher blood pressure, a cardiovascular marker (e.g., Elias, Wolf, D'Agostino, Cobb, & White, 1993; Glynn et al., 1999), and higher levels of interleukin-6, an inflammatory marker (e.g., Mooijaart et al., 2013; Economos et al., 2013). The effect sizes observed in these studies have tended to be small-to-moderate in size.

A smaller number of studies have investigated associations between change in allostatic load biomarkers and change in cognitive ability. Studies into such relationships have found, for example, that urinary cortisol increases tracked with ~28 month memory declines in women but not men, with a small-to-moderate effect size (*n* = 194, sample aged 70-79 years; Seeman et al., 1997), that systolic blood pressure increases tracked with Mini-Mental State Examination declines in normotensive Mexican-Americans over ~7 years, with a moderate effect size (*n* = 2,859, sample aged 65+ years; Insel, Palmer, Stroup-Benham, Markides, & Espino, 2005), and that extreme longitudinal variation in C-reactive protein was related to cognitive decline on the Modified Mini-Mental Status Exam in women (*HR* = 1.8) and those without an *APOE* e4 allele (*HR* = 1.6), but not men or *APOE* e4 allele carriers (Metti et al., 2014).

Studies that have not incorporated measures of allostatic load or multisystem physiological dysregulation have nonetheless suggested that risk for subsequent cognitive decline may be increased by the presence of dysregulation in multiple biomarkers and across multiple physiological systems. For example, studies have found that metabolic syndrome tends to better predict cognitive decline than single metabolic markers, and that risk is heightened for those with metabolic syndrome who also have high inflammation (Yaffe, 2007).

5.1.3 Allostatic load summary measures and cognitive ability

Relatively few studies have investigated the relationships between allostatic load summary measures and cognitive ability.

A few studies have found relationships between allostatic load and contemporaneously measured cognitive ability. In the Midlife in the United States study (MIDUS; *n* = 1,076; *Mdn* age = 57 y), allostatic load was related to lower scores on contemporaneously measured factors of episodic memory and executive function (Karlamangla et al., 2014). Also, in the MacArthur Studies of Successful Aging, at baseline in 1988 (*N* = 1,189; age *range* = 70-79 y), allostatic load had small negative correlations with general cognitive ability (*r* = -.13) and specific cognitive abilities (spatial ability, memory, and abstract reasoning; *r* range = -.11 to -.09; Seeman et al., 1997). In the Social Environment and Biomarkers of Aging Study (SEBAS; based in Taiwan; *n* = 1,023 for cited studies; range of *mean* ages = 67-68 years), an allostatic load profile involving dysregulation of cardiovascular and metabolic biomarkers was related to a contemporaneously measured cognitive summary variable (Seplaki, Goldman, Weinstein, & Lin, 2006). Plus, allostatic load scored in some ways but not others was related to worse contemporaneously measured temporal orientation (a cognitive impairment test; Seplaki et al., 2005). The present study will use data from the Lothian Birth Cohort 1936 (LBC1936), a cohort study with a focus on cognitive ageing. Previous LBC studies have found small negative associations between allostatic load and general cognitive ability at Waves 1 and 2 (Booth et al., 2015; Crook et al., 2018). At Wave 2, allostatic load also had small negative associations with the specific cognitive factors of knowledge and processing speed. and Wave 2 allostatic load also had a small negative association with age 11 IQ (Booth et al., 2015).

Research has also investigated the longitudinal relationships between allostatic load and cognitive ability measures. In the West of Scotland Twenty-07 Study's 1972-born cohort (*N* = 705), slower processing speed at age 16 (measured by choice reaction time) had a small correlation with higher allostatic load at age 36 (*r* = .13; Gale et al., 2015). In the MacArthur Studies of Successful Aging, allostatic load in 1988 was also modestly associated with memory decline between 1988 and 1991 (*r* = -.08), predicted change in general cognitive ability between 1988 and 1996 (Seeman et al., 2001), and was moderately correlated with 1998-1991 and 1991-1996 cognitive declines (canonical *r* = .29; Karlamangla et al., 2002). However, in SEBAS, allostatic load and non-clinical biomarker summary scores did not predict subsequent cognitive impairment, nor subsequent decline in cognitive impairment (*n* = 820; Goldman et al., 2006). The effect size estimate for the association between allostatic load and cognitive impairment was extremely small: one additional allostatic load biomarker in the top or bottom 10% of sample values was associated with 0.01 more cognitive tasks completely incorrectly (Goldman et al., 2006). The association between cognitive impairment and a clinical biomarker summary score was around twice as strong and was statistically significant, though the effect size was still very small (Goldman et al., 2006). In the LBC1936, Wave 1 allostatic load was related to general cognitive decline between Waves 1 and 3 (mean ages 70 and 76; beta range when jointly estimated across genotyptic groups = -.87 to to -.98; Crook et al., 2018).

5.1.4 The present study

In the present study, we used latent growth curve modelling with four waves of LBC1936 data to test contemporaneous and longitudinal relationships between allostatic load and cognitive ability in older age, from approximately 70 to 79 years. This enabled us to improve understanding of how allostatic load and cognitive ability relate over time, in particular by testing the relationship between change in allostatic load and change in cognitive ability.

The overarching research question of the current study is: how do allostatic load and cognitive ability relate over time? Specifically, we will seek to address the following research questions:

1. Does initial (age 70) allostatic load relate to subsequent change (between ages 70 and 79) in cognitive ability?
2. Does initial cognitive ability relate to subsequent change in allostatic load?
3. Does change in allostatic load correlate with change in cognitive ability?

We also fitted our latent growth curve model with age 11 cognitive ability, coded as IQ scores, added as a predictor of older age level and change in allostatic load and cognitive ability. This enabled us to test the relationships age 11 IQ has with older age level and change in allostatic load, and test whether the relationships between allostatic load and cognitive ability in older age are affected by accounting for age 11 cognitive ability.

4. Does age 11 IQ predict allostatic load at age 70 and change in allostatic load between ages 70 and 79?
5. Does statistically controlling for the effects of age 11 IQ affect the contemporaneous and longitudinal relationships between older age allostatic load and older age cognitive ability?

Our main models tested these research questions in relation to a factor of general cognitive ability. We also fitted exploratory models investigating each of these research questions in relation to domain-specific factors of cognitive ability and scores on specific cognitive tests.

5.2 Methods

5.2.1 Measures

5.2.1.1 Cognitive ability

5.2.1.1.1 Childhood cognitive ability. Most participants sat the Moray House Test No. 12 on 4th June, 1947 as part of the Scottish Mental Survey 1947. The Moray House Test No. 12 is a validated intelligence test (Scottish Council for Research in Education, 1949). It is a paper-and-pencil test, administered in groups, with 75 items of various types, a 45-minute time limit, and a maximum possible score of 76 (Scottish Council for Research in Education, 1949). We used Moray House Test No. 12 scores from the Scottish Mental Survey 1947 that have previously been corrected for age at the time assessment and converted to age 11 IQ scores.

5.2.1.1.2 Older age cognitive ability. The cognitive tests taken by the LBC1936 have previously been detailed (Deary et al., 2007). We used data from the following cognitive ability tests:

- the **National Adult Reading Test** (NART; a test of word recognition and pronunciation, requires pronouncing 50 irregular words; Nelson & Willison, 1991);

- **verbal fluency** total score (a test of executive function, requires listing as many words as possible beginning with a given letter [here C, F, and L] in a one minute period; Lezak, Howieson, Loring, Hannay, & Fischer, 2004);
- the **Wechsler Test of Adult Reading** (WTAR; another test of word recognition and pronunciation, also requires pronouncing 50 irregular words; Holdnack, 2001);
- **block design** (a test of non-verbal reasoning, requires reproducing a design using blocks; Wechsler, 1998a);
- **digit span backwards** (a test of working memory, requires recall of increasingly long lists of numbers in reverse order; Wechsler, 1998b);
- **letter-number sequencing** (a test of working memory, requires recall of increasingly long lists of letters and numbers, with numerical order recall of the numbers followed by alphabetical order recall of the letters; Wechsler, 1998a);
- **matrix reasoning** (a test of non-verbal reasoning, items require choosing the option that correctly completes the pattern in a matrix; Wechsler, 1998a);
- **spatial span forwards** (a test of non-verbal learning and memory, requires participants to repeat a sequence of block touches; Wechsler, 1998b);
- **spatial span backwards** (a test of non-verbal learning and memory, requires participants to observe a sequence of block touches and then enact the sequence in reverse order; Wechsler, 1998b);
- **digit symbol coding** (a test of processing speed, requires entering symbols next to numbers according to a digit-symbol code, with a time limit of two minutes; Wechsler, 1998a);
- **inspection time** (a test of processing speed, requires identifying the longer of two briefly displayed vertical lines on trials with

display times from 6ms to 200ms; Deary et al., 2004);

- **simple reaction time** (a test of processing speed, requires pressing a key as quickly as possible after a stimulus appears; Deary, 2001);
- **choice reaction time** (a test of processing speed, requires pressing the correct numbered key [1, 2, 3, or 4] as quickly as possible after the matching stimulus appears; Deary, 2001);
- **symbol search** (a test of processing speed, requires identifying whether either of two target symbols appears in a row of symbols, with items completed up to a time limit of two minutes; Wechsler, 1998a);
- **logical memory immediate recall** (a test of immediate verbal declarative memory, requires immediate recall of a story with 25 elements, with there being two stories and the second story delivered twice; Wechsler, 1998b);
- **logical memory delayed recall** (a test of delayed verbal declarative memory, requires delayed recall of the two stories from the logical memory immediate recall test; Wechsler, 1998b);
- **verbal paired associates immediate recall** (a test of verbal learning and memory in which participants hear unrelated word pairs, before being given the first of the pair and being asked to recall the second; Wechsler, 1998b);
- and **verbal paired associates delayed recall** (a test of verbal learning and memory in which participants are given the first word of the word pairs from the verbal paired associates immediate recall test and asked to recall the second; Wechsler, 1998b).

5.2.2 Data exclusion criteria

Participants were excluded if, during a structured interview, they self-reported having dementia. Participants were also excluded if they had an abnormal score on the Mini-Mental State Examination (MMSE;

Folstein, Folstein, & McHugh, 1975) for their level of education, as such scores were taken as indicating possible cognitive impairment due to pathological cognitive ageing. Educational level was collected as part of a structured interview at Wave 1, where participants were asked how many years of full-time formal education they had. We used Mitchell's (2013, p. 37) "education-adjusted cut-off points for an abnormal score[:] <21 for patients with a basic school education, <23 for a high school education, and <24 for graduate/university education."

5.2.3 Statistical analysis

We followed a two-step approach in the current study. First, we established plausible latent variable measurement models for the cognitive ability tests and assessed the similarity of this model across the four waves via measurement invariance testing. Following this, we fitted latent growth curve models to the cognitive and allostatic load data across waves in order to test our research hypotheses.

5.2.3.1 Cognitive ability measurement models

In our study preregistration (Crook et al., 2017b), we originally planned to fit second-order latent growth curve models using a bifactor cognitive ability model previously applied to cross-sectional LBC1936 data (Booth, Bastin, et al., 2013; Booth et al., 2015). In this model, each of the 18 cognitive tests loads of a general factor as well as on a specific cognitive factor, namely knowledge, non-verbal reasoning, processing speed or verbal memory. However, latent growth curve models based on this model structure did not converge. Our second planned approach was to fit latent growth curve models using the same set of five cognitive factors, but instead of including the factor structure in the model, the growth model would be applied to factor scores from the longitudinal confirmatory factor analysis model for the bifactor model. However, this model structure did not produce models with satisfactory fit (see Section 5.3.1).

Hence, we fitted models using the final contingency model structure from our preregistration. In our models, there was a single general cognitive ability factor in which scores from the following seven tests loaded on the general factor at each wave: block design, digit symbol coding, letter-number sequencing, matrix reasoning, symbol search, logical memory delayed recall, and the WTAR.

To identify the model, the variances of the cognitive ability factors were fixed at 1.

5.2.3.1.1 Wave 1 measurement model fit. For our Wave 1 cognitive ability measurement model, fit index results beyond the following thresholds were taken to indicate satisfactory fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003; van de Schoot, Lugtig, & Hox, 2012): confirmatory fit index (*CFI*) $\geq$ .95; root mean square error of approximation (*RMSEA*) $\leq$ .08; standardised root mean square residual (*SRMR*) $\leq$ .10. As suggested cutoffs for the *CFI* differ (Schermelleh-Engel et al., 2003; van de Schoot et al., 2012), the following *CFI* range will indicate borderline satisfactory fit: .95 > *CFI* $\geq$ .90. We planned that if, on balance, the initial model did not have acceptable fit, the model would be modified based on modification indices, with the modifications that would produce the greatest improvement in fit made one at a time until the model has acceptable fit. Note that while modification indices can be used to improve model fit, they can lead to overfitting and the introduction of nonsense parameters, particularly where there is not an empirical basis for the proposed model modifications (Bollen & Noble, 2011). In the present study, some suggested modifications did not make substantive sense and/or affected the core structure of the model; we decided not to consider these modifications.

The robust variants of the *CFI* and *RMSEA* were used throughout.

5.2.3.1.2 Longitudinal measurement model. After a model with acceptable fit to the Wave 1 data was found, data from Waves 2-4 were added to the model. In the longitudinal measurement models, we included correlations between the different waves of the general cognitive ability factor as well as correlated residuals across waves for each cognitive test (e.g., all waves of symbol search residuals were allowed to correlate). Measurement invariance of the cognitive factors across waves was then tested. Measurement invariance was assessed by sequentially adding equivalence constraints to the measurement models and comparing model fit at each step.

The following sequence of models was fitted:

1. Configural invariance: Cognitive tests are associated with the same factor across waves.
2. Metric invariance: Factor indicator loadings constrained to be equal across waves.
3. Strong invariance: Factor indicator intercepts constrained to be equal across waves.
4. Strict invariance: Factor indicator residuals and correlated residuals constrained to be equal across waves.

5.2.3.1.3 Longitudinal measurement invariance testing. To assess whether measurement invariance held, we considered various model fit indices. First, following Chen (2007), in testing the metric constraints (δ models 1 and 2), differences of $\geq$ -.010 in *CFI*, $\geq$ .015 in *RMSEA* or $\geq$ .030 in *SRMR*, were taken as indicative of noninvariance. In testing for strong and strict invariance, differences of $\geq$ -.010 in *CFI*, supplemented by a difference of $\geq$ .015 in *RMSEA* or a change of $\geq$ .010 in *SRMR* were taken to indicate noninvariance. We also consulted the Akaike Information Criterion and the Bayesian Information Criterion, for which an increase was taken to indicate noninvariance (van de Schoot et al., 2012). For completeness, we also considered the chi-square test, for which $p \leq .05$ indicated noninvariance. However, note that the chi-square test is highly sensitive to sample size as well as distributional assumption violations (Chen, 2007). In large samples like the one here, it tends to over-reject models with acceptable fit (Chen, 2007).

As these various fit measures have their own advantages and disadvantages, we considered all of them when deciding whether, on balance, invariance was achieved at each stage. We planned that if the model comparisons testing weak, strong, or strict factorial invariance suggest that full measurement invariance (invariance of all the constrained parameters) was not achieved, we would consult modification indices to find and make fit-improving changes to the model so that partial measurement invariance (invariance of some of the constrained parameters) could be achieved.

Following the guidance of Widaman, Ferrer, & Conger (2010), in our latent growth curve models we interpreted the results relating to cognitive factors that had at least strong longitudinal invariance, that is, invariance of factor indicator loadings and intercepts across waves.

5.2.3.2 Latent growth curve models

Figure 5.1 shows the general specification of the dual change latent growth curve models. Note that for general cognitive ability, the latent growth curve indicators were the latent general cognitive ability factor at each wave, rather than an observed score.

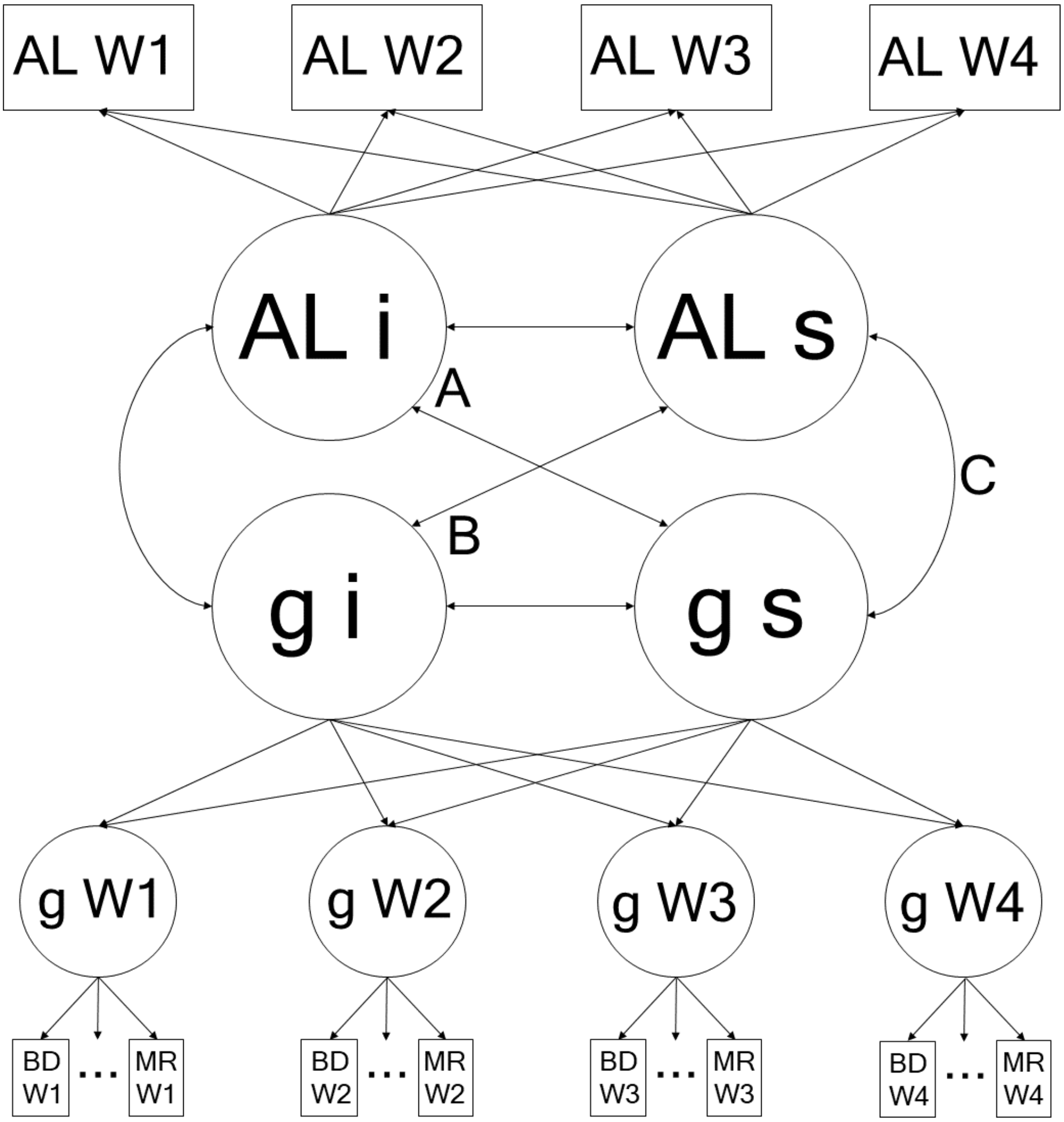

Figure 5.1: Allostatic load-general cognitive ability latent growth curve model specification. A-C are labelled paths described in the text. AL = allostatic load. BD = block design. g = general cognitive ability. i = intercept. MR = matrix reasoning. s = slope. W = wave.

Correlations between the intercept and slope growth factors tested our primary research questions. The correlation between the allostatic load intercept and cognitive ability slope (path A in Figure 5.1) tested question 1. The correlation between the cognitive ability intercept and the allostatic load slope (path B) tested question 2. The correlation between the allostatic load slope and the cognitive ability slope (path C) tested question 3.

To address our final two research questions, age 11 IQ was added to the model as a predictor of all growth factors. The regressions of the allostatic load intercept and slope on age 11 IQ tested question 4. To answer the question 5, we assessed how the growth factor correlations differed between models with and without age 11 IQ.

5.2.3.2.1 Exploratory analyses. The bifactor models we intended to use for our confirmatory models used the results from 18 cognitive tests. These bifactor models included a general factor and four specific cognitive factors. However, as these models did not fit, we used contingency models that modelled the results from only seven cognitive tests. To investigate how allostatic load relates over time to specific cognitive abilities, we fitted additional exploratory models. For each of the 18 cognitive tests, we fitted four latent growth curve models: the MAR main model; a MAR model with age 11 IQ as a predictor of all growth factors; and NMAR versions of those two models, which used to the same pattern-mixture modelling approach as our confirmatory models.

Please note that these were exploratory models that were not preregistered and were planned after the main analyses had commenced. These models should not be taken as strongly confirming or refuting hypotheses, but may be used as the basis for future confirmatory research.

5.3 Results

Table 5.1: Descriptive statistics for age, allostatic load and cognitive ability tests

	n	Mean	SD	Skew	Kurtosis
Age W1	1088	69.53	0.83	-0.04	-0.88
Age W2	861	72.49	0.71	-0.01	-0.83
Age W3	685	76.24	0.67	-0.05	-0.81
Age W4	536	79.32	0.62	0.08	-0.74
AL W1	1087	0.69	0.24	1.07	2.15
AL W2	861	0.71	0.24	1.17	3.17
AL W3	684	0.85	0.25	0.71	1.13
AL W4	535	0.87	0.25	0.59	0.45
NART W1	1086	34.53	8.09	-0.52	-0.25
NART W2	859	34.47	8.05	-0.57	0.00
NART W3	684	35.09	7.99	-0.64	0.09
NART W4	533	35.76	8.06	-0.64	0.03
Verbal fluency W1	1084	42.46	12.51	0.31	-0.12
Verbal fluency W2	860	43.22	12.89	0.24	0.03
Verbal fluency W3	685	43.04	12.66	0.28	0.10
Verbal fluency W4	534	43.95	13.18	0.07	-0.24
WTAR W1	1086	41.06	7.11	-0.94	0.53
WTAR W2	859	41.09	6.81	-0.89	0.55
WTAR W3	683	41.15	7.00	-1.12	1.42
WTAR W4	533	41.76	6.91	-1.14	1.22
Block design W1	1082	33.83	10.31	0.25	-0.27
Block design W2	859	33.71	10.06	0.48	0.11
Block design W3	679	32.39	9.79	0.36	0.25
Block design W4	523	31.48	9.36	0.51	0.39
Digit span backwards W1	1087	7.74	2.26	0.51	-0.12
Digit span backwards W2	861	7.83	2.28	0.29	-0.09
Digit span backwards W3	683	7.81	2.36	0.30	-0.38
Digit span backwards W4	535	7.62	2.15	0.39	-0.13
Letter-number sequencing W1	1076	10.95	3.13	0.17	0.01
Letter-number sequencing W2	858	10.94	3.06	0.26	0.35
Letter-number sequencing W3	678	10.51	2.98	0.21	0.41
Letter-number sequencing W4	524	10.17	2.79	0.35	0.40

Table 5.1: Descriptive statistics for age, allostatic load and cognitive ability tests *(continued)*

	n	Mean	SD	Skew	Kurtosis
Matrix reasoning W1	1083	13.52	5.12	-0.03	-1.00
Matrix reasoning W2	858	13.22	4.95	-0.04	-1.00
Matrix reasoning W3	677	13.10	4.90	0.01	-0.93
Matrix reasoning W4	523	13.03	5.00	0.09	-0.95
Spatial span forwards W1	1083	7.68	1.64	-0.18	-0.37
Spatial span forwards W2	858	7.64	1.66	-0.06	-0.47
Spatial span forwards W3	678	7.59	1.62	-0.08	-0.32
Spatial span forwards W4	524	7.42	1.60	-0.06	-0.39
Spatial span backwards W1	1081	7.05	1.73	0.02	-0.13
Spatial span backwards W2	856	7.08	1.60	-0.10	-0.13
Spatial span backwards W3	678	7.07	1.58	-0.06	-0.38
Spatial span backwards W4	524	6.76	1.58	0.01	-0.12
Digit symbol W1	1083	56.66	12.90	0.04	-0.21
Digit symbol W2	858	56.47	12.30	0.03	-0.14
Digit symbol W3	674	54.12	12.66	-0.03	-0.10
Digit symbol W4	524	51.64	12.73	-0.06	-0.05
Inspection time W1	1038	112.18	10.98	-0.79	1.91
Inspection time W2	834	111.28	11.78	-1.22	4.12
Inspection time W3	645	110.30	12.44	-0.76	1.09
Inspection time W4	461	107.06	13.58	-1.15	2.90
SRT mean W1	1082	0.28	0.06	2.18	7.79
SRT mean W2	860	0.28	0.05	1.94	6.86
SRT mean W3	677	0.28	0.05	1.60	4.28
SRT mean W4	531	0.29	0.05	1.35	2.79
CRT mean W1	1081	0.64	0.09	0.96	2.38
CRT mean W2	860	0.65	0.09	1.15	3.94
CRT mean W3	675	0.68	0.10	0.84	1.53
CRT mean W4	531	0.70	0.11	1.01	2.32
Symbol search W1	1081	24.79	6.24	0.04	0.40
Symbol search W2	857	24.67	6.13	-0.32	0.74
Symbol search W3	675	24.81	6.26	-0.13	0.66
Symbol search W4	517	22.98	6.41	-0.13	0.72

Table 5.1: Descriptive statistics for age, allostatic load and cognitive ability tests *(continued)*

	n	Mean	SD	Skew	Kurtosis
Logical memory immediate recall W1	1085	44.23	10.30	-0.33	-0.13
Logical memory immediate recall W2	859	45.77	10.15	-0.44	0.30
Logical memory immediate recall W3	681	46.02	10.69	-0.40	0.05
Logical memory immediate recall W4	533	45.19	11.08	-0.59	0.80
Logical memory delayed recall W1	1084	27.37	7.99	-0.40	0.01
Logical memory delayed recall W2	859	28.79	7.94	-0.58	0.46
Logical memory delayed recall W3	679	28.98	8.55	-0.60	0.44
Logical memory delayed recall W4	531	28.36	8.76	-0.63	0.60
VPA immediate recall W1	1057	20.24	7.41	-0.40	-0.73
VPA immediate recall W2	842	20.84	7.66	-0.50	-0.72
VPA immediate recall W3	659	20.19	7.70	-0.43	-0.72
VPA immediate recall W4	496	20.92	7.64	-0.54	-0.53
VPA delayed recall W1	1048	6.17	2.03	-0.97	-0.01
VPA delayed recall W2	840	6.39	2.02	-1.25	0.66
VPA delayed recall W3	658	6.25	2.07	-1.13	0.36
VPA delayed recall W4	493	6.32	2.06	-1.20	0.50
Age 11 IQ	1028	100.00	14.99	-0.80	0.92

Note.

AL = allostatic load. CRT = choice reaction time. NART = National Adult Reading Test. SRT = simple reaction time. VPA = verbal paired associates. W = wave. WTAR = Wechsler Test of Adult Reading.

Table 5.1 contains descriptive statistics for the study variables. The allostatic load mean increased throughout the waves, with the biggest jump in means being between Waves 2 and 3. However, there was substantial variability in scores at all waves. The cognitive tests displayed different patterns of change. For example, the WTAR mean increased across waves while the matrix reasoning mean decreased across waves. Block design, digit symbol coding and letter-number sequencing each showed little change in means between Waves 1 and 2, followed by gradual decline from Waves 2 to 4.

5.3.1 Measurement models: longitudinal CFA and unconditional latent growth curve models

We began by fitting measurement models using a bifactor cognitive ability model previously identified in this sample (Booth et al., 2013; 2015). Bifactor longitudinal measurement models had satisfactory fit well. However, when we fitted the unconditional bifactor latent growth curve model, the model did not converge. We then fitted models based on our first contingency model structure, in which growth factors reflected the level and change in the factor scores from the bifactor longitudinal measurement model. This model did not have satisfactory fit: *CFI* = 0.63, *RMSEA* = 0.23; *SRMR* = 0.18.

Finally, we fitted models based on our last contingency model structure, in which seven cognitive tests loaded on to a general cognitive factor and there were no specific cognitive factors. Table 5.2 contains the models fitted using this model structure and the *CFI*, *RMSEA* and *SRMR* for each model. Based on modification indices for the Wave 1 measurement model for this structure, we added four correlated residuals: digit symbol coding with symbol search, block design with matrix reasoning, logical memory delayed recall with WTAR, and block design with symbol search. In the longitudinal measurement models for this structure, we had to free the intercepts for Wave 4 block design and WTAR to achieve partial measurement invariance. However, the latent growth curve model based on this model had a negative variance for Wave 4 general cognitive ability. After this variance was fixed to zero, the model only barely achieved satisfactory fit: *CFI* = 0.95; *RMSEA* = 0.05; *SRMR* = 0.09. Hence, following Ferrer, Balluerka, & Widaman (2008), we tested measurement invariance within an unconditional latent growth curve model framework to see if the same model was identified as the best fitting or if a better fitting model could be found. Initially, we compared models with different reference indicators but no further constraints. Again, the Wave 4 general cognitive ability variance was consistently negative and had to be fixed at zero. The best fitting model used letter-number sequencing as the reference indicator, so it was used as the reference indicator in all subsequent models. To achieve partial measurement invariance in these models, four intercepts had to be freely estimated: Wave 4 block design, Wave 3 and Wave 4 WTAR, and Wave 1 delayed recall. These modifications to the measurement model were found via assessing modification indices and the results from models without intercepts constrained. No other modifications had to be made to achieve measurement invariance.

Table 5.2: Cognitive ability measurement model fit statistics

	CFI	RMSEA	SRMR
LCFA: Initial longitudinal measurement model with configural invariance	0.973	0.039	0.054
LCFA: Added invariance of loadings	0.965	0.044	0.224
LCFA: Add invariance of intercepts	0.951	0.050	0.111
LCFA: Freed Wave 4 block design intercept	0.951	0.050	0.111
LCFA: Freed Wave 4 WTAR intercept	0.961	0.045	0.075
LCFA: Freed Wave 4 symbol search intercept	0.961	0.045	0.075
LCFA: Added invariance of residuals	0.958	0.045	0.078
LCFA: Added invariance of correlated residuals	0.956	0.045	0.080
LGCM: Inital model based on final longitudinal measurement model	0.950	0.048	0.087
LGCM: Configural invariance, block design as reference indicator	0.973	0.039	0.055
LGCM: Configural invariance, digit symbol coding as reference indicator	0.971	0.041	0.054
LGCM: Configural invariance, letter-number sequencing as reference indicator	0.973	0.039	0.055
LGCM: Configural invariance, matrix reasoning as reference indicator	0.973	0.039	0.055
LGCM: Configural invariance, symbol search as reference indicator	0.970	0.041	0.055
LGCM: Configural invariance, delayed logical memory recall as reference indicator	0.971	0.041	0.055
LGCM: Configural invariance, WTAR as reference indicator	0.973	0.039	0.054
LGCM: Retained letter-number sequencing as reference indicator, added invariance of loadings	0.972	0.039	0.060
LGCM: Added invariance of intercepts	0.949	0.051	0.104

Table 5.2: Cognitive ability measurement model fit statistics *(continued)*

	CFI	RMSEA	SRMR
LGCM: Freed Wave 4 block design intercept	0.950	0.051	0.101
LGCM: Freed Wave 4 WTAR intercept	0.955	0.048	0.082
LGCM: Freed Wave 3 WTAR intercept	0.959	0.046	0.071
LGCM: Freed Wave 4 logical memory delayed recall intercept	0.962	0.044	0.069
LGCM: Added invariance of residuals	0.960	0.044	0.072
LGCM: Added invariance of correlated residuals	0.958	0.044	0.072

Note.

CFI and RMSEA are robust variants. CFI = comparative fit index. LCFA = longitudinal confirmatory factor analysis. LGCM = latent growth curve model. RMSEA = root mean square error of approximation. SRMR = standardised root mean square residual. WTAR = Wechsler Test of Adult Reading.

The final unconditional latent growth curve model for general cognitive ability had satisfactory fit and fitted better than the unconditional latent growth curve model based on the longitudinal measurement model: *CFI* = 0.96; *RMSEA* = 0.04; *SRMR* = 0.07. Standardised loadings in the general cognitive ability measurement part of the model ranged from 0.49 to 0.74 in absolute size. There was notable variation in baseline level (intercept mean estimate [*SE*] = 0.005 [0.027]; intercept variance estimate [*SE*] = 0.38 [0.03]). On average, general cognitive ability declined gradually over time (slope mean estimate [*SE*] = -0.041 [0.002]), although there was notable variation in individual trajectories (slope variance estimate [*SE*] = 0.0003 [0.0001]). Baseline level and subsequent trajectory were not statistically significantly correlated (*r* [*SE*] = -0.02 [0.09]).

The final unconditional latent growth curve model for allostatic load did not have satisfactory fit: *CFI* = 0.90; *RMSEA* = 0.10; *SRMR* = 0.06. There was notable variability in allostatic load at baseline (intercept mean estimate [*SE*] = 0.68 [0.01]; intercept variance estimate [*SE*] = 0.029 [0.004]). On average, allostatic load gradually increased over time (slope mean estimate [*SE*] = 0.021 [0.001]), although as with general cognitive ability, there was notable variation in individual trajectories (slope variance estimate [*SE*] = 0.00033 [0.00007]). Higher baseline allostatic load was related to a flatter trajectory in allostatic load, with a moderate strength effect size (*r* [*SE*] = -0.31 [0.10]).

5.3.2 Allostatic load-general cognitive ability latent growth curve models

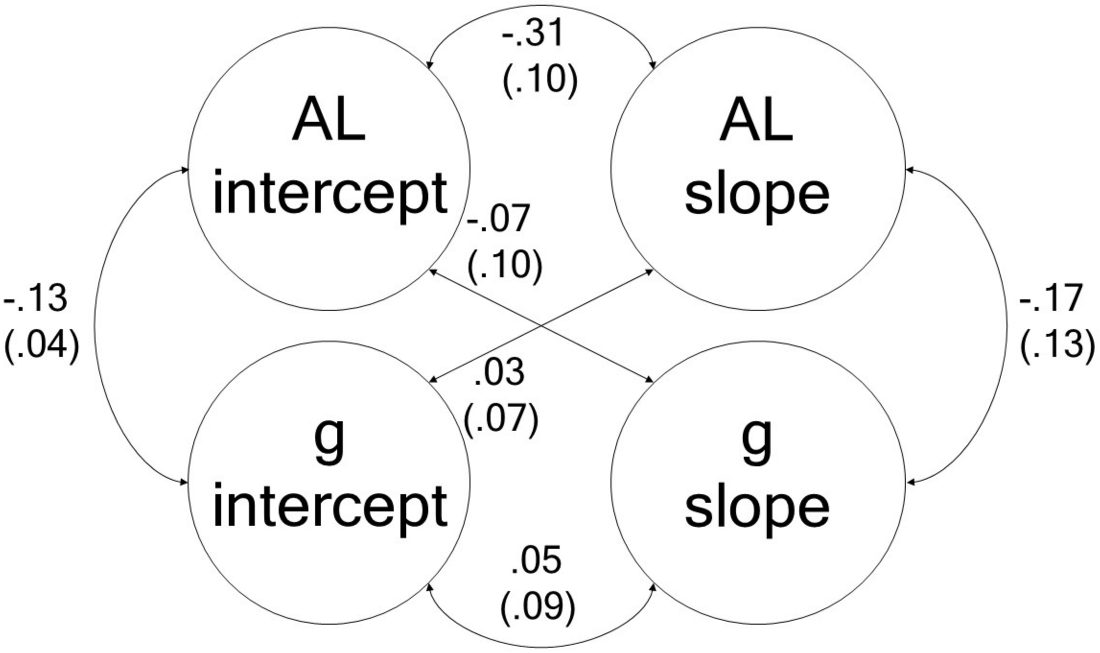

Figure 5.2: Allostatic load-general cognitive ability MAR model growth factor correlations. Correlations are displayed, with their standard errors in parentheses. AL = allostatic load. g = general cognitive ability.

Figure 5.2 contains results from latent growth curve models that modelled the initial level and trajectories over time for both general cognitive ability and allostatic load. There was a small negative correlation between the intercept growth factors (*r* [*SE*] = -0.13 [0.04]), which means that lower general cognitive ability was related to higher allostatic load levels at age 70. The allostatic load intercept and slope were moderately negatively correlated (*r* [*SE*] = -0.31 [0.10]), which means that those with higher allostatic load at baseline had less steep increases (or were more likely to decrease) in allostatic load over time.

Research question 1 was addressed by the correlation between the allostatic load intercept (allostatic load at age 70) and the general cognitive ability slope (change in general cognitive ability between ages 70 and 79). Research question 2 was addressed by the correlation between the general cognitive ability intercept (general cognitive ability at age 70) and the allostatic load slope (change in allostatic load between ages 70 and 79). Both of these correlations were very small and their standard errors were large relative to the parameter estimates. Hence, we did not find that age 70 allostatic load relates to subsequent change in general cognitive ability, nor did we find that age 70 general cognitive ability relates to subsequent change in allostatic load.

Research question 3 was addressed by the correlation between the two slopes. This research question asked about the association between changes in allostatic load and general cognitive ability between ages 70 and 79. The allostatic load and general cognitive ability slopes had a small-to-moderate negative correlation, but this correlation had a large standard error and was not statistically significant (*r* [*SE*] = -0.17 [0.13]), so further research is required on associated longitudinal change in allostatic load and general cognitive ability.

5.3.2.1 Not missing at random parallel process latent growth curve models

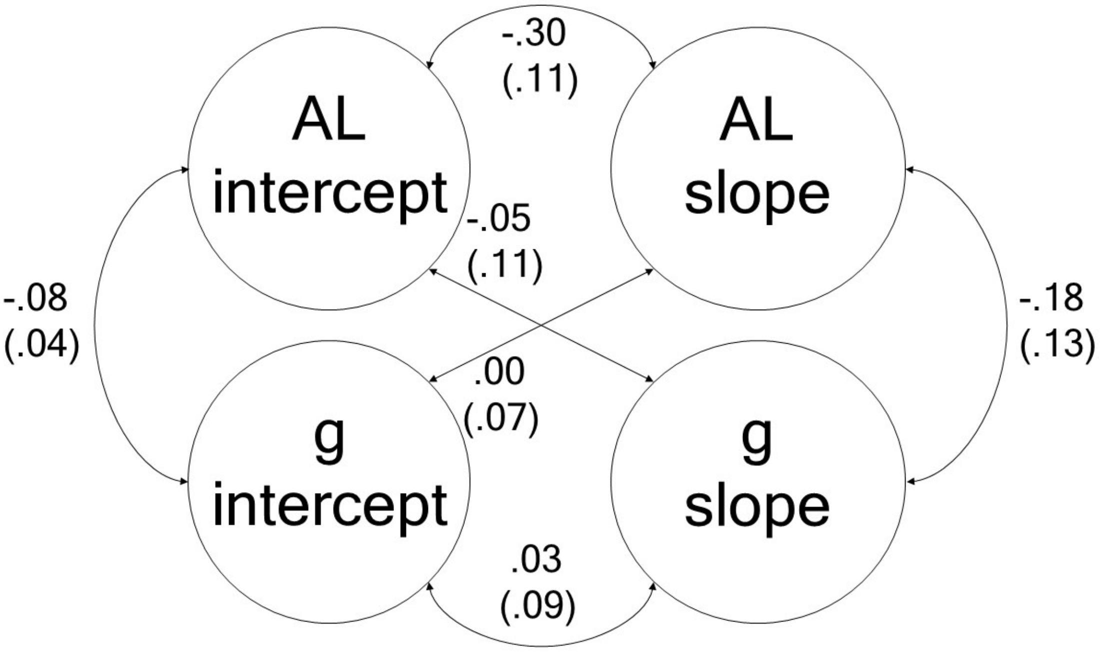

Figure 5.3: Allostatic load-general cognitive ability NMAR model growth factor correlations. Correlations are displayed, with their standard errors in parentheses. AL = allostatic load. g = general cognitive ability.

Figure 5.3 contains results from allostatic load-general cognitive ability latent growth curve models that assumed missing data were not missing at random after the data in the model was taken into account. Here, binary dummy variables (d2-d4) that indicated if a participant dropped out at Waves 2-4 respectively. Compared to those who did not drop out, those who dropped out at Wave 2 had 0.70 *SD* lower general cognitive ability (*SE* = 0.09) and 0.40 *SD* higher allostatic load at baseline (*SE* = 0.11), while those who dropped out at Wave 3 had 0.56 *SD* lower general cognitive ability (*SE* = 0.09) and 0.47 *SD* higher allostatic load at baseline (*SE* = 0.13). Also, those who dropped out at Wave 2 or 3 had 0.68 *SD* steeper decline in general cognitive ability (*SE* = 0.36). Those who dropped out at Wave 4 had 0.40 *SD* lower general cognitive ability at baseline (*SE* = 0.09) and 0.50 *SD* steeper decline in general cognitive ability (*SE* = 0.23).

Notably, the correlation between the allostatic load and general cognitive ability intercepts was attenuated in this model and was no longer statistically significant with an alpha of .05 (*r* [*SE*] = -0.08 [0.04]). This suggests that the correlation between allostatic load and general cognitive ability at age 70 is partly driven by participants who dropped out at Waves 2-4.

5.3.3 Adding control for age 11 IQ

Table 5.3: Allostatic load-general cognitive ability MAR model results after controlling for age 11 IQ

	Estimate	SE	p-value
g i regression on age 11 IQ	0.73	0.02	< 0.001
g s regression on age 11 IQ	-0.19	0.07	0.007
AL i regression on age 11 IQ	-0.05	0.04	0.197
AL s regression on age 11 IQ	0.04	0.06	0.477
AL i-g i correlation	-0.13	0.05	0.006
AL s-g s correlation	-0.17	0.14	0.226
AL i-g s correlation	-0.08	0.11	0.479
AL s-g i correlation	-0.01	0.08	0.894
AL i-AL s correlation	-0.31	0.10	0.003
g i-g s correlation	0.29	0.12	0.021

Note.
Regression betas standardised on all variables. AL = allostatic load; g = general cognitive ability; i = intercept; s = slope.

Table 5.3 contains results from the missing at random model after age 11 IQ was added as a predictor of all four growth factors. Age 11 IQ was strongly positively related to general cognitive ability at age 70 (β [*SE*] = 0.73 [0.02]). It had a weak but statistically significant relationship with greater cognitive decline between ages 70 and 79 (β [*SE*] = -0.19 [0.07]).

Research question 4 asked about the relationships of age 11 IQ with allostatic load at age 70 and change allostatic load from age 70 to 79. It was addressed by the regressions of the allostatic load intercept and slope on age 11 IQ. These regression estimates were small and their confidence intervals were wide, so age 11 IQ was not related to older age level and change in allostatic load.

Research question 5 asked whether statistically controlling for age 11 IQ would change the relationships between older age allostatic load and general cognitive ability. It was addressed by comparing the intercept and slope correlation estimates from this model to those from the model without statistical control for age 11 IQ. Comparing these estimates revealed that control for age 11 IQ did not change the older age relationships between allostatic load and general cognitive ability. The correlation between the allostatic load and general cognitive ability intercepts did not notably change (*r* [*SE*] = -0.13 [0.05]) and nor did any of the other correlations between the growth factors.

5.3.3.1 Not missing at random model with control for age 11 IQ

Table 5.4: Allostatic load-general cognitive ability NMAR model results after controlling for age 11 IQ

	Estimate	SE	p-value
g i regression on age 11 IQ	0.71	0.02	< 0.001

Table 5.4: Allostatic load-general cognitive ability NMAR model results after controlling for age 11 IQ *(continued)*

	Estimate	SE	p-value
g s regression on age 11 IQ	-0.21	0.07	0.002
AL i regression on age 11 IQ	-0.03	0.04	0.536
AL s regression on age 11 IQ	0.02	0.06	0.674
AL i-g i correlation	-0.08	0.05	0.071
AL s-g s correlation	-0.17	0.14	0.218
AL i-g s correlation	-0.06	0.11	0.581
AL s-g i correlation	-0.03	0.08	0.663
AL i-AL s correlation	-0.30	0.11	0.005
g i-g s correlation	0.29	0.13	0.021

Note.
Regression betas standardised on all variables. AL = allostatic load; g = general cognitive ability; i = intercept; NMAR = not missing at random; s = slope.

Table 5.4 contains results from the model that retained statistical control for age 11 IQ and added dummy variables for dropout. This addition resulted in only one notable difference in results. It was the same as the difference between the MAR and NMAR models that did not have control for age 11 IQ: the correlation between the allostatic load and general cognitive ability intercepts was attenuated and was no longer statistically significant (*r* [*SE*] = -0.08 [0.05]).

5.3.4 Exploratory latent growth curve models relating allostatic load and specific cognitive tests

Table 5.5: Allostatic load and specific cognitive variables: Exploratory MAR model growth factor correlations

	r	SE	p-value
AL i-block design i correlation	-0.08	0.04	0.087
AL s-block design s correlation	0.08	0.15	0.624
AL i-block design s correlation	-0.18	0.13	0.162
AL s-block design i correlation	-0.03	0.07	0.650
block design i-s correlation	-0.33	0.08	< 0.001
AL i-digit-symbol coding i correlation	-0.18	0.04	< 0.001
AL s-digit-symbol coding s correlation	-0.10	0.09	0.267
AL i-digit-symbol coding s correlation	-0.02	0.08	0.785
AL s-digit-symbol coding i correlation	0.12	0.07	0.083
digit-symbol coding i-s correlation	-0.12	0.07	0.069
AL i-letter-number sequencing i correlation	-0.04	0.05	0.400
AL s-letter-number sequencing s correlation	-0.06	0.12	0.615
AL i-letter-number sequencing s correlation	-0.01	0.09	0.898
AL s-letter-number sequencing i correlation	-0.03	0.07	0.682
letter-number sequencing i-s correlation	-0.34	0.08	< 0.001
AL i-matrix reasoning i correlation	-0.09	0.05	0.072
AL s-matrix reasoning s correlation	-0.05	1.40	0.970
AL i-matrix reasoning s correlation	0.03	1.05	0.978
AL s-matrix reasoning i correlation	-0.01	0.08	0.943
matrix reasoning i-s correlation	0.16	1.71	0.926
AL i-symbol search i correlation	-0.11	0.05	0.026
AL s-symbol search s correlation	-0.50	0.21	0.017
AL i-symbol search s correlation	0.05	0.14	0.737
AL s-symbol search i correlation	0.05	0.07	0.488
symbol search i-s correlation	-0.03	0.16	0.863
AL i-logical memory delayed recall i correlation	-0.07	0.05	0.102
AL s-logical memory delayed recall s correlation	-0.01	0.12	0.919
AL i-logical memory delayed recall s correlation	-0.06	0.09	0.549
AL s-logical memory delayed recall i correlation	0.10	0.07	0.156

Table 5.5: Allostatic load and specific cognitive variables: Exploratory MAR model growth factor correlations *(continued)*

	r	SE	p-value
logical memory delayed recall i-s correlation	0.01	0.08	0.907
AL i-WTAR i correlation	-0.07	0.04	0.092
AL s-WTAR s correlation	-0.02	0.14	0.871
AL i-WTAR s correlation	-0.02	0.12	0.883
AL s-WTAR i correlation	0.08	0.06	0.191
WTAR i-s correlation	0.00	0.10	0.998
AL i-NART i correlation	-0.06	0.04	0.186
AL s-NART s correlation	-0.09	0.13	0.489
AL i-NART s correlation	0.00	0.10	0.977
AL s-NART i correlation	0.07	0.06	0.278
NART i-s correlation	-0.09	0.09	0.300
AL i-verbal fluency i correlation	-0.09	0.04	0.031
AL s-verbal fluency s correlation	-0.37	0.16	0.026
AL i-verbal fluency s correlation	0.06	0.12	0.618
AL s-verbal fluency i correlation	0.15	0.07	0.031
verbal fluency i-s correlation	-0.09	0.09	0.324
AL i-digit span backwards i correlation	-0.01	0.05	0.853
AL s-digit span backwards s correlation	0.20	0.27	0.449
AL i-digit span backwards s correlation	-0.31	0.25	0.209
AL s-digit span backwards i correlation	-0.06	0.07	0.355
digit span backwards i-s correlation	-0.21	0.15	0.144
AL i-spatial span forwards i correlation	-0.01	0.06	0.865
AL s-spatial span forwards s correlation	0.38	0.45	0.398
AL i-spatial span forwards s correlation	-0.01	0.24	0.956
AL s-spatial span forwards i correlation	-0.09	0.09	0.280
spatial span forwards i-s correlation	0.19	0.27	0.486
AL i-spatial span backwards i correlation	-0.09	0.06	0.116
AL s-spatial span backwards s correlation	0.11	0.38	0.773
AL i-spatial span backwards s correlation	-0.19	0.35	0.590
AL s-spatial span backwards i correlation	-0.06	0.09	0.514
spatial span backwards i-s correlation	0.20	0.31	0.512

Table 5.5: Allostatic load and specific cognitive variables: Exploratory MAR model growth factor correlations *(continued)*

	r	SE	p-value
AL i-inspection time i correlation	-0.04	0.05	0.403
AL s-inspection time s correlation	-0.26	0.53	0.631
AL i-inspection time s correlation	-0.50	0.84	0.551
AL s-inspection time i correlation	-0.01	0.08	0.939
inspection time i-s correlation	0.98	2.02	0.629
AL i-SRT i correlation	0.06	0.05	0.208
AL s-SRT s correlation	-0.12	0.17	0.475
AL i-SRT s correlation	0.03	0.13	0.838
AL s-SRT i correlation	0.04	0.08	0.621
SRT i-s correlation	-0.20	0.18	0.251
AL i-CRT i correlation	0.13	0.05	0.005
AL s-CRT s correlation	0.02	0.13	0.908
AL i-CRT s correlation	0.12	0.11	0.293
AL s-CRT i correlation	-0.02	0.07	0.819
CRT i-s correlation	0.32	0.16	0.042
AL i-logical memory immediate recall i correlation	-0.06	0.05	0.184
AL s-logical memory immediate recall s correlation	0.06	0.11	0.605
AL i-logical memory immediate recall s correlation	-0.04	0.09	0.681
AL s-logical memory immediate recall i correlation	0.12	0.07	0.114
logical memory immediate recall i-s correlation	-0.04	0.08	0.576
AL i-Verbal Paired Associates I i correlation	-0.05	0.04	0.204
AL s-Verbal Paired Associates I s correlation	-0.13	0.10	0.208
AL i-Verbal Paired Associates I s correlation	-0.08	0.08	0.326
AL s-Verbal Paired Associates I i correlation	0.10	0.07	0.130
Verbal Paired Associates I i-s correlation	-0.18	0.08	0.021
AL i-Verbal Paired Associates II i correlation	-0.08	0.04	0.084
AL s-Verbal Paired Associates II s correlation	-0.15	0.12	0.220
AL i-Verbal Paired Associates II s correlation	0.01	0.09	0.929
AL s-Verbal Paired Associates II i correlation	0.12	0.07	0.092
Verbal Paired Associates II i-s correlation	-0.12	0.10	0.265
AL i-knowledge i correlation	-0.03	0.04	0.447

Table 5.5: Allostatic load and specific cognitive variables: Exploratory MAR model growth factor correlations *(continued)*

	r	SE	p-value
AL s-knowledge s correlation	0.06	0.11	0.571
AL i-knowledge s correlation	0.02	0.08	0.815
AL s-knowledge i correlation	0.01	0.06	0.870
knowledge i-s correlation	-0.04	0.07	0.549
AL i-nonverbal reasoning i correlation	-0.02	0.05	0.595
AL s-nonverbal reasoning s correlation	0.44	0.41	0.287
AL i-nonverbal reasoning s correlation	-0.14	0.20	0.473
AL s-nonverbal reasoning i correlation	-0.15	0.07	0.041
nonverbal reasoning i-s correlation	-0.87	0.52	0.095
AL i-processing speed i correlation	-0.12	0.05	0.012
AL s-processing speed s correlation	-0.04	0.10	0.657
AL i-processing speed s correlation	-0.03	0.08	0.670
AL s-processing speed i correlation	-0.02	0.07	0.754
processing speed i-s correlation	-0.08	0.07	0.257
AL i-verbal memory i correlation	-0.03	0.04	0.541
AL s-verbal memory s correlation	-0.06	0.13	0.658
AL i-verbal memory s correlation	-0.04	0.10	0.685
AL s-verbal memory i correlation	0.02	0.07	0.714
verbal memory i-s correlation	-0.43	0.08	< 0.001

Note.

AL = allostatic load; i = intercept; s = slope.

We fitted separate exploratory latent growth curve models relating allostatic load growth factors to growth factors for each of 18 cognitive tests and four specific cognitive factors. Table 5.5 contains the results from these models when data were assumed to be missing at random. Most correlations between allostatic load and cognitive growth factors were not statistically significant, although there were some notable relationships. Higher allostatic load at baseline was related to poorer verbal fluency (*r* [*SE*] = -0.09 [0.04]), as well as slower contemporaneously measured processing speed (*r* [*SE*] = -0.12 [0.05]) and worse performance on two processing speed tasks: digit symbol coding (*r* [*SE*] = -0.18 [0.04]) and choice reaction time (*r* [*SE*] = 0.13 [0.05]; higher choice reaction time indicates slower processing and therefore worse performance). Increase in allostatic load over time was related to decline on the symbol search test (*r* [*SE*] = -0.50 [0.21]), which is also a test of processing speed, as well as decline on a verbal fluency task, which taps executive function (*r* [*SE*] = -0.37 [0.16]). Finally, better baseline nonverbal reasoning was related to having a lower allostatic load slope (i.e., less increase or more decrease; *r* [*SE*] = -0.15 [0.07]), while better baseline verbal fluency was related to having a higher allostatic load slope (i.e., more increase; *r* [*SE*] = 0.15 [0.07]).

Table 5.6: Allostatic load and specific cognitive variables: Exploratory MAR model growth factor correlations after controlling for age 11 IQ

	r	SE	p-value
AL i-block design i correlation	-0.06	0.05	0.219
AL s-block design s correlation	0.10	0.16	0.543
AL i-block design s correlation	-0.21	0.14	0.126
AL s-block design i correlation	-0.07	0.07	0.331
block design i-s correlation	-0.20	0.10	0.052
AL i-digit-symbol coding i correlation	-0.18	0.04	< 0.001
AL s-digit-symbol coding s correlation	-0.11	0.09	0.263
AL i-digit-symbol coding s correlation	-0.03	0.08	0.743
AL s-digit-symbol coding i correlation	0.12	0.07	0.098
digit-symbol coding i-s correlation	-0.11	0.07	0.127
AL i-letter-number sequencing i correlation	-0.01	0.05	0.780
AL s-letter-number sequencing s correlation	-0.06	0.12	0.655
AL i-letter-number sequencing s correlation	-0.03	0.09	0.770
AL s-letter-number sequencing i correlation	-0.07	0.08	0.392
letter-number sequencing i-s correlation	-0.28	0.09	0.002

Table 5.6: Allostatic load and specific cognitive variables: Exploratory MAR model growth factor correlations after controlling for age 11 IQ *(continued)*

	r	SE	p-value
AL i-matrix reasoning i correlation	-0.07	0.05	0.151
AL s-matrix reasoning s correlation	-0.04	0.49	0.930
AL i-matrix reasoning s correlation	-0.01	0.38	0.986
AL s-matrix reasoning i correlation	-0.03	0.08	0.708
matrix reasoning i-s correlation	-0.02	0.44	0.970
AL i-symbol search i correlation	-0.10	0.05	0.055
AL s-symbol search s correlation	-0.50	0.21	0.015
AL i-symbol search s correlation	0.04	0.14	0.787
AL s-symbol search i correlation	0.04	0.08	0.582
symbol search i-s correlation	0.07	0.19	0.716
AL i-logical memory delayed recall i correlation	-0.05	0.05	0.286
AL s-logical memory delayed recall s correlation	-0.01	0.12	0.928
AL i-logical memory delayed recall s correlation	-0.06	0.09	0.514
AL s-logical memory delayed recall i correlation	0.10	0.07	0.204
logical memory delayed recall i-s correlation	0.07	0.09	0.480
AL i-WTAR i correlation	-0.05	0.04	0.197
AL s-WTAR s correlation	-0.02	0.15	0.884
AL i-WTAR s correlation	-0.03	0.12	0.830
AL s-WTAR i correlation	0.07	0.07	0.300
WTAR i-s correlation	-0.05	0.10	0.646
AL i-NART i correlation	-0.03	0.04	0.473
AL s-NART s correlation	-0.09	0.13	0.503
AL i-NART s correlation	-0.01	0.10	0.900
AL s-NART i correlation	0.04	0.06	0.499
NART i-s correlation	-0.17	0.09	0.061
AL i-verbal fluency i correlation	-0.08	0.04	0.065
AL s-verbal fluency s correlation	-0.37	0.17	0.024
AL i-verbal fluency s correlation	0.06	0.12	0.632
AL s-verbal fluency i correlation	0.15	0.07	0.032
verbal fluency i-s correlation	-0.07	0.10	0.465
AL i-digit span backwards i correlation	0.02	0.05	0.708

Table 5.6: Allostatic load and specific cognitive variables: Exploratory MAR model growth factor correlations after controlling for age 11 IQ *(continued)*

	r	SE	p-value
AL s-digit span backwards s correlation	0.18	0.23	0.426
AL i-digit span backwards s correlation	-0.29	0.20	0.140
AL s-digit span backwards i correlation	-0.10	0.07	0.173
digit span backwards i-s correlation	-0.16	0.16	0.312
AL i-spatial span forwards i correlation	0.00	0.06	0.941
AL s-spatial span forwards s correlation	0.33	0.36	0.352
AL i-spatial span forwards s correlation	-0.01	0.22	0.949
AL s-spatial span forwards i correlation	-0.11	0.09	0.229
spatial span forwards i-s correlation	0.21	0.26	0.419
AL i-spatial span backwards i correlation	-0.08	0.06	0.197
AL s-spatial span backwards s correlation	0.11	0.43	0.794
AL i-spatial span backwards s correlation	-0.23	0.43	0.603
AL s-spatial span backwards i correlation	-0.08	0.10	0.421
spatial span backwards i-s correlation	0.37	0.35	0.290
AL i-inspection time i correlation	-0.04	0.05	0.497
AL s-inspection time s correlation	-0.40	1.28	0.756
AL i-inspection time s correlation	-0.66	2.00	0.743
AL s-inspection time i correlation	-0.01	0.08	0.889
inspection time i-s correlation	1.23	4.30	0.775
AL i-SRT i correlation	0.05	0.05	0.334
AL s-SRT s correlation	-0.12	0.17	0.455
AL i-SRT s correlation	0.04	0.13	0.761
AL s-SRT i correlation	0.06	0.09	0.524
SRT i-s correlation	-0.16	0.19	0.394
AL i-CRT i correlation	0.12	0.05	0.010
AL s-CRT s correlation	0.02	0.13	0.881
AL i-CRT s correlation	0.11	0.11	0.297
AL s-CRT i correlation	0.00	0.08	0.950
CRT i-s correlation	0.31	0.16	0.050
AL i-logical memory immediate recall i correlation	-0.04	0.05	0.434
AL s-logical memory immediate recall s correlation	0.05	0.11	0.631

Table 5.6: Allostatic load and specific cognitive variables: Exploratory MAR model growth factor correlations after controlling for age 11 IQ *(continued)*

	r	SE	p-value
AL i-logical memory immediate recall s correlation	-0.04	0.09	0.683
AL s-logical memory immediate recall i correlation	0.11	0.08	0.151
logical memory immediate recall i-s correlation	-0.03	0.09	0.764
AL i-Verbal Paired Associates I i correlation	-0.04	0.05	0.387
AL s-Verbal Paired Associates I s correlation	-0.13	0.10	0.185
AL i-Verbal Paired Associates I s correlation	-0.08	0.08	0.333
AL s-Verbal Paired Associates I i correlation	0.10	0.07	0.171
Verbal Paired Associates I i-s correlation	-0.20	0.08	0.012
AL i-Verbal Paired Associates II i correlation	-0.06	0.05	0.174
AL s-Verbal Paired Associates II s correlation	-0.16	0.13	0.199
AL i-Verbal Paired Associates II s correlation	0.01	0.09	0.919
AL s-Verbal Paired Associates II i correlation	0.11	0.07	0.112
Verbal Paired Associates II i-s correlation	-0.15	0.11	0.172
AL i-knowledge i correlation	-0.02	0.04	0.724
AL s-knowledge s correlation	0.06	0.11	0.615
AL i-knowledge s correlation	0.03	0.08	0.746
AL s-knowledge i correlation	-0.01	0.06	0.895
knowledge i-s correlation	-0.13	0.07	0.085
AL i-nonverbal reasoning i correlation	-0.02	0.05	0.727
AL s-nonverbal reasoning s correlation	0.43	0.40	0.276
AL i-nonverbal reasoning s correlation	-0.15	0.20	0.456
AL s-nonverbal reasoning i correlation	-0.16	0.07	0.032
nonverbal reasoning i-s correlation	-0.85	0.49	0.080
AL i-processing speed i correlation	-0.11	0.05	0.014
AL s-processing speed s correlation	-0.05	0.10	0.616
AL i-processing speed s correlation	-0.03	0.08	0.727
AL s-processing speed i correlation	-0.02	0.07	0.734
processing speed i-s correlation	-0.09	0.07	0.210
AL i-verbal memory i correlation	-0.03	0.04	0.523
AL s-verbal memory s correlation	-0.06	0.13	0.621
AL i-verbal memory s correlation	-0.03	0.10	0.742

Table 5.6: Allostatic load and specific cognitive variables: Exploratory MAR model growth factor correlations after controlling for age 11 IQ *(continued)*

	r	SE	p-value
AL s-verbal memory i correlation	0.03	0.07	0.698
verbal memory i-s correlation	-0.43	0.08	< 0.001

Note.
AL = allostatic load; i = intercept; s = slope.

Table 5.6 contains growth factor correlations from exploratory models with age 11 IQ added as a prediction of all growth factors. Results were similar after controlling for age 11 IQ. The correlation between verbal fluency and allostatic load at baseline was slightly attenuated and was no longer statistically significant (*r* [*SE*] = -0.08 [0.04]).

Table 5.7: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations

	r	SE	p-value
AL i-block design i correlation	-0.04	0.04	0.358
AL s-block design s correlation	0.09	0.16	0.562
AL i-block design s correlation	-0.21	0.13	0.113
AL s-block design i correlation	-0.05	0.07	0.432
block design i-s correlation	-0.31	0.09	< 0.001
AL i-digit-symbol coding i correlation	-0.14	0.04	0.001
AL s-digit-symbol coding s correlation	-0.10	0.10	0.299
AL i-digit-symbol coding s correlation	-0.02	0.08	0.757
AL s-digit-symbol coding i correlation	0.09	0.07	0.166
digit-symbol coding i-s correlation	-0.11	0.07	0.085
AL i-letter-number sequencing i correlation	0.00	0.05	0.943
AL s-letter-number sequencing s correlation	-0.05	0.13	0.689
AL i-letter-number sequencing s correlation	-0.03	0.10	0.773
AL s-letter-number sequencing i correlation	-0.06	0.07	0.423
letter-number sequencing i-s correlation	-0.34	0.08	< 0.001

Table 5.7: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations *(continued)*

	r	SE	p-value
AL i-matrix reasoning i correlation	-0.05	0.05	0.352
AL s-matrix reasoning s correlation	0.14	3.01	0.964
AL i-matrix reasoning s correlation	-0.14	3.02	0.963
AL s-matrix reasoning i correlation	-0.03	0.08	0.676
matrix reasoning i-s correlation	0.25	5.96	0.967
AL i-symbol search i correlation	-0.07	0.05	0.148
AL s-symbol search s correlation	-0.52	0.23	0.022
AL i-symbol search s correlation	0.04	0.15	0.803
AL s-symbol search i correlation	0.03	0.07	0.688
symbol search i-s correlation	0.02	0.18	0.909
AL i-logical memory delayed recall i correlation	-0.03	0.05	0.553
AL s-logical memory delayed recall s correlation	-0.01	0.13	0.956
AL i-logical memory delayed recall s correlation	-0.06	0.09	0.505
AL s-logical memory delayed recall i correlation	0.07	0.07	0.328
logical memory delayed recall i-s correlation	0.04	0.08	0.618
AL i-WTAR i correlation	-0.05	0.04	0.238
AL s-WTAR s correlation	-0.02	0.14	0.876
AL i-WTAR s correlation	0.00	0.12	0.967
AL s-WTAR i correlation	0.07	0.06	0.271
WTAR i-s correlation	-0.04	0.09	0.689
AL i-NART i correlation	-0.04	0.04	0.402
AL s-NART s correlation	-0.09	0.13	0.470
AL i-NART s correlation	0.02	0.10	0.845
AL s-NART i correlation	0.05	0.06	0.378
NART i-s correlation	-0.12	0.08	0.148
AL i-verbal fluency i correlation	-0.07	0.04	0.086
AL s-verbal fluency s correlation	-0.38	0.17	0.027
AL i-verbal fluency s correlation	0.07	0.12	0.561
AL s-verbal fluency i correlation	0.14	0.07	0.046
verbal fluency i-s correlation	-0.10	0.09	0.274
AL i-digit span backwards i correlation	0.02	0.05	0.632

Table 5.7: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations *(continued)*

	r	SE	p-value
AL s-digit span backwards s correlation	0.22	0.28	0.429
AL i-digit span backwards s correlation	-0.35	0.27	0.196
AL s-digit span backwards i correlation	-0.08	0.07	0.228
digit span backwards i-s correlation	-0.19	0.16	0.230
AL i-spatial span forwards i correlation	0.02	0.06	0.704
AL s-spatial span forwards s correlation	0.38	0.39	0.334
AL i-spatial span forwards s correlation	-0.04	0.22	0.842
AL s-spatial span forwards i correlation	-0.12	0.09	0.183
spatial span forwards i-s correlation	0.23	0.24	0.336
AL i-spatial span backwards i correlation	-0.05	0.06	0.369
AL s-spatial span backwards s correlation	0.14	0.38	0.702
AL i-spatial span backwards s correlation	-0.24	0.37	0.509
AL s-spatial span backwards i correlation	-0.09	0.09	0.342
spatial span backwards i-s correlation	0.28	0.28	0.325
AL i-inspection time i correlation	-0.01	0.05	0.836
AL s-inspection time s correlation	-0.31	0.76	0.685
AL i-inspection time s correlation	-0.57	1.24	0.644
AL s-inspection time i correlation	-0.02	0.08	0.749
inspection time i-s correlation	1.17	2.99	0.695
AL i-SRT i correlation	0.03	0.05	0.533
AL s-SRT s correlation	-0.14	0.17	0.414
AL i-SRT s correlation	0.05	0.13	0.724
AL s-SRT i correlation	0.06	0.08	0.447
SRT i-s correlation	-0.18	0.19	0.327
AL i-CRT i correlation	0.10	0.05	0.044
AL s-CRT s correlation	0.02	0.14	0.880
AL i-CRT s correlation	0.11	0.11	0.322
AL s-CRT i correlation	0.01	0.07	0.937
CRT i-s correlation	0.36	0.17	0.031
AL i-logical memory immediate recall i correlation	-0.02	0.05	0.724
AL s-logical memory immediate recall s correlation	0.07	0.12	0.550

Table 5.7: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations *(continued)*

	r	SE	p-value
AL i-logical memory immediate recall s correlation	-0.05	0.09	0.619
AL s-logical memory immediate recall i correlation	0.08	0.07	0.250
logical memory immediate recall i-s correlation	-0.01	0.08	0.856
AL i-Verbal Paired Associates I i correlation	-0.02	0.04	0.634
AL s-Verbal Paired Associates I s correlation	-0.13	0.10	0.211
AL i-Verbal Paired Associates I s correlation	-0.08	0.08	0.324
AL s-Verbal Paired Associates I i correlation	0.08	0.07	0.228
Verbal Paired Associates I i-s correlation	-0.18	0.08	0.021
AL i-Verbal Paired Associates II i correlation	-0.04	0.05	0.380
AL s-Verbal Paired Associates II s correlation	-0.16	0.13	0.219
AL i-Verbal Paired Associates II s correlation	0.01	0.09	0.937
AL s-Verbal Paired Associates II i correlation	0.10	0.07	0.168
Verbal Paired Associates II i-s correlation	-0.10	0.11	0.348
AL i-knowledge i correlation	-0.04	0.04	0.423
AL s-knowledge s correlation	0.07	0.12	0.569
AL i-knowledge s correlation	0.01	0.08	0.868
AL s-knowledge i correlation	0.01	0.06	0.836
knowledge i-s correlation	-0.03	0.08	0.646
AL i-nonverbal reasoning i correlation	-0.01	0.05	0.890
AL s-nonverbal reasoning s correlation	0.47	0.47	0.310
AL i-nonverbal reasoning s correlation	-0.19	0.23	0.421
AL s-nonverbal reasoning i correlation	-0.16	0.07	0.029
nonverbal reasoning i-s correlation	-0.88	0.59	0.132
AL i-processing speed i correlation	-0.10	0.05	0.032
AL s-processing speed s correlation	-0.05	0.10	0.591
AL i-processing speed s correlation	-0.02	0.08	0.819
AL s-processing speed i correlation	-0.03	0.07	0.663
processing speed i-s correlation	-0.09	0.07	0.193
AL i-verbal memory i correlation	-0.02	0.05	0.731
AL s-verbal memory s correlation	-0.06	0.13	0.626
AL i-verbal memory s correlation	-0.04	0.10	0.676

Table 5.7: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations *(continued)*

	r	SE	p-value
AL s-verbal memory i correlation	0.02	0.07	0.766
verbal memory i-s correlation	-0.43	0.08	< 0.001

Note.
AL = allostatic load; i = intercept; NMAR = not missing at random; s = slope.

Table 5.7 contains growth factor correlations from the exploratory models which added dummy variables for dropout. The results were similar. Again, the correlation between verbal fluency and allostatic load at baseline was slightly attenuated and was no longer statistically significant (*r* [*SE*] = -0.07 [0.04]).

Table 5.8: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations after controlling for age 11 IQ

	r	SE	p-value
AL i-block design i correlation	-0.03	0.05	0.478
AL s-block design s correlation	0.10	0.16	0.516
AL i-block design s correlation	-0.23	0.14	0.102
AL s-block design i correlation	-0.08	0.07	0.250
block design i-s correlation	-0.19	0.11	0.076
AL i-digit-symbol coding i correlation	-0.15	0.04	0.001
AL s-digit-symbol coding s correlation	-0.10	0.10	0.289
AL i-digit-symbol coding s correlation	-0.03	0.08	0.735
AL s-digit-symbol coding i correlation	0.10	0.07	0.162
digit-symbol coding i-s correlation	-0.10	0.07	0.160
AL i-letter-number sequencing i correlation	0.02	0.05	0.687
AL s-letter-number sequencing s correlation	-0.05	0.13	0.700
AL i-letter-number sequencing s correlation	-0.04	0.10	0.709
AL s-letter-number sequencing i correlation	-0.08	0.08	0.276

Table 5.8: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations after controlling for age 11 IQ *(continued)*

	r	SE	p-value
letter-number sequencing i-s correlation	-0.27	0.09	0.004
AL i-matrix reasoning i correlation	-0.04	0.05	0.414
AL s-matrix reasoning s correlation	0.01	0.44	0.981
AL i-matrix reasoning s correlation	-0.05	0.35	0.894
AL s-matrix reasoning i correlation	-0.05	0.08	0.553
matrix reasoning i-s correlation	0.00	0.44	0.998
AL i-symbol search i correlation	-0.07	0.05	0.165
AL s-symbol search s correlation	-0.53	0.23	0.020
AL i-symbol search s correlation	0.03	0.15	0.819
AL s-symbol search i correlation	0.03	0.08	0.689
symbol search i-s correlation	0.12	0.21	0.567
AL i-logical memory delayed recall i correlation	-0.02	0.05	0.761
AL s-logical memory delayed recall s correlation	-0.01	0.13	0.952
AL i-logical memory delayed recall s correlation	-0.07	0.10	0.490
AL s-logical memory delayed recall i correlation	0.07	0.08	0.350
logical memory delayed recall i-s correlation	0.10	0.10	0.310
AL i-WTAR i correlation	-0.05	0.04	0.221
AL s-WTAR s correlation	-0.02	0.15	0.882
AL i-WTAR s correlation	0.00	0.12	0.989
AL s-WTAR i correlation	0.07	0.07	0.292
WTAR i-s correlation	-0.07	0.10	0.472
AL i-NART i correlation	-0.03	0.04	0.505
AL s-NART s correlation	-0.09	0.13	0.477
AL i-NART s correlation	0.01	0.10	0.902
AL s-NART i correlation	0.05	0.06	0.476
NART i-s correlation	-0.18	0.09	0.043
AL i-verbal fluency i correlation	-0.07	0.04	0.100
AL s-verbal fluency s correlation	-0.38	0.17	0.024
AL i-verbal fluency s correlation	0.07	0.12	0.552
AL s-verbal fluency i correlation	0.15	0.07	0.035
verbal fluency i-s correlation	-0.08	0.10	0.435

Table 5.8: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations after controlling for age 11 IQ *(continued)*

	r	SE	p-value
AL i-digit span backwards i correlation	0.04	0.05	0.435
AL s-digit span backwards s correlation	0.19	0.24	0.418
AL i-digit span backwards s correlation	-0.31	0.21	0.128
AL s-digit span backwards i correlation	-0.11	0.07	0.137
digit span backwards i-s correlation	-0.14	0.16	0.383
AL i-spatial span forwards i correlation	0.03	0.06	0.629
AL s-spatial span forwards s correlation	0.34	0.33	0.308
AL i-spatial span forwards s correlation	-0.04	0.21	0.857
AL s-spatial span forwards i correlation	-0.12	0.09	0.170
spatial span forwards i-s correlation	0.24	0.24	0.320
AL i-spatial span backwards i correlation	-0.05	0.06	0.416
AL s-spatial span backwards s correlation	0.13	0.41	0.746
AL i-spatial span backwards s correlation	-0.26	0.44	0.548
AL s-spatial span backwards i correlation	-0.10	0.10	0.318
spatial span backwards i-s correlation	0.42	0.35	0.226
AL i-inspection time i correlation	-0.01	0.05	0.877
AL s-inspection time s correlation	-0.49	2.19	0.824
AL i-inspection time s correlation	-0.80	3.50	0.819
AL s-inspection time i correlation	-0.03	0.08	0.744
inspection time i-s correlation	1.56	7.51	0.836
AL i-SRT i correlation	0.03	0.05	0.611
AL s-SRT s correlation	-0.14	0.17	0.413
AL i-SRT s correlation	0.05	0.13	0.696
AL s-SRT i correlation	0.07	0.09	0.418
SRT i-s correlation	-0.14	0.20	0.469
AL i-CRT i correlation	0.09	0.05	0.054
AL s-CRT s correlation	0.02	0.14	0.861
AL i-CRT s correlation	0.11	0.11	0.325
AL s-CRT i correlation	0.01	0.08	0.875
CRT i-s correlation	0.35	0.17	0.037
AL i-logical memory immediate recall i correlation	0.00	0.05	0.953

Table 5.8: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations after controlling for age 11 IQ *(continued)*

	r	SE	p-value
AL s-logical memory immediate recall s correlation	0.06	0.12	0.583
AL i-logical memory immediate recall s correlation	-0.05	0.09	0.631
AL s-logical memory immediate recall i correlation	0.09	0.08	0.266
logical memory immediate recall i-s correlation	0.00	0.09	0.991
AL i-Verbal Paired Associates I i correlation	-0.01	0.05	0.756
AL s-Verbal Paired Associates I s correlation	-0.14	0.10	0.186
AL i-Verbal Paired Associates I s correlation	-0.08	0.08	0.335
AL s-Verbal Paired Associates I i correlation	0.08	0.07	0.244
Verbal Paired Associates I i-s correlation	-0.20	0.08	0.012
AL i-Verbal Paired Associates II i correlation	-0.03	0.05	0.464
AL s-Verbal Paired Associates II s correlation	-0.17	0.13	0.196
AL i-Verbal Paired Associates II s correlation	0.01	0.10	0.923
AL s-Verbal Paired Associates II i correlation	0.10	0.07	0.167
Verbal Paired Associates II i-s correlation	-0.14	0.11	0.229
AL i-knowledge i correlation	-0.03	0.04	0.465
AL s-knowledge s correlation	0.06	0.12	0.593
AL i-knowledge s correlation	0.02	0.08	0.835
AL s-knowledge i correlation	0.01	0.06	0.931
knowledge i-s correlation	-0.12	0.08	0.101
AL i-nonverbal reasoning i correlation	0.00	0.05	0.946
AL s-nonverbal reasoning s correlation	0.47	0.45	0.299
AL i-nonverbal reasoning s correlation	-0.19	0.23	0.413
AL s-nonverbal reasoning i correlation	-0.16	0.07	0.025
nonverbal reasoning i-s correlation	-0.86	0.55	0.115
AL i-processing speed i correlation	-0.10	0.05	0.033
AL s-processing speed s correlation	-0.06	0.10	0.573
AL i-processing speed s correlation	-0.02	0.08	0.843
AL s-processing speed i correlation	-0.03	0.07	0.658
processing speed i-s correlation	-0.10	0.07	0.166
AL i-verbal memory i correlation	-0.02	0.05	0.719
AL s-verbal memory s correlation	-0.07	0.13	0.613

Table 5.8: Allostatic load and specific cognitive variables: Exploratory NMAR model growth factor correlations after controlling for age 11 IQ *(continued)*

	r	SE	p-value
AL i-verbal memory s correlation	-0.04	0.10	0.698
AL s-verbal memory i correlation	0.02	0.07	0.757
verbal memory i-s correlation	-0.43	0.08	< 0.001

Note.

AL = allostatic load; i = intercept; NMAR = not missing at random; s = slope.

Table 5.8 contains growth factor correlations from exploratory models with dummy variables for dropout and statistical control for age 11 IQ. Results were similar to the missing at random exploratory models controlling for age 11 IQ. The only difference in the pattern of statistical significance was that baseline allostatic load was not statistically significantly related to slower baseline choice reaction time (*r* [*SE*] = 0.09 [0.05]).

5.4 Discussion

In this study, we fitted latent growth curve models to investigate how allostatic load relates to cognitive ability in older age. Four waves of data were collected between ages 70 and 79. General cognitive ability was modelled as a latent factor based on seven cognitive tests, which enabled fine-grained analysis of non-pathological cognitive ageing. We replicated a relationship between contemporaneously measured allostatic load and poorer general cognitive ability. The effect size was relatively small, but similar to those observed in some other samples (e.g., Seeman et al., 1997). However, we found that baseline allostatic load was higher in those who dropped out at Wave

2 or 3 and baseline general cognitive ability was poorer in those who dropped out at Wave 2, 3 or 4. When this was controlled for, the relationship between contemporaneously measured allostatic load and poorer general cognitive ability was attenuated by around a third. This suggests that the correlation at baseline was partly driven by participants with higher allostatic load and poorer general cognitive ability who went on to drop out of the study at Waves 2-4, many of whom dropped out due to poor health or mortality. It could be that the correlation is affected by a threshold effect, whereby the relationship is stronger beyond certain high levels of allostatic load, or low levels of general cognitive ability. Overall, it seems that this relationship is relatively weak overall and is weaker, if it exists at all, in older adults who are in better health.

We also fitted models in which the growth factors, which represented baseline and change in allostatic load and general cognitive ability in older age, were regressed on age 11 IQ. Previous studies in this sample have found that higher age 11 IQ had a small-to-moderate association with lower allostatic load at age 73 (LBC1936 Wave 2; Booth et al., 2015; Ritchie et al., 2017) and that there was no significant association between age 73 allostatic load and general cognitive change between ages 11 and 73 (Booth et al., 2015). In a novel test, we did not find there was a relationship between age 11 IQ and older age change in allostatic load. In these models, the growth factor correlation estimates were similar to those from models without age 11 IQ. This means that poorer general cognitive change between age 11 and age 70 was related to allostatic load at age 70, with a small effect size. Another way of thinking about this relationship is that at age 70, those who had lower general cognitive ability than was expected based on their age 11 IQ tended to have higher allostatic load. This association could have emerged in various ways. It could be that these individuals' allostatic load was also higher at age 11, and/or that causal links between allostatic load and general cognitive ability have affected each other between ages 11 and 70, and/or that other variables have affected change in general cognitive ability between ages ~11 and 70 as well as allostatic load up to age 70. Further longitudinal research is required to clarify these relationships. Again, after dropout was controlled for, the correlation was attenuated by around a third, which suggested that it was boosted by participants who dropped out at Waves 2-4.

This study also tested how change in allostatic load and change in cognitive ability relate over time. This relationship was estimated as small-to-moderate and negative, but there was a wide confidence interval around the estimate, so the result was inconclusive. More data is required to clarify this relationship.

Originally, we intended to fit a bifactor latent growth curve model to investigate how allostatic load relates to general as well as specific factors of cognitive ability. This bifactor model did not converge, so our confirmatory analyses tested relationships only with a factor of general cognitive ability. Hence, to assess how allostatic load relates to specific cognitive ability factors and test results, we conducted exploratory analyses. These analyses were not preregistered and so they should not be thought of confirming or denying any hypotheses, but rather as providing additional results that can provide the basis for confirmatory tests in other samples. The results from these models suggested that the relationship between allostatic load and processing speed should be a focus of future research. At age 70, higher allostatic load was related to slower processing speed (measured on a specific factor extracted from a bifactor cognitive ability measurement model) as well as poorer performance on two processing speed tests: digit symbol coding and choice reaction time. Additionally, increase in allostatic load between was related to steeper decline in performance on a symbol search task, another test of processing speed. To better understand this relationship, future research could investigate it at a tighter timescale, such as with hours or days rather than years between measurements. Brain white matter may help us to understand the mechanisms underlying the relationship. White matter transports nerve impulses throughout the brain and so its structure is related to processing speed (Kuznetsova et al., 2016). Aspects of white matter structure have been found to be related to allostatic load (Booth et al., 2015; Ottino-González et al., 2018; Savransky et al., 2017). A previous LBC1936 study did not find that the relationship between allostatic load and processing speed was mediated by total white matter volume (Booth et al., 2015), so it may be that more narrow white matter measures must be measured to detect the effects or simply that any effect sizes involved are very small. Although some of our exploratory analyses pointed towards future focuses for research, it is worth noting that most of the exploratory analysis correlations had very small-to-small estimates and had standard errors that were large relative to their estimates.

Overall, our study suggests that while there are relationships between allostatic load and cognitive ability measures, the effect sizes involved tend to be small. Hence, allostatic load summary scores will likely have limited utility as a marker of risk for cognitive ageing.

5.4.1 Strengths and limitations

Different aspects of the study had advantages and disadvantages.

A strength of the study is that gold-standard intelligence tests were administered. These tests assess a broad range of ability and do not have problematic floor or ceiling effects, which meant that we were able to assess how allostatic load relates to non-pathological cognitive ability and cognitive ageing. Additionally, using structural equation modelling enabled us to separate reliable variance from noise, which increased the statistical power of our confirmatory analyses. A limitation of our modelling approach is that the bifactor latent growth curve model we intended to use was too complex and would not converge.

A further strength of our study is that cognitive ability was tested at age 11, which enabled us to test associations with cognitive change between ages 11 and 70, as well as to test the relationship between age 11 IQ and allostatic load in older age.

Our study benefitted from allostatic load being measured at four waves, which enabled us to measure relationships between change in allostatic load and initial level and change in cognitive measures. We calculated allostatic load using the *z* score method, which preserves the continuous nature of biomarkers and has been favoured in a comparison of allostatic load operationalisation methods (Seplaki et al., 2005). A limitation of our allostatic load operationalisation is that no HPA axis biomarkers were included, as these were not measured in this study.

Our sample size was large enough to enable a structural equation modelling approach with measurement invariance testing, and reasonable statistical power for some pathways, but it was not large enough to offer adequate power for some pathways, including many involving change over time.

An additional strength of our study is that we fitted models using both the missing at random and not missing at random assumptions about the missing data. More often, studies assume that the missing data are missing at random and do not fit models that assume the missing data are not missing at random. Comparing models fitted using the different assumptions enabled us to assess how our results related to participant dropout.

5.4.2 Directions for future research

Our study provided a novel test of dual change in allostatic load and general cognitive ability. This parameter, though, was imprecisely estimated, so further tests with larger samples and more time points are required to clarify this relationship. Statistical power to detect correlated change can increase substantially with each additional measurement, particularly when there have been fewer than 10 measurements (Hertzog, Lindenberger, Ghisletta, & Oertzen, 2006). It may also be profitable for researchers to rerun longitudinal analyses after new waves of data are released, so that the estimates are updated and the precision of the estimates is increased.

As noted above, a limitation of our study is that no HPA axis markers were used as indicators of allostatic load. It would be informative to compare our results to those from another study with a comparable design that did use HPA axis markers as indicators of allostatic load. Such a study could compare the results of models using two different allostatic load operationalisations: one calculated with HPA axis markers and one without. The study could then test whether the relationship between allostatic load and cognitive ability is stronger, weaker or similar when HPA axis markers are used.

The results of our non-preregistered exploratory analyses could provide the basis for preregistered confirmatory analyses. Future studies, then, could attempt to confirm the association between allostatic load and measures of processing speed in other samples.

Future longitudinal studies should carefully consider how participant dropout relates to the levels and trajectories of allostatic load and cognitive ability, as well how it relates to the relationships being assessed. To find out why modelling dropout affects the relationship between allostatic load and general cognitive ability, future studies could use threshold regression to find out whether the relationship changes beyond certain levels of each variable, and/or quantile regression to test the strength of the association at different levels of each variable. The present study benefitted from the use of pattern-mixture modelling, which assumes that missing data are not missing at random. Future research may wish to take this or another approach, such as one that also models the reasons for dropout.

5.4.3 Conclusion

In summary, this study tested the relationship of allostatic load and cognitive ability between ages 70 and 79. Our study went beyond previous research by measuring both allostatic load and cognitive ability at four waves, which meant that we could test associated change relationships. We found that allostatic load was associated with lower general cognitive ability at age 70, with a small effect size. However, this association was partly driven by participants dropped out at Waves 2-4 and it was attenuated by around a third after dropout was taken into account. We also found that allostatic load at age 70 was related to poorer general cognitive change between ages 11 and 70, with a small effect size. Regarding specific cognitive ability factors, exploratory analyses suggested that allostatic load may be related to slower processing speed. Finally, the association between dual change in allostatic load and general cognitive ability was estimated as small-to-moderate and negative but it was estimated imprecisely, so further research is required to clarify this relationship.

Chapter 6

How do allostatic load and depressive symptoms relate over time in the Lothian Birth Cohort 1936?

6.1 Introduction

In late life, depressive symptoms are common (Luppa et al., 2012) and often occur in conjunction with physical ill health (Büchtemann, Luppa, Bramesfeld, & Riedel-Heller, 2012). It has been suggested that understanding the relationship between summary measures of allostatic load and depressive symptoms could help explicate the potential mediating role of allostatic load in the relationship between depression and adverse physical health outcomes (Penninx, 2017). Further, it has been put forward that understanding the relationship between allostatic load and depressive symptoms could help to improve the diagnosis and treatment of depressive symptoms and physiological dysregulation, enabling more personalised and effective treatment

(Lopresti et al., 2014). However, despite such suggestions, few studies to date have investigated the relationship between allostatic load and depressive symptoms, and even fewer have tested this relationship longitudinally.

Here, we will first introduce allostatic load and depressive symptoms. Then, we will discuss previous cross-sectional and longitudinal research relating allostatic load and depressive symptoms. Finally, we will introduce the present study, which will test the contemporaneous and longitudinal relationships between allostatic load and depressive symptoms in older age.

6.1.1 Depressive symptoms

Depressive symptoms include low mood, low energy, hopelessness, irritability and suicidality (National Health Service, 2016). Depressive disorders are a major public health issue (Cassano & Fava, 2002; Marcus, Yasamy, van Ommeren, Chisholm, & Saxena, 2012) and non-clinical levels of depressive symptoms are thought to “[represent] a less severe manifestation of the same disorder” (Luppa et al., 2012, p. 217). It is important, therefore, to study depressive symptoms as a continuous variable. Compared to earlier in life, in older age “less severe forms of depression seem to be far more frequent” (Büchtemann et al., 2012, p. 177). One of the main variables related to the development of depressive symptoms in older age is physical ill health (Büchtemann et al., 2012).

6.1.2 Allostatic load biomarkers and depressive symptoms

In efforts to better understand the relationship between depressive symptoms and physical ill health, research has investigated the relationships between depressive symptoms and physiological markers. It is well established that certain putative allostatic load biomarkers tend to be dysregulated in people with depression. This includes endocrine (Murri et al., 2014), inflammatory (Valkanova et al., 2013), and metabolic (Marazziti et al., 2014) biomarkers. Physiological dysregulation in people with depression is not uniform, as profiles of dysregulation differ between subtypes of depression (Lamers et al., 2013). One study found that compared to control participants and those with severe melancholic depression (characterised by lowered appetite and decreased weight), participants with severe atypical depression (which is characterised by excessive eating and increased weight) had higher inflammation and greater metabolic dysregulation (Lamers et al., 2013). Also, those with severe melancholic depression had higher cortisol production during the waking period and a higher diurnal cortisol slope than control participants and those with severe atypical depression (Lamers et al., 2013).

6.1.3 Allostatic load summary scores and contemporaneously measured depressive symptoms

Allostatic load summary measures have been statistically significantly related to contemporaneously measured depressive symptoms in several studies. For example, in a small sample of 58 older adults aged 52-80 in Quebec, Canada, allostatic load explained 20% of the variance in depressive symptoms (data from the Douglas Hospital Longitudinal Study of Normal and Pathological Aging; participants were healthy at baseline in 1988, this contemporaneous relationship was tested using 1994 data; Juster et al., 2011). Also, a clinically relevant effect was found in a sample of 125 community-dwelling older adults aged 67 to 94, where allostatic load was found to be related to depressive symptoms in adjusted models and the effect size was "comparable (but opposite in direction) to that of antidepressant use" (data from former participants of the Mindfulness to Improve Elders' Immune and Health Status [MIEIHS] study; Kobrosly et al., 2014, p. 223). In sum, strong effects have been reported in smaller samples. However, while studies using larger samples have also reported finding a relationship between allostatic load and depressive symptoms, they have reported smaller effect sizes than small sample studies.

In a larger sample of 705 adults in the 1972-born cohort of the West of Scotland Twenty-07 Study, allostatic load had a small correlation with depressive symptoms at age 36 (r = .167; Gale et al., 2015). Plus, in a sample of 958 adults in Taiwan (M [SD] age = 68 [8] y), models adjusted for age and sex with different allostatic load operationalisations predicted between 2.9% and 5.0% of the variance in depressive symptoms (Social Environment and Biomarkers of Aging Study [SEBAS]; Seplaki et al., 2005). Finally, in a large sample of 2405 older adults aged 60 and above in the United States, allostatic load had a small relationship with depressive symptoms in adjusted models (data from the National Health and Nutrition Examination Survey [NHANES] 2005-2008; Kobrosly et al., 2013). In a fully adjusted model, having 1 additional allostatic load biomarker beyond a clinical cutoff point was associated with a score 0.23 higher on the Patient Health Questionnaire-9, which measures depressive symptoms (nine-item questionnaire with a possible score range of 0-27; Kobrosly et al., 2013). To summarise, allostatic load summary measures have consistently been related to contemporaneously measured depressive symptoms, although the effect sizes reported have been smaller in studies using larger samples.

6.1.4 Allostatic load summary scores and subsequently measured depressive symptoms

Little research to date has investigated how allostatic load and depressive symptoms relate over time.

In the aforementioned sample in Taiwan, baseline allostatic load predicted depressive symptoms measured three years later, and this relationship remained statistically significant after adjusting for baseline depressive symptoms and other baseline health covariates (n = 926; Goldman et al., 2006). In this study, 1 additional allostatic biomarker in the highest or lowest sample decile predicted a score 0.41 higher on the CES-D-10 (the 10-item Center for Epidemiologic Studies Depression Scale, which has a possible score range of 0-30; Goldman et al., 2006). The effect size, though, was roughly halved after controlling for baseline CES-D and other baseline health covariates, as then the same increase predicted a score 0.20 higher on the CES-D-10 (Goldman et al., 2006). The relationship between allostatic load and subsequently measured depressive symptoms was not always statistically significant, though, in the aforementioned small-sample Quebec-based study. Allostatic load explained 7% of the variance in depressive symptoms measured three years later (*n* = 58), but only 1% of the variance in depressive symptoms measured six years later (beta non-significant; *n* = 32; Juster et al., 2011).

Research has only very recently begun to study how allostatic load predicts responses to depression medication. A recent study measured baseline allostatic load in 67 healthy controls and 34 participants with depression who were about to under go eight weeks of selective serotonin reuptake inhibitor (SSRI) treatment (Hough et al., 2017). Compared to healthy controls and participants with depression who experienced more than 50% improvement in depression, there was higher baseline allostatic load in participants with depression who experienced lower than 50% improvement in depression severity during treatment (Hough et al., 2017).

Research has not yet fully explored how allostatic load and depressive symptoms relate over time in older age. Notably, the relationship between change in allostatic load and change in depressive symptoms has not been tested, and nor has the relationship between depressive symptoms and subsequent change in allostatic load.

6.1.5 The present study

In the present study, we fitted latent growth curve models to data from the Lothian Birth Cohort (LBC1936) study, which has measured allostatic load and depressive symptoms in a large sample at approximately ages 70, 73, 76 and 79. The main question being asked in this study is: how do allostatic load and depressive symptoms relate over time?

The first two specific research questions being asked in this study are:

1. How strongly do allostatic load and depressive symptoms relate at age 70?
2. Does allostatic load at age 70 relate to change in depressive symptoms between ages 70 and 79?

The sample size of the LBC1936 enabled our study to more precisely estimate contemporaneous relationships than many previous studies (research question 1). The LBC1936 study data also enabled us to investigate the relationship between allostatic load and change in subsequently measured depressive symptoms over more time points and a longer period of time than previous studies (research question 2). Also, crucially, having four waves of data for both variables enabled us to ask research questions about two previously unexplored relationships:

3. Do depressive symptoms at age 70 relate to change in allostatic load between ages 70 and 79?
4. Between ages 70 and 79, does change in allostatic load correlate with change in depressive symptoms?

Previous research suggests that plausibly, the correlations involved are more likely to be positive, meaning higher allostatic load is related to more depressive symptoms. Plausible effect sizes are between approximately zero and moderate in size; the effect sizes are unlikely to be large.

6.2 Methods

6.2.1 Measures

6.2.1.1 Depressive symptoms

Depressive symptoms were measured using the Hospital Anxiety and Depression Scale-Depression (HADS-D) subscale (Zigmond & Snaith, 1983). The HADS-D has seven items, which are scored from zero to three using different ordinal response options (Zigmond & Snaith, 1983). Because the HADS-D was designed to be used in nonpsychiatric hospital inpatient settings, the questions focus on anhedonia symptoms and do not cover somatic symptoms, such as insomnia and fatigue (Zigmond & Snaith, 1983). In our models, we used the total HADS-D score, which is the sum of the item scores.

6.2.2 Statistical analysis

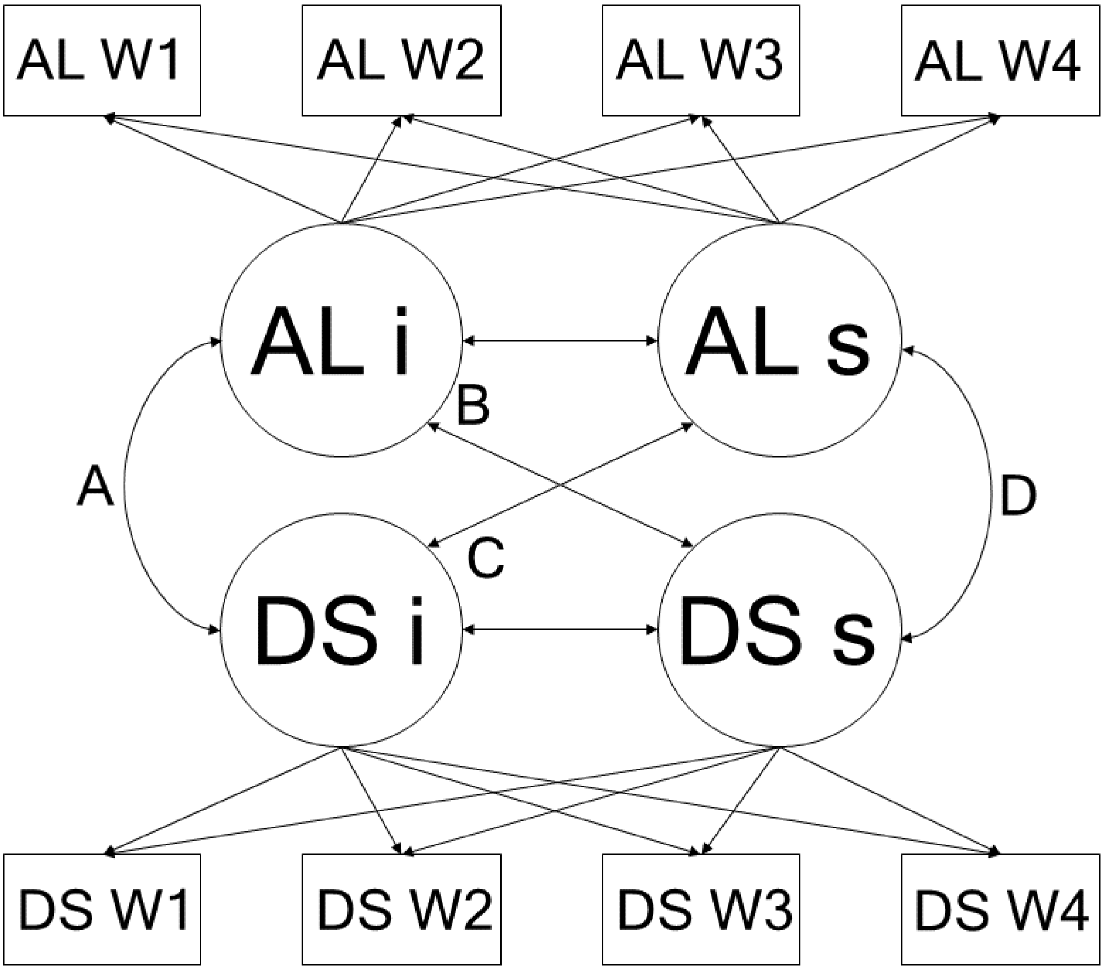

Figure 6.1: Allostatic load-depressive symptoms latent growth curve model specification. A-D are labelled paths described in the text. AL = allostatic load. DS = depressive symptoms. i = intercept. s = slope. W = wave.

In this study, the parallel process latent growth curve models included growth factors for two variables: allostatic load and depressive symptoms. Figure 6.1 shows the general specification of the dual change latent growth curve models.

Research question 1 was tested by the intercept-intercept correlation (path A in Figure 6.1). The next two research questions were tested by the intercept-slope correlations. Question 2 was tested by the correlation between the allostatic load intercept and the depressive symptoms slope (path B). Question 3 was tested by the correlation between depressive symptoms intercept and the allostatic slope tested question 3 (path C). Finally, question 4 was tested by the correlation between the allostatic load and depressive symptoms slopes (path D).

6.3 Results

Table 6.1: Descriptive statistics for age, allostatic load and depressive symptoms

	n	Mean	SD	Skew	Kurtosis
Age W1	1091	69.53	0.83	-0.05	-0.88
Age W2	866	72.49	0.71	-0.02	-0.84
Age W3	697	76.25	0.68	-0.05	-0.83
Age W4	550	79.32	0.62	0.06	-0.76
Allostatic load W1	1090	0.69	0.24	1.07	2.12
Allostatic load W2	866	0.71	0.24	1.17	3.17
Allostatic load W3	696	0.85	0.25	0.71	1.16
Allostatic load W4	549	0.87	0.26	0.62	0.48
HADS-D W1	1086	2.80	2.23	1.35	2.30
HADS-D W2	865	2.63	2.22	1.37	1.95
HADS-D W3	696	2.88	2.32	1.22	1.24
HADS-D W4	548	2.98	2.34	1.35	2.35

Note.

HADS-D = Hospital Anxiety and Depression Scale - Depression subscale.

Table 6.1 contains descriptive statistics for age, allostatic load and depressive symptoms at each wave. The allostatic load mean increased only very slightly between Waves 1 and 2 and between Waves 3 and

4, with the largest increase by far being between Waves 2 and 3. The depressive symptoms mean decreased between Waves 1 and 2 but increased between Waves 2 and 4.

Table 6.2: Allostatic load and depressive symptoms correlations

	AL W1	AL W2	AL W3	AL W4	DS W1	DS W2	DS W3	DS W4
AL W1		< .001	< .001	< .001	< .001	.03	.24	.83
AL W2	.45 (.03)		< .001	< .001	.14	.12	.24	.29
AL W3	.31 (.04)	.46 (.03)		< .001	.74	.83	.24	.74
AL W4	.27 (.04)	.40 (.04)	.47 (.04)		.12	.24	.34	.30
DS W1	.13 (.03)	.09 (.03)	.04 (.04)	.11 (.04)		< .001	< .001	< .001
DS W2	.11 (.03)	.09 (.03)	.03 (.04)	.10 (.04)	.65 (.03)		< .001	< .001
DS W3	.08 (.04)	.09 (.04)	.09 (.04)	.08 (.04)	.62 (.03)	.68 (.03)		< .001
DS W4	.02 (.04)	.09 (.04)	.06 (.04)	.08 (.04)	.61 (.03)	.62 (.03)	.72 (.03)	

Note.

Lower triangle contains pairwise Pearson's correlations, with standard errors in parentheses (n range = 536-1090). Upper triangle contains p-values for the correlations. AL = allostatic load. DS = depressive symptoms. W = Wave.

Table 6.2 contains correlations between the allostatic load and depressive symptoms variables. Allostatic load levels at adjacent waves were moderately-to-strongly positively correlated, while allostatic load levels at non-adjacent waves were moderately positively correlated. Depressive symptoms at different waves were strongly positively correlated. Contemporaneously measured allostatic load and depressive symptoms had a modest positive correlation. Allostatic load and depressive symptoms variables measured at different waves had positive correlations. all of which were small or very small in size.

6.3.1 One process latent growth curve models

Table 6.3: Unconditional one process MAR model results for allostatic load and depressive symptoms: Growth factor means and variances

	Estimate	SE	p-value
AL i mean	0.682	0.007	< 0.001
AL i variance	0.029	0.004	< 0.001
AL s mean	0.0211	0.0012	< 0.001
AL s variance	0.0003	0.0001	< 0.001
HADS-D i mean	2.73	0.07	< 0.001
HADS-D i variance	3.45	0.31	< 0.001
HADS-D s mean	0.039	0.008	< 0.001
HADS-D s variance	0.014	0.005	0.003

Note.
AL = allostatic load. HADS-D = Hospital Anxiety and Depression Scale - Depression subscale. i = intercept [growth factor]. s = slope [growth factor].

Table 6.3 contains the growth factor means and variances from the unconditional allostatic load and depressive symptoms growth curve models. As these models were unconditional, no covariates were included. Here, the intercept mean represents the average level at baseline (age 70) and the slope mean represents the average change per year between ages 70 and 79, while the variances represent individual variation around the mean levels and trajectories. Notably, the average participant increased in allostatic load and depressive symptoms over the course of the study, but there was significant variation in the trajectories observed for both variables.

Table 6.4: Conditional one process MAR model results for allostatic load and depressive symptoms: Growth factor and regression estimates

	Estimate	SE	p-value
AL W1 on age W1 regression	0.008	0.009	0.369
AL W2 on age W2 regression	0.024	0.011	0.022
AL W3 on age W3 regression	0.045	0.014	0.001
AL W4 on age W4 regression	0.005	0.017	0.755
AL i on male regression	-0.015	0.082	0.852
AL s on male regression	-0.031	0.127	0.807
HADS-D W1 on age W1 regression	0.07	0.07	0.364
HADS-D W2 on age W2 regression	0.26	0.09	0.004
HADS-D W3 on age W3 regression	0.27	0.11	0.011
HADS-D W4 on age W4 regression	0.22	0.12	0.076
HADS-D i on male regression	0.09	0.07	0.206
HADS-D s on male regression	-0.13	0.13	0.302
AL i mean	0.683	0.010	< 0.001
AL i variance	0.028	0.003	< 0.001
AL s mean	0.0214	0.0016	< 0.001
AL s variance	0.0003	0.0001	< 0.001
HADS-D i mean	2.65	0.09	< 0.001
HADS-D i variance	3.44	0.31	< 0.001
HADS-D s mean	0.047	0.012	< 0.001
HADS-D s variance	0.014	0.005	0.003

Note.
AL = allostatic load. HADS-D = Hospital Anxiety and Depression Scale - Depression subscale. i = intercept [growth factor]. s = slope [growth factor]. Regressions on male are estimates with the latent variables standardised and the male variable coded as 1 for males and 0 for females. For all other parameters, raw estimates are provided.

Table 6.4 contains the results from conditional models, which added sex as a time-invariant covariate of the growth factors and age at each wave as a time-varying covariate of the corresponding allostatic load and depressive symptoms variables. Here, the intercept and slope means represent the estimates for a female participant who was the average age of observation at each wave, and the male regression coefficients can be used to obtain the estimates for other participant profiles. Within Waves 2 and 3, older age at the time of testing was modestly related to higher allostatic load and more depressive symptoms. Other covariates did not have notable effects.

Table 6.5: Conditional one process NMAR model results for allostatic load and depressive symptoms: Pattern-mixture model regressions, growth factor means and growth factor variances

	Estimate	SE	p-value
AL i on d2 regression	0.38	0.11	0.001
AL i on d3 regression	0.50	0.13	< 0.001
AL i on d4 regression	0.16	0.12	0.171
AL s on d2/d3 regression	-0.73	0.32	0.023
AL s on d4 regression	-0.05	0.20	0.800
HADS-D i on d2 regression	0.35	0.10	0.001
HADS-D i on d3 regression	0.19	0.10	0.062
HADS-D i on d4 regression	0.03	0.10	0.791
HADS-D s on d2/d3 regression	0.11	0.34	0.747
HADS-D s on d4 regression	0.44	0.23	0.062
AL i mean	0.655	0.012	< 0.001
AL i variance	0.028	0.003	< 0.001
AL s mean	0.0235	0.0017	< 0.001
AL s variance	0.0003	0.0001	< 0.001
HADS-D i mean	2.45	0.11	< 0.001
HADS-D i variance	3.36	0.31	< 0.001
HADS-D s mean	0.047	0.012	< 0.001
HADS-D s variance	0.014	0.005	0.004

Note.

AL = allostatic load. HADS-D = Hospital Anxiety and Depression Scale - Depression subscale. i = intercept [growth factor]. s = slope [growth factor]. Growth factor means and variances are raw estimates. Regressions are estimates with the latent variables standardised and the d2-d4 dummy variables coded as 1 for those who dropped out at Waves 2-4 respectively and 0 otherwise.

Table 6.5 contains the results from models that assumed that missing data were *not* missing at random. These were pattern-mixture models in which the growth factors were regressed on dummy variables d2-d4 that were coded as 1 if the participant dropped out at the corresponding wave and 0 otherwise. The regressions of the slope growth factors on d2 and d3 were constrained to equality to identify the models. The dropout variable regressions have been standardised on the latent variables, so the estimates presented are in standard deviation units.

Note that here, the growth factor means are the expected values for participants who did not drop out. Dropping out at Wave 2 or 3 was related to having higher allostatic load at Wave 1 (alternatively phrased, participants with higher allostatic load at Wave 1 were more likely to drop out at Wave 2 or 3). These participants also had a much lower allostatic load slope (i.e., a flatter or even declining trajectory), although this parameter had a wide confidence interval. This could seem a counter-intuitive finding. However, it becomes more interpretable after considering that the mean allostatic load only rose slightly between Waves 1 and 2 before a large rise between Waves 2 and 3, where these participants had at least some missing data.

Additionally, dropping out at Wave 2 was related to scoring higher on depressive symptoms at Wave 1.

6.3.2 Parallel process latent growth curve models

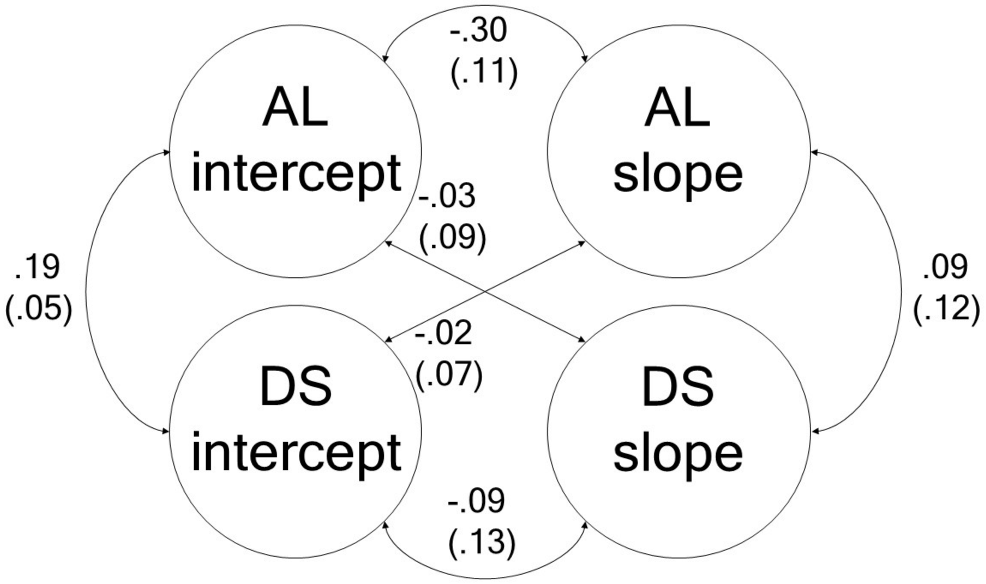

Figure 6.2: Allostatic load-depressive symptoms MAR model growth factor correlations. Correlations are displayed, with their standard errors in parentheses. AL = allostatic load. DS = depressive symptoms.

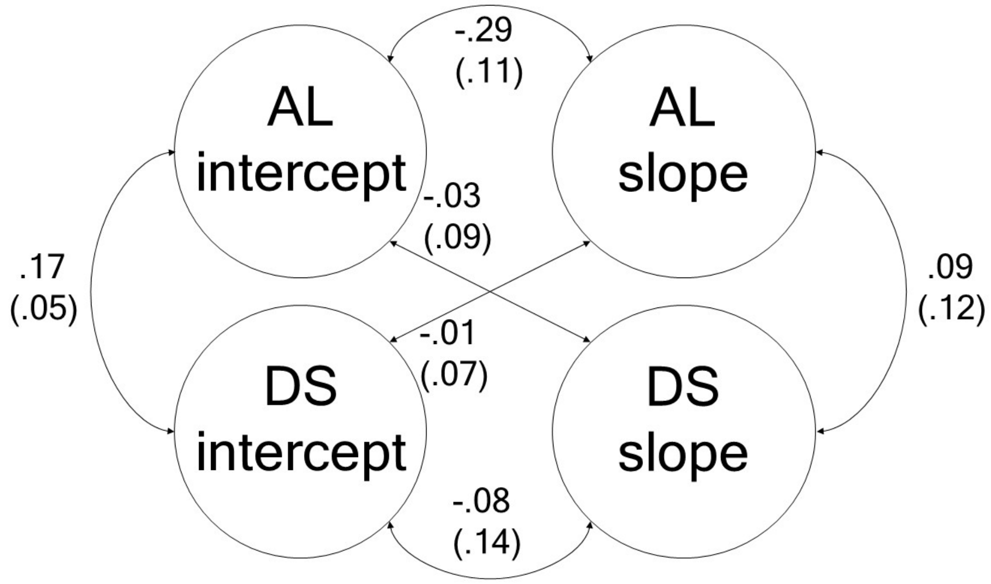

Figure 6.3: Allostatic load-depressive symptoms NMAR model growth factor correlations. Correlations are displayed, with their standard errors in parentheses. AL = allostatic load. DS = depressive symptoms.

Figures 6.2 and 6.3 display the results from parallel process latent growth curve models with the MAR and NMAR assumptions respectively. The growth factor correlations were similar in the MAR and NMAR models. Initial level in allostatic load was moderately correlated with a more negative (i.e., flatter rather than increasing) allostatic load slope.

The intercept-intercept correlation addresses research question 1 about contemporaneous relationships at age 70. At baseline, higher allostatic load was associated with reporting more depressive symptoms, with a small-to-moderate effect size. Research questions 2-4 were addressed by the other three growth factor correlations in the model. All of these estimates were small and had standard errors that were large relative to the parameter estimates. Our results suggested that allostatic load and depressive symptoms are not strongly related over time.

Table 6.6: Model fit statistics for allostatic load and depressive symptoms

Model	CFI	RMSEA	SRMR
AL MAR unconditional	0.904	0.098	0.054
AL MAR conditional	0.905	0.051	0.028
AL NMAR conditional	0.901	0.045	0.037
HADS-D MAR unconditional	0.985	0.063	0.033
HADS-D MAR conditional	0.991	0.025	0.018
HADS-D NMAR conditional	0.985	0.027	0.030
Parallel process MAR	0.964	0.037	0.026
Parallel process NMAR	0.959	0.035	0.038

Note.

AL = allostatic load. CFI = comparative fit index. HADS-D = Hospital Anxiety and Depression Scale - Depression subscale. RMSEA = root mean square error of approximation. SRMR = square root mean residual. For CFI and RMSEA, robust variants were used.

Table 6.6 contains model fit statistics from all models fitted. The unconditional allostatic load model had borderline satisfactory fit. The conditional allostatic load models had good fit according to the *RMSEA* and *SRMR*, but borderline satisfactory fit according to the *CFI*. The depressive symptoms and parallel process models had excellent fit.

6.4 Discussion

Our study investigated the contemporaneous and longitudinal relationships between allostatic load and depressive symptoms in older age. Importantly, our study had a large, age-homogeneous sample and four waves of data for both measures. This enabled us to conduct a novel test of the relationship between change in allostatic load and change in depressive symptoms.

We found that at baseline, higher allostatic load was correlated with reporting more depressive symptoms, with a small-to-moderate effect size. This replicated previous findings but was one of the smaller published estimates of the effect, in line with the estimates from other studies that have used larger samples (Kobrosly et al., 2013; Seplaki et al., 2005).

We also refitted our model assuming missing data were not missing at random. We found that those who dropped out at Wave 2 and 3 had higher allostatic load and reported more depressive symptoms at baseline. In this model, the baseline correlation between allostatic load and depressive symptoms was of a similar size and remained statistically significant with an alpha of .05. This suggests that the baseline correlation was not strongly driven by participants, perhaps in poorer health, who later dropped out of the study.

Furthermore, we also provided unadjusted correlations between contemporaneously measured allostatic load and depressive symptoms at each of four waves. Higher allostatic load was correlated with reporting depressive symptoms at each wave, but the effect size was consistently small (*r* range = .08 to .13).

Our novel test of the relationship between change in allostatic load and change in depressive symptoms produced small positive estimates. Standard errors for these tests were wide and the relationship was not statistically significant. Our results suggest that there is not a strong dual change relationship. Further research using larger samples and improved measures is required to determine the presence and strength of this relationship.

The correlation table for allostatic load and depressive symptoms at Waves 1-4 in our study provides places for comparison with some previous longitudinal studies. Goldman et al. (2006) and Juster et al. (2011) found that allostatic load was related to depressive symptoms measured three years later, with a moderate strength effect size. In our study, the waves also had roughly three-year gaps. Regarding the lagged relationships, Wave 1 and 2 allostatic load were correlated with Wave 2 and 3 depressive symptoms respectively, although both correlations had small effect sizes. However, Wave 3 allostatic load and Wave 4 depressive symptoms were not significantly correlated.

Previously, Juster et al. (2011) found that allostatic load explained 1% of the variance in depressive symptoms measured six years later. This test was in a small sample of only 32 and the relationship was not statistically significant (Juster et al., 2011). In our study, the correlations between allostatic load and depressive symptoms measured six years later had a similarly small effect size (*r*s = .08 and .09), but due to our larger sample size they were statistically significant with an alpha of .05. Note that the correlation is likely non-zero and positive: whether it is found to be statistically significance depends in part on the sample size of the study. Studies should focus not on confirming the presence of the relationship, but instead on clarifying its strength.

Altogether, the results of our study echo previous results that higher allostatic load is related to greater depressive symptomatology. However, like other studies that have used larger samples, we found the relationship consistently had only a small effect size. In their study, Juster et al. (2011) posited that “using the AL index could allow for the development of individualized interventions in order to prevent the emergence and/or maintenance of depressive symptoms” (Juster et al., 2011, p. 363). Our results suggest that allostatic load assessments would only have a modest contribution to such interventions. Considering the cost of measuring many allostatic load markers, it is unlikely that the widespread administration of allostatic load tests would be a cost-effective depressive symptoms prevention strategy. If allostatic load biomarkers are to be useful in treating depressive symptoms, it may be as part of an approach that considers patterns of physiological dysregulation and depressive symptoms, rather than the overall levels of each (Lamers et al., 2013). Methods that consider the differential risk conferred by different biomarkers, unlike most allostatic load operationalisations, could also help to improve prediction of depressive symptoms and bolster prevention efforts. For example, a recent study applied machine learning to cohort study and identified additional novel biomarkers associated with depressive symptoms (Dipnall et al., 2016), including the allostatic load biomarker serum glucose (Juster et al., 2010). Machine learning could likely be effectively applied to predicting depressive symptoms electronic health records; a well-calibrated algorithm would be able to use the data to improve prediction more than the modest benefits in prediction given by an allostatic load summary measure.

6.4.1 Strengths and limitations

Our study benefitted from having four waves of longitudinal data for both allostatic load and depressive symptoms. This enabled us to fit bivariate latent growth curve models that went beyond previous research by testing the relationship between change in allostatic load and change in depressive symptoms.

Furthermore, using full information maximum likelihood estimation enabled us to use all our data and avoid producing biased estimates, which can result from complete case analyses (Enders, 2010). We also fitted pattern-mixture models that assumed data were not missing at random, which showed that some growth factor estimates differed for participants who dropped out of the study, but also confirmed that our estimates of allostatic load-depressive symptoms relationships were not biased by including these participants in our models.

A limitation of our study is that the biomarker panel used to operationalise allostatic load did not have any HPA axis markers, nor any primary mediators of the stress response. The biomarkers used were instead secondary outcomes (biomarkers whose levels change in response to primary mediator changes) and were mostly metabolic or cardiovascular biomarkers.

A further limitation is that only sum scores were available for our depressive symptoms measure, the HADS-D, which precluded analysing allostatic load's relationships with specific symptoms and groups of symptoms. This also precluded analysis methods that require item- and interitem-level data, such as factor analysis. Additionally, the combined HADS has been criticised for having a latent structure that varies based on the methods used to detect it (Cosco, Doyle, Ward, & McGee, 2012) and for primarily measuring general distress rather than depression and anxiety, as its subscales are intended to do (Norton, Cosco, Doyle, Done, & Sacker, 2013). This may in part be due to four problematic items that seem to measure restlessness rather than anxiety or depression (Straat, van der Ark, & Sijtsma, 2013).

6.4.2 Directions for future research

Our study used an age-homogeneous sample who were observed between approximately 70 to 79 years old. This means our study has a strong focus on the relationships between allostatic load and depressive symptoms in people's seventies, but our findings may not generalise to other age groups. As stress affects allostatic load and depressive symptoms throughout the lifespan (Danese & McEwen, 2012), the relationship between allostatic load and depressive symptoms is likely to also be found in childhood and earlier adulthood. However, the effect size may vary depending on age. Future studies using mixed-age samples could investigate this possibility by testing whether the contemporaneous and longitudinal relationships between allostatic load and depressive symptoms are moderated by age.

Future studies on the allostatic load-depressive symptoms relationships should aim to clarify the presence and strength of their dual change relationship by using more data points than were used in this study. This could involve having more participants and/or having both measures at more time points (Hertzog et al., 2006).

Our study, and others, have used gaps of around three years between measures. It would be beneficial to study this relationship at different timescales, for example with hours, days or weeks between measurements, rather than years. For example, a recent stress reactivity study collected cortisol and stress questionnaire data up to 10 times a day for six consecutive days (Vaessen et al., 2018). A similar approach could be taken with depressive symptoms items and multiple allostatic load biomarkers being measured throughout the day and throughout the week, to better understand how the two relate on shorter timescales.

In addition, future research would benefit from analysing specific symptoms and subcategories of symptoms, as some previous research has done (e.g, Kobrosly et al., 2013), to further clarify which depressive symptoms allostatic load is more strongly related to.

6.4.3 Conclusion

In summary, we replicated the contemporaneous relationship between higher allostatic load and more depressive symptoms in a large sample of older age adults. In line with previous large studies, the effect size was small. Using four waves of data over ~10 years, we conducted a novel test of the correlation between change in allostatic load and change in depressive symptoms. The estimate was positive but small and not statistically significant, which suggested the two variables do not have a strong dual change relationship. Raw correlations showed that higher allostatic load was associated with more depressive symptoms being reported around three years later in two out of three cases, replicating previous associations. Higher allostatic load was also associated with more depressive symptoms being reported around six years later. Effect sizes for these correlations were consistently small. Altogether, our study confirms some relationships between allostatic load and depressive symptoms, but suggests the relationship is less strong than some smaller studies have previously reported.

Chapter 7

Does APOE e4 moderate the relationships of cortisol levels, depressive symptoms and neuroticism with cognitive ability? A replication study in the 36-Day Sample

Formatting note

This chapter was initially prepared as a manuscript for submission to *BMC Psychology*. It is presented here with an added prelude and then only minimal changes made to match the formatting and requirements of this thesis.

7.1 Prelude

The modest effect sizes in the previous two studies suggest that allostatic load summary scores offer limited clinical utility for cognitive ageing and depressive symptoms. However, as we have noted, the allostatic load model could still prove fruitful for clinicians assessing risk for these conditions. It may be that markers of HPA axis activity and function will be the most important in leveraging the allostatic load model. However, the previous studies were limited in that no HPA axis markers were available to be used in the calculation of allostatic load scores. In this chapter, the final study of this thesis will test older age associations between the levels and dynamics of an important HPA axis marker - cortisol - and cognitive ability.

Furthermore, the clinical utility of the allostatic load model for cognitive ageing will be enhanced if moderating variables can be found that help to identify mechanisms and improve prediction. Previously, we used data from the LBC1936 and found that *APOE* e4 allele possession did not deleteriously moderate the relationships of allostatic load, depressive symptoms, and neuroticism with cognitive ability and subsequent cognitive change (Crook et al., 2018). However, while previous research had reported *APOE* e4-by-cortisol interactions, our study used an allostatic load measure that did not incorporate any measures of cortisol (Crook et al., 2018). The 36-Day Sample, a different follow-up study of the Scottish Mental Survey 1947, collected salivary cortisol data in older age, which enables us to attempt replication of previous *APOE* e4-by-cortisol results. Hence, the present study will test whether *APOE* e4 allele possession moderates the relationships of cortisol, depressive symptoms, and neuroticism with cognitive ability.

7.2 Introduction

Cognitive ageing is an increasingly important public health issue (Deary et al., 2009). It is therefore crucial that researchers attempt to better understand risk factors for adverse cognitive outcomes in older age, including how they are modified.

7.2.1 *APOE* e4, neuroticism, depressive symptoms and cortisol levels: How they relate to each other and to cognitive ability

Four such risk factors are the e4 allele of the apolipoprotein E (*APOE*) gene (Davies et al., 2015, 2014), depressive symptomatology (Chodosh, Miller-Martinez, Aneshensel, Wight, & Karlamangla, 2010; Dotson, Resnick, & Zonderman, 2008; Ganguli, Du, Dodge, Ratcliff, & Chang, 2006; Salthouse, 2014), neuroticism (Luchetti, Terracciano, Stephan, & Sutin, 2016) and various measures of cortisol levels (Beluche et al., 2010; Franz et al., 2011; Lee et al., 2007), each of which has been linked to worse cognitive ability in cross-sectional research using older age samples. In addition, longitudinal studies have related cognitive decline to *APOE* e4 (Davies et al., 2014), neuroticism (Luchetti et al., 2016), cortisol levels (Franz et al., 2011; Li et al., 2006), and, in some (Chodosh, Kado, Seeman, & Karlamangla, 2007; Dotson et al., 2008; Paterniti, Verdier-Taillefer, Dufouil, & Alpérovitch, 2002; Sachs-Ericsson, Joiner, Plant, & Blazer, 2005; Wilson et al., 2002; Wilson, Leon, Bennett, Bienias, & Evans, 2004) but not all studies (Bunce, Batterham, Mackinnon, & Christensen, 2012; Chodosh et al., 2010; Ganguli et al., 2006), depressive symptomatology.

As well as being related to cognitive ability, neuroticism, depressive symptoms, and cortisol levels have been linked to each other.

Neuroticism is a normal-range personality trait which encompasses various closely related facets, including the tendency to depression (Costa Jr. & McCrae, 1995). Higher neuroticism is a risk factor for depressive disorders (Kotov, Gamez, Schmidt, & Watson, 2010). Some studies have found high neuroticism to be related to higher levels of cortisol secretion (Nater, Hoppmann, & Klumb, 2010; Portella, Harmer, Flint, Cowen, & Goodwin, 2005), although others have reported no relationship (Riese, Rijsdijk, Rosmalen, Snieder, & Ormel, 2009; see the introduction of Garcia-Banda et al., 2014 for a brief review). Depression has been related to muted as well as greater cortisol awakening responses (Chida & Steptoe, 2009), a flatter diurnal cortisol slope (Adam et al., 2017), to increased cortisol in response to psychological stress (Burke, Fernald, Gertler, & Adler, 2005), and to high cortisol levels in older individuals (Murri et al., 2014) but not in a set of mixed-age samples (Knorr, Vinberg, Kessing, & Wetterslev, 2010).

7.2.2 Cortisol: Differentiating measures

The aforementioned studies relating cortisol to other variables used different cortisol measures. Some measures assess overall cortisol exposure over a sustained period (e.g., urinary cortisol, half a day or a day; hair cortisol, periods of months), some test cortisol levels at a particular time of day or during a particular event (e.g., waking cortisol, evening cortisol, during a lab visit), and some measures aim to quantify particular dynamics of cortisol secretion (e.g., the cortisol awakening response, the diurnal cortisol slope, responses to stress; Nicolson, 2008; Russell, Koren, Rieder, & van Uum, 2012). Here, we will consider what cortisol is, what cortisol measures are used and how these measures, which index various aspects of cortisol secretion, have different relationships with cognitive ability.

Cortisol is the primary glucocorticoid hormone in humans (Kirschbaum & Hellhammer, 1989). Cortisol secretion is heightened in response to stress or challenge and varies according to the time of day (Kirschbaum & Hellhammer, 1989). Chronic under- or over-secretion of cortisol can be harmful to health (McEwen, 1998).

Generally, higher cortisol production has been associated with lower cognitive ability (Franz et al., 2011; Lee et al., 2007; Li et al., 2006; but see Fonda, Bertrand, O'Donnell, Longcope, & McKinlay, 2005). The relationship between cortisol and cognition may be partially mediated by the effects cortisol has on the brain. One specific example is that the relationship between cortisol and declarative memory may be mediated through effects on the hippocampus. Experiments have found that increasing cortisol can negatively affect declarative memory, which the hippocampus is crucial for (Quervain, Roozendaal, Nitsch, McGaugh, & Hock, 2000; Squire, 1992). Plus, higher cortisol was related, although weakly, to lower hippocampal volume in a recent meta-analysis of late-life depression patients and controls (Geerlings & Gerritsen, 2017). This effect has also been observed within individuals. A small longitudinal study into the treatment of Cushing's syndrome, which is caused by prolonged and excessive glucocorticoid exposure (Newell-Price, Bertagna, Grossman, & Nieman, 2006), found that lowering patients' cortisol levels can lead to their hippocampal volume increasing (Starkman et al., 1999 *n*s = 18 for plasma cortisol, 17 for urinary cortisol). Other effects on the brain may also help to explain relationships with cognitive ability. Effects of cortisol on the prefrontal cortex may partially mediate the relationships cortisol has with working memory and executive function (Lupien, Gillin, & Hauger, 1999; McCormick, Lewis, Somley, & Kahan, 2007; Qin, Hermans, van Marle, Luo, & Fernández, 2009). Additionally, higher cortisol in response to a stressor has been related to reduced brain white matter integrity (Cox, Bastin, et al., 2015).

Although higher cortisol levels are typically linked to lower cognitive ability, studies relating morning cortisol to cognitive ability have reported both positive and negative relationships (Beluche et al., 2010; Comijs et al., 2010; Evans et al., 2011; Geoffroy, Hertzman, Li, & Power, 2012; Kuningas et al., 2007; MacLullich et al., 2005; Stawski et al., 2011), as well as null results. Methodological differences may partially explain the variety of results from studies relating morning cortisol and cognitive measures, including different cognitive domains being researched and cortisol samples being collected at different points during the morning (Beluche et al., 2010; Comijs et al., 2010; Evans et al., 2011; Geoffroy et al., 2012; Kuningas et al., 2007; MacLullich et al., 2005; Stawski et al., 2011). On multiple occasions, higher morning cortisol has been found to be related to better executive functioning (Evans et al., 2011; Stawski et al., 2011), but worse general cognitive functioning, slower processing speed, (Kuningas et al., 2007; MacLullich et al., 2005) and lower verbal fluency (Beluche et al., 2010; Geoffroy et al., 2012). Relationships with memory in both directions have been reported (Evans et al., 2011; Geoffroy et al., 2012).

Aspects of the dynamics of cortisol secretion, such as the cortisol awakening response and the diurnal cortisol slope, have also been studied in relation to cognitive ability. In a person with a healthy salivary cortisol diurnal pattern, salivary cortisol levels are high immediately after waking, then rise higher to a peak reached after around 30 to 45 minutes, before declining gradually to much lower levels by the end of the day (Adam & Kumari, 2009). A greater cortisol awakening response, that is, a larger increase in cortisol levels in the 30-45 minutes after waking, has been linked to higher cognitive ability (e.g., Stawski et al., 2011; Ennis, Moffat, & Hertzog, 2016; Evans et al., 2011; Evans, Hucklebridge, Loveday, & Clow, 2012; Law, Evans, Thorn, Hucklebridge, & Clow, 2015; but see Almela, van der Meij, Hidalgo, Villada, & Salvador, 2012). In addition, a flatter diurnal cortisol slope, that is, a smaller decline from morning to evening cortisol levels, has been related to worse performance on some cognitive measures (e.g., Stawski et al., 2011; Evans et al., 2011; but see the null result in Hidalgo, Almela, Pulopulos, & Salvador, 2016) and greater cognitive decline on some measures (e.g., Beluche et al., 2010; but see the null result using serum cortisol in Comijs et al., 2010), although these studies included null results for relationships with other cognitive domains. In a previous study of cortisol measures and cognitive ability in the 36-Day Sample, most of the associations that were statistically significant suggested higher cortisol levels were related to higher cognitive ability, although the associations were no longer significant after correction for multiple testing (Harris, Cox, Brett, Deary, & MacLullich, 2017). In a male subsample of the Lothian Birth Cohort 1936, which like the 36-Day-Sample has followed up Scottish Mental Survey 1947 participants in older age, cognitive change since age 11 was related to reactive cortisol after a cognitive test but unrelated to the diurnal cortisol slope (Cox, MacPherson, et al., 2015).

7.2.3 Does *APOE* e4 moderate the relationships of neuroticism, depressive symptoms and cortisol levels with cognitive ability and decline?

The aforementioned relationships of neuroticism, depressive symptoms and cortisol levels with cognitive outcomes may be moderated by possession of the *APOE* e4 allele.

Studies using the Ginkgo Evaluation of Memory study sample have suggested that *APOE* e4 moderates the deleterious effect of neuroticism on cognitive ability and decline in older age, such that the effect is greater in e4 allele carriers (Dar-Nimrod, Chapman, Franks, et al., 2012; Dar-Nimrod, Chapman, Robbins, et al., 2012). In addition, some studies in other samples have suggested that in e4 carriers, depressive symptomatology is more strongly related to subsequent mild cognitive impairment risk (Geda et al., 2006), dementia risk (Irie et al., 2008; Karlsson et al., 2015; Kim et al., 2011) and non-pathological cognitive decline (Corsentino, Sawyer, Sachs-Ericsson, & Blazer, 2009; Niti, Yap, Kua, & Ng, 2009; Rajan, Wilson, Skarupski, Leon, & Evans, 2014). However, our recent study found no e4-by-neuroticism or e4-by-depressive symptoms moderation when testing associations with contemporaneously measured cognitive ability and subsequent cognitive decline (Crook et al., 2018). Also, a 2010 study reported no e4-by-depressive symptoms moderation in models predicting cognitive decline (Köhler et al., 2010), and a 2015 study used a large sample ($n > 18{,}000$) and found no e4-by-depression or e4-by-psychological distress symptoms moderations in models predicting various cognitive test scores (Luciano et al., 2015).

Studies have also suggested that the effect of cortisol on cognitive ability may be greater in e4 carriers (Gerritsen et al., 2011; Lee et al., 2008; Singh-Manoux et al., 2014). In models predicting various cognitive test scores, interactions of e4 status with the following cortisol measures have been observed: pretest cortisol, AUC cortisol during study visit, morning cortisol, diurnal cortisol slope and evening cortisol (Gerritsen et al., 2011; Lee et al., 2008). Also, one study found that flatness of diurnal cortisol slope and high bedtime cortisol were related to verbal fluency decline in *APOE* e4 carriers (Singh-Manoux et al., 2014). However, this study conducted analyses stratified by genotype and the *APOE*-by-cortisol interaction was not formally tested (Singh-Manoux et al., 2014).

7.2.4 The present study

In the present study, we investigate the main effects of neuroticism, depressive symptoms and cortisol levels on cognitive ability and cognitive change, as well as the moderation of these effects by *APOE* e4 allele possession. Data from the 36-Day Sample enabled us to test the main effects on older age cognition of depressive symptoms, a latent neuroticism factor, and various cortisol measures from three salivary samples across the course of a day. Participants were genotyped for *APOE*, which enabled us to test the moderation of those effects by *APOE* e4 status. Several cognitive tests were administered in older age, which allowed the estimation of associations with both a latent general cognitive ability factor and with aspects of cognitive ability unique to individual tests. Additionally, we were able to fit our models controlling for age 11 IQ to represent cognitive change over an average of 67 years.

Some previous studies have reported non-significant effects. However, it was not clear whether these non-significant results provided evidence for the lack of an effect or were inconclusive as to the presence of an effect. To overcome this limitation, we used a Bayes factor approach. For each effect of interest, we calculated Bayes factors to quantify the relative strength of evidence for the null hypothesis, which assumed no relationship, and the alternative hypothesis, which was based on effect estimates from previous research (see Section 7.3.4).

We hypothesised that higher cognitive ability would be related to greater morning cortisol production and change (waking cortisol, cortisol awakening response, total cortisol production during the waking period), while lower cognitive ability would be related to depressive symptoms, neuroticism, and greater cortisol production and change over the entire waking day (total cortisol production during the waking day, diurnal cortisol slope). Regarding the gene-by-environment moderation of these relationships, we hypothesised that all relationships of cognitive ability measures with cortisol measures, depressive symptoms, and neuroticism would be more negative in *APOE* e4 allele carriers.

7.3 Methods

7.3.1 Participants

The study used the 36-Day Sample, a cohort who participated in the Scottish Mental Survey 1947 (SMS1947) as well as related sociological data collection. The sample has been followed up in older age to study healthy ageing (*n* range for our models = 280-332; *M* [*SD*] age in years at time of cognitive testing = 78.2 [0.9]).

The SMS1947 tested the intelligence of almost all children born in 1936 who were attending school in Scotland (Scottish Council for Research in Education, 1949). In all, 70,805 children took the Moray House Test No. 12, a validated intelligence test (Scottish Council for Research in Education, 1949). Further sociological data were collected from the 36-Day Sample (*n* = 7,380), children born on the first three days of each month. Further data still was collected from a subset of the 36-Day Sample known as the 6-Day-Sample (*n* = 1,208), children born on the first day of all evenly numbered month, who were followed up almost annually until 1963 (Scottish Council for Research in Education, 1949). These samples have been followed up in older age to investigate influences on old-age cognitive and physical health. The 6-Day-Sample was traced in 2012, where possible, using the United Kingdom National Health Service Central Register (NHSCR). Those still living in Great Britain (*n* = 634), along with one emigrant, were recruited and sent a home questionnaire. Later in 2012 and in 2013, those who chose to participate further were sent a physical testing kit with associated instructions (*n* = 174; 92 females). Some participants also went on to take a phone interview (*n* = 131; 72 females). The members of the 36-Day Sample who were not part of the 6-Day-Sample, also known as the 30-Day-Sample, were traced using the NHSCR in 2013 and those still living in Great Britain were invited to participate in the study. Further participants from the 30-Day-Sample then answered the home questionnaire (*n* = 548; 253 females), underwent physical testing in 2014 (*n* = 249; 99 females) and took a phone interview (*n* = 234; 96 females). Further details about the tracing and recruitment stages for the follow-up of the 36-Day Sample has been detailed elsewhere (Brett & Deary, 2014; Deary & Brett, 2015; Harris, Brett, Deary, & Starr, 2016).

7.3.2 Measures

7.3.2.1 Demographic information

Sex and date of birth were recorded at the time of the SMS1947. Dates of subsequent data collection were noted so that participants' ages could be computed.

7.3.2.2 Genotyping

Participants collected their own saliva sample using Oragene OG-500 kits (DNA Genotek, Ottawa, ON, Canada), from which DNA was extracted. The single nucleotide polymorphisms rs7412 and rs429358 were genotyped to determine *APOE* genotype. This was done using TaqMan technology by the Wellcome Trust Clinical Research Facility Genetics Core at the Western General Hospital in Edinburgh, UK.

7.3.2.3 Cortisol

Each participant self-collected three saliva samples over the course of a single day using a Salivette kit (Sarstedt, Rommelsdorf, Germany). They were instructed to provide samples upon awakening, 45 minutes later and at 10:00pm that day, and to record the times that each sample was taken on a questionnaire. Cortisol samples were sent to the University of Dresden for assaying. They were initially placed in a centrifuge and revolved at 3000 revolutions per minute for 5 minutes at 4°C, before they were frozen at -80°C. Cortisol levels were measured at the University of Dresden with an IBL-Hamburg (Hamburg, Germany) immunoassay kit that uses chemiluminescence detection. The practical lower sensitivity limit was 0.3 nmol/l and values lower than this were assigned based on interpolation of the standard curve.

7.3.2.4 Neuroticism

The 6-Day Sample completed a 50-item International Personality Item Pool questionnaire ("Administering IPIP measures, with a 50-item sample questionnaire," n.d.) based on Goldberg's Big Five factor markers (Goldberg, 1992). The 30-Day Sample completed the 20-item Mini-International Personality Item Pool questionnaire (Donnellan, Oswald, Baird, & Lucas, 2006). The four emotional stability/neuroticism items contained in both scales were used to measure neuroticism. Participants responded on a Likert scale from "Very Inaccurate" to "Accurate" how well the statements in the items described them, with responses scored from 1 to 5.

7.3.2.5 Disease and medication information

The home questionnaire asked participants whether they had various medical conditions and asked them to indicate any unlisted health problems or illnesses they had. Participants were also asked to list all medications they were currently taking and the associated doses.

7.3.2.6 Depressive symptoms

The Hospital Anxiety and Depression Scale-Depression subscale (HADS-D; Zigmond & Snaith, 1983) was administered during the phone interview to assess participants' depressive symptoms during the previous week. This subscale has seven items, each with four ordinal response options. Each item is scored from 0 to 3, so the possible range of HADS-D scores is 0 to 21.

7.3.2.7 Cognitive ability

All participants sat the Moray House Test No. 12 (MHT) at age 11 on June 4, 1947 as part of the SMS1947. The MHT is a validated intelligence test that contains various types of items (Scottish Council for Research in Education, 1933). To remove the effect of age at the time of the SMS1947 on MHT scores, we fitted a linear regression of MHT scores on age at the time of the SMS1947. The residuals from this regression were used as the age 11 IQ variable in hypothesis-testing models.

We used scores from the following cognitive tests that were administered during the phone interview: the National Adult Reading Test (NART; Nelson & Willison, 1991), which tests participants' pronunciations of irregularly pronounced words; a semantic fluency test, which required naming as many animals as possible in one minute; the total number of words recalled on five immediate recall trials on the Rey Auditory Verbal Learning Test (RAVLT; Schmidt, 1996), a memory test that requires recall of sets of unrelated words (15 words per trial, maximum possible score of 75); the total number of words recalled on the RAVLT delayed recall trial (maximum possible score of 15); the

Symbol Digit Modalities Test (SDMT; Smith, 1982), a time-limited test of processing speed that requires matching numbers and abstract symbols based on a provided key (120 symbol-digit matches to make, time limit of 90 seconds); and Raven's Standard Progressive Matrices (Raven, Raven, & Court, 2000), a nonverbal reasoning test that requires participants to complete increasingly complex matrices by identifying patterns (60 items, time limit of 20 minutes).

7.3.3 Statistical analysis

As our interest was in non-pathological cognitive ageing, the 4 genotyped participants who indicated they had Parkinson's disease or dementia and provided cognitive test data were excluded from analyses. Further, the 13 genotyped participants who provided cognitive test data while using glucocorticoid medication were excluded from analyses involving cortisol.

We used R Version 3.3.0 (R Core Team, 2016) for data cleaning, visualisation and statistical analysis. This was facilitated by the R packages memisc (Elff, 2016), psych (Revelle, 2016), plyr (Wickham, 2011), likert (Bryer & Speerschneider, 2015), gvlma (Pena & Slate, 2014), MplusAutomation (Hallquist & Wiley, 2014) and Hmisc (Harrell Jr, Charles Dupont, & others., 2016). Latent variable measurement models and hypothesis testing models were fitted with Mplus Version 7.4 (Muthén & Muthén, n.d.). We assumed missing data were missing at random. Hence, we used full information maximum likelihood estimation, as it has been found to perform well when data are missing at random (Enders, 2010). Mardia's tests indicated there was some multivariate skewness in the general cognitive ability tests, and some multivariate skewness and kurtosis in the neuroticism items, so robust standard errors were computed in all models using a sandwich estimator.

Bayes factors were calculated using an R script (Christie, 2011). This script is a modified version of a previous R script (Baguley & Kaye, 2010) that emulates an online Bayes factor calculator (Dienes, 2008).

7.3.3.1 Cortisol measures

Following previous research, we made adjustments to cortisol variable values so that the variable distributions were approximately normal. Raw cortisol values more than three standard deviations from the mean when values were log-transformed were changed to values that were three standard deviations from the mean on the log-transformed scale (Stalder et al., 2016). We then attempted to normalise the distributions of the cortisol variables using square-root, log and various power-based transformations. The most effective at producing normally distributed cortisol variables was a transformation previously recommended for cortisol time series values (Miller & Plessow, 2013). The formula used was the following, where c’ is the transformed cortisol variable and c is the raw cortisol variable: $c' = (c^{0.26} - 1)/0.26$

In our hypothesis testing models, we used four different area under the curve (AUC) cortisol measures calculated with established formulae (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). Measures calculated with respect to the ground are affected by the difference between samples as well as how far the samples are from zero, while measures calculated with respect to the increase are only affected by the difference between samples (Pruessner et al., 2003). Measures with respect to the increase index the dynamics of cortisol secretion, while measures with respect to the ground index both the dynamics and the amount of cortisol secretion. By studying all four of these measures, we can tease apart relationships involving the dynamics and amount of cortisol secretion, in the morning as well as across the entire waking day.” We also used waking cortisol in our hypothesis testing models, because other key studies investigating cortisol and e4-by-cortisol effects investigated waking cortisol, while they had different timings than our study for other cortisol samples (Gerritsen et al., 2011; Singh-Manoux et al., 2014; Stawski et al., 2011). Two of the AUC measures used the first two samples, and two used the first and third samples. The AUC measures were:

- **cortisol awakening response (CAR):** AUC with respect to the increase between the first two samples
- **total cortisol production during the waking period (TCPWP):** AUC with respect to the ground (zero) for the first two samples
- **diurnal cortisol slope (DCS):** AUC with respect to the increase between the first and the third sample
- **total cortisol production during the waking day (TCPWD):** AUC with respect to the ground (zero) for the first and third samples

The varying time differences between samples were adjusted for by multiplying each individual's AUC value by the minimum time difference in the sample divided by their individual time difference. This adjustment recentres the AUC values by changing them to the value they would be if each participant had the minimum time difference in the sample. Hence, the adjustment statistically removes the effects of differences in sample timing on AUC values.

The distributions of the raw TCPWP and TCPWD variables were non-normal and so the same transformation used for the raw cortisol values was applied to the raw TCPWP and TCPWD values.

Linear regression models were used to residualise cortisol measures for the following covariates because they have been found to be related to cortisol production (information was taken from the home questionnaire): whether participants had diabetes (Hackett, Steptoe, & Kumari, 2014); time of the first cortisol sample; how often participants drank alcohol in the previous 12 months; how often participants had four or more alcoholic drinks in the last year; whether participants had an alcoholic drink in the last week (Badrick et al., 2008); whether participants currently smoked; and how many cigarettes per day participants currently smoked (Badrick, Kirschbaum, & Kumari, 2007). The residuals from these models were transformed to have a mean of 0 and a standard deviation of 1, to avoid any model fitting problems due to the sizes of variances. The resulting variables were used in the hypothesis testing models.

7.3.3.2 General cognitive ability measurement models

We modelled general cognitive ability and neuroticism as latent variables, estimated from the covariances between test scores and item responses respectively, using structural equation modelling (SEM). This was done in order to evaluate how well the latent models fit our data, and also to assess the measurement invariance of the models across the two genotypic groups. Measurement invariance is an important analytic step as it ensures that the same construct is being measured in both groups. Measurement models were fitted using only data from participants who had been genotyped and who had reported data on the identified covariates (age at phone interview, sex, stroke).

A series of models were estimated to establish the plausibility of the cognitive ability model. Each model contained a single latent factor and the six cognitive test scores. The latent factor was identified by constraining the loading of NART on the latent factor to 1.0. First, the single factor model was fitted in the e4 allele carriers and non-carriers separately. Then, a multiple group model was estimated with all parameters free to vary across groups. This provided the baseline model fit for subsequent tests of measurement invariance. The groups for this model were defined by genotype status (e4 carriers and non-carriers). Models were evaluated based on three model fit criteria. We aimed for the following cutoffs defined as indicating acceptable fit (Schermelleh-Engel et al., 2003): $CFI \geq .95$; $RMSEA \leq .08$; $SRMR \leq .10$. If there was model misfit, modification indices were consulted and changes made as necessary until a satisfactory fit was obtained.

Measurement invariance was tested to check that our analyses would be appropriate and that all participants could reasonably be included in one-group models. We assessed measurement invariance by sequentially including sets of equality constraints on the model parameters, and testing for the difference in fit between these models. Following standard procedures, we tested for measurement invariance placing constraints on: (1) loadings; (2) intercepts; (3) residuals: and (4) the correlated residual. To ascertain whether measurement invariance assumptions held across groups, the sequentially constrained models were compared based on a chi-square difference test with degrees of freedom equal to the number of constrained parameters. Note that the group sizes meant that our measurement invariance tests could detect major deviations from measurement invariance, but they had limited power to detect smaller deviations from measurement invariance.

The residual variances for each of the six cognitive tests in the single-group measurement model of *g* in genotyped participants were extracted for use in subsequent models.

7.3.3.3 Neuroticism measurement models

The procedure followed for the neuroticism measurement models was the same as that for the *g* measurement models, except no residual variances were extracted. The factor loading of the first item was fixed to 1.0 for identification purposes.

7.3.3.4 Hypothesis testing models

To statistically control for their effects, all models were fitted with the following variables as covariates of the cognitive outcome: age; sex (effects coded as -0.5 for females, 0.5 for males); and a dichotomous indicator of having reported having had a stroke or a mini-stroke coded as 0 for no and 1 for yes (Kase et al., 1998; Pendlebury & Rothwell, 2009). Following recent recommendations (Stalder et al., 2016), for models with the cortisol awakening response as the predictor of interest, we also fitted models with waking cortisol and the e4-by-waking cortisol interaction added as predictors.

Along with our main models, we fitted an additional set of childhood intelligence controlled (CIC) models in which age 11 IQ was added as a predictor. In the CIC models, the effect of a predictor on older age cognitive ability can be interpreted as the best available approximation of the effect of that predictor on cognitive change since age 11.

To properly statistically control for all of the covariates mentioned here, including age 11 IQ in the CIC models, each model also included all gene-by-covariate and environment-by-covariate interactions as predictors of the cognitive outcome (Keller, 2014). So, for example, in the main model for neuroticism, the following covariate interactions were included: e4-by-age; e4-by-sex; e4-by-stroke; neuroticism-by-age; neuroticism-by-sex; and neuroticism-by-stroke.

7.3.4 Bayes factor analyses

In the current analysis, we formulated specific hypotheses about the direction and magnitude for each of the focal effects of interest in the study. To ascertain how much evidence the data provided for each of these hypotheses (alternative hypotheses) compared to the standard null hypothesis of no effect, we calculated Bayes factors.

Bayes factors compare two theories as explanations of the data and indicate how strongly each theory is supported by the data compared to the other. Bayes factors are a continuous measure of support for one hypothesis relative to another hypothesis. They indicate how one's prior odds for the probabilities of the theories underpinning each of the hypotheses should be adjusted in light of the data. A Bayes factor of 1 indicates the data do not provide evidence for one theory versus the other. When comparing a null hypothesis and an alternative hypothesis, as we do here, a Bayes factor below 1/3 indicates substantial relative evidence for the null hypothesis, a Bayes factor above 3 indicates substantial relative evidence for the alternative hypothesis, and a Bayes factor between 1/3 and 3 indicates no substantial evidence for either hypothesis (Jeffreys et al., 1961, cited in Dienes, 2014).

The theories compared by Bayes factors are represented by probability distributions. Note that these distributions represent probability densities for the effect size of interest and are not the expected distributions of data under each hypothesis. Here, the null hypothesis was represented by a distribution with all of the probability density at an effect size of zero, i.e., a nil null. Each alternative hypothesis was represented by a half-normal distribution where the standard deviation was the expected effect size. The expected effect sizes were based on the previous research cited later in this section. As the alternative hypotheses were represented by half-normal distributions, all the probability density in these representations was on the hypothesised side of zero (i.e., all on the positive or negative side depending on the hypothesised direction of the effect) and most of the density was below the expected effect size, which made this a conservative choice of distribution (Dienes, 2014). Thus, the chosen distribution for each alternative hypothesis specified both the hypothesised direction and the hypothesised magnitude of the effect.

The primary focus of the current study is evaluating evidence for the interaction with *APOE* e4 status. Previous research reporting some of these interactions has found them to be larger in size than the associated main effects (e.g., Corsentino et al., 2009; Dar-Nimrod, Chapman, Franks, et al., 2012; Rajan et al., 2014). We took a conservative approach and considered that this may be an overestimate of the size of the effect (Ioannidis, 2008), so we limited the expected interaction effect sizes at the associated expected main effect sizes.

For the AUC cortisol measures, our alternative hypotheses had an expected main effect size of 0.12. This is the mean effect size of diurnal variability on four cognitive measures for which coefficients could be standardised in Gerritsen et al. (2011). The expected effect size for the AUC cortisol measures interaction, -0.09, was the mean standardised difference between e4 and non-e4 groups for models predicting two cognitive measures with diurnal variability that were reported in full by group in Gerritsen et al. (2011). For waking cortisol, we limited the expected effect size to 0.08, the size of the association between morning cortisol and reasoning in Singh-Manoux et al. (2014).

The main effect sizes for depressive symptoms and neuroticism were taken from Crook et al. (2018), which used a sample closely related to this study's sample.

In summary, our expected effect sizes (as standardised betas) were as follows:

- 0.08: waking cortisol
- -0.08: *APOE* e4-by-waking cortisol interaction
- 0.12: total cortisol production during the waking period, cortisol awakening response, cortisol awakening response (waking cortisol controlled)
- -0.12: total cortisol production during the waking day, diurnal cortisol slope

- -0.09: all *APOE* e4-by-cortisol interactions except for e4-by-waking cortisol
- -0.18: depressive symptoms
- -0.18: *APOE* e4-by-depressive symptoms interaction
- -0.25: neuroticism
- -0.25: *APOE* e4-by-neuroticism interaction

7.4 Results

Table 7.1: Descriptive statistics for cortisol, depressive symptoms, neuroticism and cognitive tests

	n	M	SD	Skew	Kurtosis
age at phone interview (years)	330	78.19	0.89	-0.55	-1.26
raw cortisol sample 1 (ug/ml)	307	32.20	27.35	4.32	25.36
raw cortisol sample 2 (ug/ml)	306	32.41	20.58	3.10	18.10
raw cortisol sample 3 (ug/ml)	304	6.73	8.98	3.87	15.83
CAR	297	2.80	26.65	0.25	2.16
TCPWP	297	11.04	0.94	-0.52	2.79
DCS	293	-2.28	1.21	0.52	2.76
TCPWD	293	1.97	0.50	-0.24	4.04
HADS-D	329	2.36	1.85	1.27	1.90
Mini-IPIP item 4	329	2.05	1.12	0.69	-0.59
Mini-IPIP item 9	329	2.65	1.16	0.62	-0.60
Mini-IPIP item 14	328	3.15	1.33	-0.16	-1.20
Mini-IPIP item 19	329	2.93	1.24	0.16	-1.09
National Adult Reading Test	327	34.23	8.32	-0.73	0.18
Semantic fluency	329	18.38	5.25	0.46	0.08
RAVLT immediate recall	326	46.23	11.33	-0.07	-0.43
Symbol Digit Modalities Test	322	42.30	8.92	-0.28	-0.03
Raven's SPM	326	32.82	7.83	-0.50	-0.13
RAVLT delayed recall	324	9.28	3.42	0.30	-0.41
Moray House Test	316	48.74	10.81	-0.84	0.94

Note.

CAR = cortisol awakening response. TCPWP = total cortisol production during the waking period. DCS = diurnal cortisol slope. TCPWD = total cortisol production during the waking day. HADS-D = Hospital Anxiety and Depression Scale-Depression. IPIP = International Personality Item Pool. RAVLT = Rey Auditory Verbal Learning Test. Raven's SPM = Raven's standard progressive matrices.

Table 7.1 contains descriptive statistics. The sample was 56% male (144 females, 186 males). There were 250 e4 non-carriers and 80 e4 carriers. For the non-cortisol variables, those who indicated having Parkinson's or dementia or who had missing data for the sex, age at phone interview, *APOE* e4 status or the stroke variable were excluded. Note that for the cortisol variables, those for whom residualised cortisol variables were not created were also excluded, i.e. those who were taking glucocorticoid medication or had missing data on the cortisol covariates. Atypically, most included participants had higher cortisol values upon awakening than they did 45 minutes later. As expected, though, for almost all participants cortisol values were higher in the early samples than in the late sample. The *n* range for the models fitted was 280-332, which meant that our measurement invariance tests were likely powered to detect major but not minor deviations from measurement invariance.

7.4.1 Measurement models

7.4.1.1 General cognitive ability measurement models

One correlated residual, between RAVLT immediate recall and RAVLT delayed recall, was added to the general cognitive ability measurement model to improve fit based on its modification index. This correlated residual was retained in subsequent models. Chi-square tests indicated that measurement invariance of test loadings, intercepts and residuals held across e4 carriers and non-carriers (all *p*s $\geq$ .05). The results of these model indicated that there were no statistically identifiable differences in the measurement of general cognitive ability across the e4 status groups. As a result, it was reasonable to conduct subsequent analyses on a single combined sample. The final measurement model for general cognitive ability, fitted in a single group model, fit well: *CFI* = .986; *RMSEA* = .056; *SRMR* = .031. The standardised loadings in this model were .54 for NART, .39 for semantic fluency, .44 for immediate recall, .32 for the Symbol Digit Modalities Test, .74 for Raven's matrices and .67 for delayed recall. The correlation between the recall residuals was .75.

7.4.1.2 Neuroticism measurement models

One correlated residual, between Mini-IPIP items 4 and 14, was added to the model. It made the best improvement to fit of two correlated residuals we considered, each of which had near-identical modification indices. Chi-square tests indicated that measurement invariance of test loadings, intercepts and residuals held across e4 carriers and non-carriers (all *p*s $\geq$ .05). The final neuroticism measurement model, fitted in a single group, fit well: *CFI* = 0.999; *RMSEA* = .015; *SRMR* = .014. The standardised item loadings were -.28, .52, .58 and .43 for Mini-IPIP items 4, 9,14 and 19 respectively, and the correlation between the residuals of items 4 and 4 was .18. Again, as with the general cognitive ability models, the results indicated that there were no statistically identifiable differences in the measurement of neuroticism across groups, and support the use of a combined single sample for subsequent analyses.

setcounterpage189

7.4.2 Hypothesis testing via Bayes factors

Tables 2-7 contain the results from models predicting general cognitive ability, and the predicting of specific cognitive abilities over and above *g*. Each table presents standardised betas, standard errors and Bayes factors from 16 models, as two model types were fitted for each of the eight predictors. For each predictor, a main model was fitted. Each main model included the main effect of the predictor, the interaction of the predictor with *APOE* e4 status, and all health and other covariates. For each predictor, we also fitted a model which added statistical control for age 11 IQ on the older age cognitive measures. For each predictor, comparing the models with and without control for childhood intelligence enabled us to consider whether controlling for prior ability attenuates effects. The range of *n*s for the main models was 294 to 332. The range of *n*s for the models with childhood intelligence controlled was 280 to 317.

In the sections that follow, we first present the results for general cognitive ability, before discussing the models predicting the residuals for each of the cognitive tests. The latter set of models essentially investigated the effects of cortisol, depression and neuroticism on specific cognitive domains after controlling for general ability (for example, is there a specific effect of depressive symptoms on semantic fluency?).

7.4.2.1 General cognitive ability

Table 7.2: General cognitive ability predicted by cortisol levels, depressive symptoms and neuroticism

	MM main beta	MM main SE	MM main BF	MM int. beta	MM int. SE	MM int. BF	CICM main beta	CICM main SE	CICM main BF	CICM int. beta	CICM int. SE	CICM int. BF
WC	0.12	0.07	2.40	-0.08	0.08	1.35	0.06	0.11	1.06	-0.05	0.06	1.17
CAR	0.03	0.07	0.70	0.00	0.08	0.66	0.00	0.12	0.69	0.00	0.05	0.52
CAR (WCC)	0.15	0.09	2.49	-0.10	0.10	1.40	0.04	0.12	0.87	-0.06	0.07	1.15
TCPWP	0.14	0.07	5.23	-0.11	0.09	1.59	-0.01	0.11	0.66	-0.05	0.07	1.02
DCS	-0.02	0.08	0.66	0.02	0.08	0.58	-0.05	0.10	0.91	-0.05	0.05	1.08
TCPWD	0.06	0.07	0.30	-0.01	0.10	0.79	-0.03	0.10	0.81	-0.04	0.06	0.86
DS	-0.15	0.08	3.47	0.08	0.07	0.21	-0.04	0.08	0.61	0.02	0.05	0.21
N	0.11	0.10	0.19	-0.17	0.11	2.07	0.24	0.18	0.30	-0.06	0.10	0.60

Note.

MM = main model. CICM = childhood intelligence controlled model. SE = standard error. BF = Bayes factor. WC = waking cortisol. CAR = cortisol awakening response. WCC = waking cortisol controlled. TCPWP = total cortisol production during the waking period. DCS = diurnal cortisol slope. TCPWD = total cortisol production during the waking day. DS = depressive symptoms. N = neuroticism.

Table 7.2 contains the results from models with the general cognitive ability factor as the outcome. We used Bayes factors to compare the null hypothesis of no effect to pre-specified alternative hypotheses. This means that for each effect tested, we can assess whether there was relative evidence for the null compared to the alternative ($BF \leq 1/3$), relative evidence for the alternative compared to the null ($BF \geq 3$), or no strong relative evidence for either hypothesis ($1/3 < BF < 3$). The table shows that most Bayes factors produced were between 1/3 and 3, which means that for most tests, neither the alternative nor the null hypothesis was strongly supported. Here, we will note the hypothesised effects that were supported ($BF \geq 3$, evidence for the alternative hypothesis) and effects where results suggested no relationship was more likely than the hypothesised relationship ($BF \leq 1/3$, evidence for the null hypothesis).

7.4.2.1.1 Depressive symptoms and general cognitive ability. Depressive symptoms were related to lower contemporaneous general cognitive ability (β [*SE*] = -0.15 [0.08], *BF* = 3.47). However, in the model controlling for childhood intelligence, the results were inconclusive regarding whether depressive symptoms were related to worse general cognitive change since childhood (β [*SE*] = -0.04 [0.08], *BF* = 0.61). Both before and after controlling for childhood intelligence, the effect of depressive symptoms on general cognitive ability did not vary depending on *APOE* e4 status (main model β [*SE*] = 0.08 [0.07], *BF* = 0.21; childhood intelligence controlled model β [*SE*] = 0.02 [0.05], *BF* = 0.21).

7.4.2.1.2 Total cortisol produced during the waking period and general cognitive ability. The total cortisol produced during the waking period was related to higher general cognitive ability (β [*SE*] = 0.14 [0.07], *BF* = 5.23). Total cortisol produced during the waking period was not related to change in general cognitive ability (β [*SE*] = -0.01 [0.11], *BF* = 0.66).

7.4.2.1.3 Total cortisol produced during the waking day and general cognitive ability. The total cortisol produced during the waking day was not related to general cognitive ability (β [*SE*] = 0.06 [0.07], *BF* = 0.30). However, there was no strong evidence regarding this measure's relationship with general cognitive change since childhood (β [*SE*] = -0.03 [0.10], *BF* = 0.81).

7.4.2.1.4 Neuroticism and general cognitive ability. The null hypotheses that neuroticism was not related to general cognitive ability or general cognitive change were supported (main model β [*SE*] = 0.11 [0.10], *BF* = 0.19; childhood intelligence controlled model β [*SE*] = 0.24 [0.18], *BF* = 0.30).

7.4.2.2 Cognitive test residual models

Table 7.3: National Adult Reading Test residuals predicted by cortisol levels, depressive symptoms and neuroticism

	MM main beta	MM main SE	MM main BF	MM int. beta	MM int. SE	MM int. BF	CICM main beta	CICM main SE	CICM main BF	CICM int. beta	CICM int. SE	CICM int. BF
WC	-0.17	0.08	0.47	0.09	0.07	0.36	-0.14	0.11	0.50	0.10	0.06	0.25
CAR	0.11	0.08	2.19	0.00	0.06	0.56	0.10	0.11	1.26	0.00	0.06	0.55
CAR (WCC)	0.11	0.10	1.55	0.09	0.09	0.43	0.04	0.12	0.85	0.13	0.08	0.33
TCPWP	-0.03	0.08	0.44	0.11	0.08	0.33	-0.04	0.12	0.59	0.15	0.07	0.21
DCS	-0.01	0.07	0.54	-0.03	0.07	0.80	-0.14	0.10	1.82	-0.07	0.06	1.44
TCPWD	-0.08	0.08	1.21	0.06	0.08	0.47	-0.23	0.10	5.17	0.05	0.06	0.33
DS	0.05	0.07	0.22	0.04	0.05	0.18	0.04	0.06	0.22	0.00	0.05	0.26
N	-0.17	0.10	2.11	-0.06	0.08	0.57	-0.28	0.14	4.45	-0.01	0.08	0.34

Note.

MM = main model. CICM = childhood intelligence controlled model. SE = standard error. BF = Bayes factor. WC = waking cortisol. CAR = cortisol awakening response. WCC = waking cortisol controlled. TCPWP = total cortisol production during the waking period. DCS = diurnal cortisol slope. TCPWD = total cortisol production during the waking day. DS = depressive symptoms. N = neuroticism.

Table 7.4: Semantic fluency residuals predicted by cortisol levels, depressive symptoms and neuroticism

	MM main beta	MM main SE	MM main BF	MM int. beta	MM int. SE	MM int. BF	CICM main beta	CICM main SE	CICM main BF	CICM int. beta	CICM int. SE	CICM int. BF
WC	-0.13	0.07	0.29	0.09	0.07	0.35	-0.21	0.09	0.29	0.10	0.07	0.34
CAR	0.00	0.07	0.52	-0.03	0.07	0.80	-0.02	0.10	0.59	-0.02	0.07	0.76
CAR (WCC)	-0.14	0.09	0.26	0.09	0.08	0.41	-0.16	0.11	0.34	0.11	0.09	0.38
TCPWP	-0.13	0.07	0.20	0.07	0.08	0.39	-0.21	0.10	0.27	0.09	0.08	0.35
DCS	0.02	0.08	0.46	0.04	0.08	0.52	-0.06	0.11	0.96	0.04	0.08	0.52
TCPWD	-0.13	0.06	4.77	0.15	0.06	0.17	-0.26	0.09	20.32	0.17	0.06	0.17
DS	0.02	0.07	0.27	0.00	0.06	0.34	0.06	0.07	0.21	0.01	0.07	0.30
N	-0.05	0.08	0.55	-0.05	0.08	0.52	0.07	0.10	0.24	-0.07	0.08	0.70

Note.

MM = main model. CICM = childhood intelligence controlled model. SE = standard error. BF = Bayes factor. WC = waking cortisol. CAR = cortisol awakening response. WCC = waking cortisol controlled. TCPWP = total cortisol production during the waking period. DCS = diurnal cortisol slope. TCPWD = total cortisol production during the waking day. DS = depressive symptoms. N = neuroticism.

Table 7.5: RAVLT immediate recall residuals predicted by cortisol levels, depressive symptoms and neuroticism

	MM main beta	MM main SE	MM main BF	MM int. beta	MM int. SE	MM int. BF	CICM main beta	CICM main SE	CICM main BF	CICM int. beta	CICM int. SE	CICM int. BF
WC	-0.04	0.07	0.49	0.15	0.08	0.30	0.11	0.12	1.30	0.15	0.07	0.28
CAR	0.04	0.08	0.75	-0.17	0.06	13.41	0.03	0.15	0.86	-0.17	0.06	14.66
CAR (WCC)	0.11	0.08	0.69	-0.16	0.09	3.04	0.03	0.13	0.84	-0.18	0.09	3.54
TCPWP	0.02	0.06	0.38	0.11	0.08	0.34	0.06	0.12	0.95	0.11	0.08	0.35
DCS	-0.02	0.07	0.59	-0.14	0.07	5.17	0.00	0.10	0.64	-0.14	0.07	3.82
TCPWD	-0.04	0.06	0.80	0.14	0.08	0.28	0.03	0.12	0.61	0.15	0.08	0.28
DS	-0.20	0.06	79.60	0.02	0.06	0.25	-0.18	0.08	6.62	0.01	0.06	0.28
N	-0.02	0.09	0.40	0.16	0.06	0.07	0.07	0.12	0.29	0.15	0.06	0.06

Note.

MM = main model. CICM = childhood intelligence controlled model. SE = standard error. BF = Bayes factor. WC = waking cortisol. CAR = cortisol awakening response. WCC = waking cortisol controlled. TCPWP = total cortisol production during the waking period. DCS = diurnal cortisol slope. TCPWD = total cortisol production during the waking day. DS = depressive symptoms. N = neuroticism.

Table 7.6: Symbol Digit Modalities Test residuals predicted by cortisol levels, depressive symptoms and neuroticism

	MM main beta	MM main SE	MM main BF	MM int. beta	MM int. SE	MM int. BF	CICM main beta	CICM main SE	CICM main BF	CICM int. beta	CICM int. SE	CICM int. BF
WC	0.15	0.08	2.75	-0.13	0.07	3.48	0.18	0.12	1.94	-0.16	0.07	6.92
CAR	-0.11	0.07	0.24	0.11	0.06	0.24	-0.12	0.12	0.44	0.11	0.06	0.22
CAR (WCC)	-0.04	0.10	0.53	0.06	0.09	0.50	-0.05	0.14	0.66	0.04	0.09	0.57
TCPWP	0.12	0.08	1.96	-0.07	0.07	1.32	0.09	0.14	1.12	-0.11	0.08	1.95
DCS	-0.04	0.09	0.81	0.04	0.08	0.49	0.13	0.14	0.51	0.05	0.09	0.54
TCPWD	0.13	0.07	0.22	-0.11	0.06	3.40	0.23	0.11	0.28	-0.15	0.06	7.11
DS	-0.08	0.07	1.18	0.06	0.07	0.20	-0.01	0.07	0.43	0.05	0.06	0.21
N	0.09	0.35	0.73	-0.03	0.36	0.85	0.09	0.12	0.26	-0.01	0.07	0.28

Note.

MM = main model. CICM = childhood intelligence controlled model. SE = standard error. BF = Bayes factor. WC = waking cortisol. CAR = cortisol awakening response. WCC = waking cortisol controlled. TCPWP = total cortisol production during the waking period. DCS = diurnal cortisol slope. TCPWD = total cortisol production during the waking day. DS = depressive symptoms. N = neuroticism.

Table 7.7: Raven's Standard Progressive Matrices residuals predicted by cortisol levels, depressive symptoms and neuroticism

	MM main beta	MM main SE	MM main BF	MM int. beta	MM int. SE	MM int. BF	CICM main beta	CICM main SE	CICM main BF	CICM int. beta	CICM int. SE	CICM int. BF
WC	0.09	0.07	1.72	−0.15	0.07	4.75	0.08	0.12	1.14	−0.16	0.07	5.63
CAR	0.04	0.06	0.72	0.01	0.05	0.47	0.00	0.11	0.69	0.01	0.05	0.43
CAR (WCC)	0.15	0.08	3.25	−0.18	0.08	4.41	0.11	0.11	1.30	−0.16	0.08	3.55
TCPWP	0.10	0.07	2.03	−0.22	0.08	12.14	0.02	0.11	0.74	−0.22	0.07	16.72
DCS	0.03	0.08	0.43	0.07	0.07	0.37	0.00	0.13	0.74	0.05	0.07	0.42
TCPWD	0.07	0.07	0.26	−0.13	0.07	3.05	0.10	0.15	0.57	−0.15	0.07	4.40
DS	0.04	0.06	0.19	−0.05	0.06	0.63	0.02	0.06	0.25	−0.06	0.06	0.77
N	0.15	0.09	0.14	−0.10	0.07	1.25	0.28	0.15	0.20	−0.11	0.09	1.16

Note.

MM = main model. CICM = childhood intelligence controlled model. SE = standard error. BF = Bayes factor. WC = waking cortisol. CAR = cortisol awakening response. WCC = waking cortisol controlled. TCPWP = total cortisol production during the waking period. DCS = diurnal cortisol slope. TCPWD = total cortisol production during the waking day. DS = depressive symptoms. N = neuroticism.

Table 7.8: RAVLT delayed recall residuals predicted by cortisol levels, depressive symptoms and neuroticism

	MM main beta	MM main SE	MM main BF	MM int. beta	MM int. SE	MM int. BF	CICM main beta	CICM main SE	CICM main BF	CICM int. beta	CICM int. SE	CICM int. BF
WC	0.04	0.08	0.99	0.04	0.09	0.59	0.14	0.13	1.39	0.01	0.08	0.69
CAR	0.01	0.07	0.54	-0.08	0.06	1.92	0.00	0.13	0.71	-0.08	0.06	2.02
CAR (WCC)	0.01	0.09	0.66	-0.10	0.11	1.33	0.03	0.13	0.83	-0.14	0.11	1.73
TCPWP	0.06	0.08	0.94	0.04	0.10	0.59	0.05	0.13	0.91	-0.01	0.09	0.76
DCS	-0.01	0.08	0.60	-0.10	0.07	2.04	0.08	0.10	0.44	-0.07	0.07	1.27
TCPWD	0.06	0.07	0.31	0.01	0.09	0.67	0.10	0.12	0.46	-0.01	0.09	0.75
DS	-0.12	0.06	2.90	0.00	0.06	0.33	-0.12	0.09	1.71	-0.02	0.07	0.44
N	-0.05	0.09	0.53	0.17	0.07	0.08	-0.07	0.12	0.72	0.16	0.07	0.07

Note.

MM = main model. CICM = childhood intelligence controlled model. SE = standard error. BF = Bayes factor. WC = waking cortisol. CAR = cortisol awakening response. WCC = waking cortisol controlled. TCPWP = total cortisol production during the waking period. DCS = diurnal cortisol slope. TCPWD = total cortisol production during the waking day. DS = depressive symptoms. N = neuroticism.

Tables 7.3-7.8 contain 192 coefficients from 96 models. For most of these coefficients (130), there was not strong evidence for either the null or the alternative hypothesis. Nonetheless, 24 show evidence in favour of the alternative and 38 show evidence in favour of the null. Here, we will describe some of the most notable results in these tables, including depressive symptoms and neuroticism results, as well as results from tests of e4-by-cortisol interactions.

7.4.2.2.1 Depressive symptoms and cognitive test residuals. Depressive symptoms were related to worse (residualised) immediate recall, both before and after controlling for childhood intelligence (main model β [*SE*] = -0.20 [0.06], *BF* = 79.60; childhood intelligence controlled model β [*SE*] = -0.18 [0.08], *BF* = 6.62). The null hypothesis was supported in models where depressive symptoms predicted National Adult Reading Test, semantic fluency and Raven's matrices residuals. The null hypotheses that the relationships of depressive symptoms with cognitive test residuals did not vary according to *APOE* e4 status was supported in the models with and without childhood intelligence controlled for the National Adult Reading test, immediate recall and Symbol Digit Modalities Test residuals.

7.4.2.2.2 Neuroticism and cognitive test residuals. Of the 24 neuroticism effects tested, 13 provided no strong evidence for either the null or the alternative hypothesis, 10 provided evidence for the null, and only one provided evidence for the alternative hypothesis. Notably, there was strong evidence for the null hypotheses about the e4-by-neuroticism interaction predicting immediate and delayed recall residuals (β range = 0.15-0.17, *BF* range = 0.06-0.08, all *p*-values < .05). These results suggested that after general cognitive ability was taken into account, possessing the e4 allele was related to higher recall scores, the opposite to what was hypothesised.

7.4.2.2.3 *APOE* e4-by-cortisol interactions and immediate recall residuals. The relationship between the cortisol awakening response and immediate recall residuals was more negative in e4 carriers (main model β [*SE*] = -0.17 [0.06], *BF* = 13.41; childhood intelligence controlled model β [*SE*] = -0.17 [0.06], *BF* = 14.66), including after waking cortisol levels were controlled for (main model β [*SE*] = -0.16 [0.09], *BF* = 3.04; childhood intelligence controlled model β [*SE*] = -0.18 [0.09], *BF* = 3.54). The diurnal cortisol slope, the other cortisol dynamics measure studied, was also more strongly negatively related to immediate recall residuals in e4 carriers than E4 non-carriers (main model β [*SE*] = -0.14 [0.07], *BF* = 5.17; childhood intelligence controlled model β [*SE*] = -0.14 [0.07], *BF* = 3.82). On the other hand, the total cortisol produced during the waking day interaction coefficients were positive, rather than negative, as was expected. There was evidence for the null hypothesis about this interaction (main model β [*SE*] = 0.14 [0.08], *BF* = 0.28; childhood intelligence controlled model β [*SE*] = 0.15 [0.08], *BF* = 0.28).

7.4.2.2.4 *APOE* e4-by-cortisol interactions and Raven's matrices residuals. Measures of total cortisol production related more negatively to Raven's matrices residuals in E4 carriers. The evidence for the e4 interaction was stronger for the total cortisol production during the waking period (main model β [*SE*] = -0.22 [0.08], *BF* = 12.14; childhood intelligence controlled model β [*SE*] = -0.22 [0.07], *BF* = 16.72) than the total cortisol production during the waking day (main model β [*SE*] = -0.13 [0.07], *BF* = 3.05; childhood intelligence controlled model β [*SE*] = -0.15 [0.07], *BF* = 4.40).

7.4.2.3 Age 11 IQ effects in cognitive test residual models

Briefly, it is of interest to note the magnitude of the effects of age 11 IQ predicting the residuals for each cognitive test after controlling for *g*. These coefficients give an indication of how strongly prior general ability as measured by the Moray House Test No. 12 relates to specific abilities over and above general ability in later life. The β ranges were: 0.44 to 0.51 for NART residuals; 0.07 to 0.11 for semantic fluency residuals; -0.07 to -0.03 for immediate recall residuals; -0.03 to 0.00 for symbol digit residuals; 0.14 to 0.19 for Raven's residuals; and -0.14 to -0.13 for delayed recall residuals.

7.5 Discussion

In this replication-and-extension study, we tested the older-age associations of depressive symptoms, neuroticism and six cortisol measures with general cognitive ability and the specific variance associated with six cognitive tests. We also tested whether these associations were moderated by possession of the *APOE* e4 allele. Additionally, we repeated each model with age 11 cognitive ability as a covariate of the cognitive outcome to test associations with cognitive change since childhood.

We replicated previous findings that depressive symptoms are related to worse general cognitive ability, with a small-to-moderate effect size. However, unlike in some previous studies, this negative association was not stronger in e4 carriers. We also found that producing more cortisol during the waking period (the first ~45 minutes after waking) was related to higher general cognitive ability, with a small-to-moderate effect size. We failed to replicate an association between neuroticism and worse general cognitive ability. Also against expectations, the total cortisol produced during the waking day was not related to worse general cognitive ability.

We will now discuss the notable results from models relating depressive symptoms, neuroticism and cortisol to the specific variance of cognitive tests, categorised by predictor type.

7.5.1 Depressive symptoms

Depressive symptoms seem to be mostly related to cognitive test scores through a factor of general cognitive ability.

After removing the variance in cognitive tests associated with the general factor, we found depressive symptoms were not related to results on the National Adult Reading Test, a semantic fluency test or Raven's matrices. For those three tests, there was relative support for the null hypothesis (versus the alternative hypothesis that a negative association does exist).

However, we did find that depressive symptoms were related to lower residualised immediate recall test scores. Future research could also investigate the relationship between depressive symptoms and delayed recall residuals (*BF* = 2.90). This result is consistent with previous research that compared depression cases and controls, as a meta-analysis found memory was impaired in depression and there was greater impairment on immediate recall than delayed recall (Burt, Zembar, & Niederehe, 1995).

Overall, results suggested that depressive symptoms were not more strongly related to cognitive ability or cognitive change since childhood in e4 carriers. Interestingly, the lack of moderation matches most previous studies that have measured cognition with multiple tests along with an ordinal or continuous depressive symptoms variable (Crook et al., 2018; Köhler et al., 2010; Luciano et al., 2015; but see

Rajan et al., 2014) and contrasts with studies that have used cognitive impairment screeners or cognitive impairment diagnoses as outcomes (Corsentino et al., 2009; Geda et al., 2006; Irie et al., 2008; Karlsson et al., 2015; Kim et al., 2011; Niti et al., 2009). It is possible, then, that depressive symptoms are more strongly negatively related to pathological cognitive ageing in e4 carriers, but are not more strongly negatively related to ability in cognitively healthy individuals. This would follow from studies of amyloid-associated depression, which is characterised by a high amyloid-β (Aβ) 40 to Aβ 42 peptides (Sun et al., 2008). Amyloid-associated depression has been related to broader cognitive impairment than nonamyloid depression (Sun et al., 2008) and it may indicate greater risk for Alzheimer's disease in e4 carriers (Qiu et al., 2015).

7.5.2 Neuroticism

Overall, the results suggested neuroticism did not have a moderate strength relationship to poorer cognitive ability.

The expected effect size for the neuroticism effects was -0.25, which was the estimate for the relationship between neuroticism and general cognitive ability in the Lothian Birth Cohort 1936 (Crook et al., 2018). This was stronger than the other expected effect sizes. Future studies should adjust the expected effect for this main effect based on our results, as our study suggests it is unlikely that the relationship is this strong.

After removing general cognitive variance, neuroticism seemed to be related to *better* immediate and delayed recall in e4 carriers, both before and after controlling for childhood intelligence; these interactions were statistically significant and *positive*, the opposite to the hypothesised direction. Future studies into the e4-by-neuroticism interaction should consider the cognitive tests used.

7.5.3 Cortisol

The cortisol measure with the most notable results was the total cortisol production during the waking day. Against expectations, it was not related to worse general cognitive ability. It also displayed interesting patterns of results in the models predicting cognitive test residuals. For example, as hypothesised, it was related to poorer semantic fluency ability, both before and after controlling for childhood intelligence. Against expectations, though, this relationship was stronger in e4 non-carriers rather than e4 carriers; the e4 interactions in these models were statistically significant in the unexpected direction (positive). Also against expectations, producing more cortisol over the day was not related to poorer symbol digit modalities residuals. This relationship was positive, so although the e4 interaction was supported in these models, it made the association less positive rather than more negative in e4 carriers. If the effects studied here exist and are related primarily to the total amount of cortisol produced rather than cortisol dynamics, it would follow that the total cortisol production during the waking day, which incorporates all cortisol production throughout the waking day, should be the cortisol measure most closely linked to cognitive variables. Future studies on cortisol-cognition relationships should consider using this cortisol measure and should aim to clarify relationships involving it and cognitive abilities.

Notably, the relationship between a flat diurnal cortisol slope and poorer immediate recall was stronger in e4 carriers, which replicated one of Gerritsen et al. (2011)'s findings.

A steeper cortisol awakening response (both with and without control for waking cortisol) and a flatter diurnal cortisol slope were related to poorer immediate recall residual scores. The latter finding replicates Gerritsen et al. (2011), who used raw immediate recall scores rather than residuals from a general factor model. On the other hand, waking cortisol and the total cortisol produced during the waking day were not related to worse performance on immediate recall in e4 carriers; parameter estimates for these interactions were positive and small-to-moderate in effect size.

Finally, we did not replicate Evans et al. (2012). Their study found a greater cortisol awakening response was related to better executive functioning (Evans et al., 2012). However, in our main model without control for waking cortisol levels, the association between the cortisol awakening response and semantic fluency residuals was close to zero. After controlling for waking cortisol, the association was negative and the Bayes factor provided notable evidence for the null hypothesis. Hence, our data suggested no relationship is more likely than a positive relationship regarding cortisol awakening response and executive functioning.

7.5.4 Strengths and weaknesses

One methodological strength of this study was the calculation of Bayes factors for each hypothesis. This made it possible for us to find evidence for the null hypothesis, and made it clear when this was the case compared to when there was simply no clear evidence for either the null or the alternative hypothesis. We encourage the interested reader to consider their own alternative hypotheses for the main and interaction effects being tested here and calculate their own Bayes factors. One can do this by entering their prior expectations about the relevant effect, along with the estimate and standard error from this study, into an online calculator (Dienes, n.d.). Alternatively, or in addition, the reader can decide on their prior odds for the alternative theories we present. These prior odds can then be multiplied by the Bayes factors obtained to produce the reader's posterior odds for the probability of the theory.

A strength of this study is that a measure of age 11 intelligence was available, which enabled us to test the relationships of older age depressive symptoms, neuroticism and cortisol with long-term change in cognitive ability. Another strength was the number and range of measurements available in older age. Having three cortisol samples enables different aspects of cortisol secretion to be investigated. Plus, having several cognitive test results enabled a strong general factor of cognitive ability to be fitted using structural equation modelling, and relationships with general and specific cognitive variance to be separately tested. In addition, having item-level neuroticism data enabled a latent factor of neuroticism, rather than a sum score, to be used in model testing. Unlike sum scores, latent variable scores weight the contributions of indicators by how strongly they relate to the factor (DiStefano, Zhu, & Mîndrilă, 2009). This means that the resulting factor scores are more valid estimates of the factor than sum scores, because they are not contaminated by measurement error or by the variance specific to each indicator (DiStefano et al., 2009).

The main weakness of this study was the sample size, which was insufficient to provide substantial evidence for either the null or the alternative hypothesis in most cases. The study would also have benefitted from longitudinal measurements of the predictors and the outcome in older age, as this would have enabled the relationships between changes in these variables over time to be analysed.

Another limitation of the study is that the cognitive test used in childhood differed from those used in older age. This meant that our tests of effects on cognitive change since childhood were only the best available approximations of such effects. The approximation worked well for cognitive outcomes where age 11 IQ was strongly linked to older age ability, such as general cognitive ability, but for the residuals of some cognitive tests, not having childhood results on the same test limited the study's capability to test effects on change.

Finally, a limitation of this study is that it did not test whether the relationships between the predictors and the outcomes were nonlinear. Spurious interaction effects are sometimes found when one of the predictors has a nonlinear relationship with the outcome and this nonlinearity is omitted from the model (Belzak & Bauer, 2019). This was less likely in our study because the sample size was relatively large and the *APOE*-predictor correlations were not very strong (Belzak & Bauer, 2019). Future studies should consider testing nonlinear relationships between the interacting predictors and the outcome, particularly in studies where the sample size is small.

7.5.5 Directions for future research

Future research could apply a similar design to a larger sample of older adults, so that substantial evidence would be produced for more hypotheses. Researchers could also recalibrate the effect sizes expected by alternative hypotheses based on our results and/or their personal beliefs.

Ideally, future studies should include continuous subclinical measures, like the questionnaire and tests used in our study, as well as clinical measures of depression and cognition. This would enable a nuanced approach to be taken to studying these relationships.

7.5.5.1 Incorporating amyloid-beta measurement

Also worth exploring are possible pathways through which these potential moderations could function. A prime candidate pathway would be Aβ load.

Studies have reported that the relationship between Aβ load and cognition is moderated by e4 status such that in e4 carriers, Aβ load has a medium-strength negative relationship with cognition, but in non-e4 carriers there is only a weak negative relationship if there is one at all (Kantarci et al., 2012; Lim et al., 2013). Depression has also been consistently linked to Aβ load (Harrington, Lim, Gould, & Maruff, 2015), and one study using a sample with various cognitive statuses found a positive relationship between plasma cortisol and Aβ load (Toledo et al., 2012). In addition, a recent study using a sample of cognitively healthy older adults reported that plasma cortisol moderated the effect of Aβ load on some cognitive declines such that in those with high Aβ load, decline was greater in those with high plasma cortisol (Pietrzak et al., 2016).

A study with longitudinal measures of all relevant variables could help to clarify these relationships and understand which particular relationships, if any, are moderated by e4 status. This could also help to clarify the directions of the relationships between variables, for example, whether Aβ plaques tend to precede or follow depressive episodes. In a previous example of this approach, autoregressive latent trajectory modelling was used to find that depressive symptoms are related to later memory decline, but memory was not related to later change in depressive symptoms (Zahodne, Stern, & Manly, 2014).

7.5.5.2 Glucocorticoids: Experimentation

With regard to the relationships involving cortisol, a range of related options are available if the effects of glucocorticoids in general are studied instead.

Researchers could also take an experimental approach to testing the effect and the moderation of the effect of glucocorticoids on cognition. For example, previous studies have tested the effects of acute and chronic hydrocortisone administration on cognitive ability (Porter, Barnett, Idey, McGuckin, & O'Brien, 2002; Young, Sahakian, Robbins, & Cowen, 1999). A similar design could be used with the addition of genotyping participants.

7.5.5.3 Glucocorticoids: Observing exposure

Another option would be to study people who have already been exposed to high levels of glucocorticoids. For example, one study looked at the relationship between glucocorticoid medication and psychiatric disorders (Fardet, Petersen, & Nazareth, 2012), while other researchers have used a case-control design to look at the effects of previous Cushing's disease on psychopathology and cognition (Tiemensma, Biermasz, et al., 2010; Tiemensma, Kokshoorn, et al., 2010). To study the moderation of relations by e4 status, the genotyping of participants would be added to the design.

7.5.6 Conclusion

In summary, we investigated the relationships of depressive symptoms, neuroticism and cortisol with cognitive ability in older age, including the moderation of these relationships by *APOE* e4 allele possession. Depressive symptoms were related to poorer general cognitive ability, and, beyond that, to poorer immediate recall ability. However, depressive symptoms were not more strongly negatively related to general cognitive ability or immediate recall ability in *APOE* e4 carriers, in both models with and without control for childhood intelligence. Against expectations, neuroticism and the total cortisol produced during the waking day were not related to poorer general cognitive ability. A steeper cortisol awakening response and a flatter diurnal cortisol slope were related to worse immediate recall ability in e4 carriers. However, there were also cortisol effects on specific cognitive abilities that were not moderated by *APOE* e4 status.

Chapter 8

Conclusion

This chapter will begin by summarising the key findings of each study in the thesis. The chapter will then consider the wider applicability of the findings of this thesis. The limitations of our findings will be discussed. Following this, we will identify potentially fruitful future directions for allostatic load research. Finally, we will summarise the conclusions and discussion points of this chapter.

8.1 Summary of findings

The studies in Chapters 4-6 applied latent growth curve modelling to LBC1936 study data to test the contemporaneous (age ~70) and associated change (ages ~70 to ~79) relationships of allostatic load and other physiological and psychological variables.

In Chapter 4, we tested the contemporaneous and longitudinal relationships of allostatic load and three other physiological weathering measures: biological age, epigenetic age, and telomere length. The two most closely related measures were allostatic load and biological age, which are calculated using mostly the same biomarkers. Contemporaneously measured allostatic load and biological age were moderately correlated. Change in the two variables over time was strongly correlated. However, aside from allostatic load and biological age, the physiological weathering measures tended to be only weakly correlated, so the measures seem to mostly index different aspects of physiological weathering.

The next two chapters tested the associations of allostatic load with psychological variables. The Chapter 5 study found that higher allostatic load was correlated with lower general cognitive ability, although with only a small effect size. After controlling for participant dropout, the effect size was reduced by approximately a third, which suggested that the correlation was partially driven by those who dropped out of the study. The study also tested the correlation between change in allostatic load and change in general cognitive ability, but the estimates for this parameter were not large and standard errors were wide relative to the estimates. Next, the Chapter 6 study found that higher allostatic load was correlated with contemporaneously depressive symptoms and the effect size was small-to-moderate. As with the previous chapters, estimates of longitudinal relationships produced small effect sizes and wide standard errors.

Finally, the study in Chapter 7 fitted structural equation models to 36-Day Sample study data. This study tested the relationships of cognitive ability measures with cortisol measures, depressive symptoms, and neuroticism, and whether these relationships were moderated by possession of the APOE e4 allele. The study used Bayes factors to calculate the evidence the data provided for alternative and null hypotheses about the relationships. The total cortisol production during the waking period was related to higher general cognitive ability, with a small effect size. The total cortisol produced during the entire waking day was not related to general cognitive ability. Additionally, depressive symptoms were related to lower general cognitive ability and this relationship was not moderated by APOE e4.

8.2 Is allostatic load clinically useful for psychological conditions?

Allostatic load researchers have often suggested that allostatic load summary scores may be clinically useful for psychological conditions. For example, Seeman et al. (1997) studied cognitive ageing, as well as physical functioning and cardiovascular disease, and concluded that "the concept of allostatic load may provide the basis for a more comprehensive assessment of major risks in the aging process" (Seeman et al., 1997, p. 2259). Regarding depressive symptoms, Juster et al. (2011) posited that allostatic load scores could be used to identify and more effectively treat individuals at risk of depressive symptoms.

The studies in Chapters 5 and 6 used a non-clinical sample and measures of allostatic load, cognitive ability and depressive symptoms that provided substantial information about sub-clinical individual differences. The results of these studies, then, cannot be generalised to clinical populations. The results can, however, inform as to whether allostatic load scores would greatly help clinicians hoping to identify older adults at risk of cognitive ageing and depressive symptoms. We did find that allostatic load was related to contemporaneously measured general cognitive ability and depressive symptoms. Importantly, though, effect sizes for these relationships were small. Estimates were also small for the dual change relationships in these studies and the confidence intervals for these parameters were wide. Due to the small sizes of these effects, we must conclude that allostatic load scores offer limited clinical utility for cognitive ageing and depressive illness.

Nevertheless, the allostatic load *concept* may still have some clinical utility for these conditions. As the effect sizes are small, large datasets will be required to leverage the allostatic load model. For example, as referenced in Chapter 6, a machine learning algorithm helped to identify the allostatic load biomarker serum glucose as a novel marker for depressive symptoms in a large cohort study (Dipnall et al., 2016). Analyses of even larger datasets, such as national electronic health records (see Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014 for an overview of big data and health research), could benefit from the allostatic load model. For example, a recent study using electronic health records from the UK National Health Service to predict first cardiovascular events found that machine learning algorithms incorporating additional allostatic load biomarkers from the metabolic system improved prediction of cardiovascular events relative to established formulas (Weng, Reps, Kai, Garibaldi, & Qureshi, 2017). This study may have benefitted from including primary mediators of allostatic load, such as cortisol, as plasma cortisol was found to be a cause of cardiovascular disease in a recent Mendelian randomisation study (Crawford, Timpson, Smith, & Walker, 2015). Similar approaches could potentially be applied to predicting cognitive ageing and depressive symptoms.

In summary, allostatic load scores seem to offer little clinical utility for cognitive ability and depressive symptoms. The allostatic load model, though, points towards markers and pathways that can improve prediction of non-psychological conditions and show promise for the prediction of psychological conditions.

8.3 Relationships with cognitive ability: allostatic load vs. cortisol

If allostatic load summary scores are to be clinically useful, then they should relate more strongly to outcomes of interest than single biomarkers. In this thesis, we used data from two similar older age samples to test how cognitive ability relates to allostatic load scores and various cortisol measures. Comparing the effect sizes from these studies enables us to roughly assess whether the effect sizes are similar or suggest that cognitive ability relates substantially more strongly to allostatic load or cortisol. For contemporaneously measured general cognitive ability, the correlation with allostatic load in the LBC1936 was -.13 (mean age at Wave 1 = 69.5 years), while the mean standardised beta for cortisol measures in the 36-Day-Sample was .09 (mean age = 78.2 years). After dropout was added to the model, the allostatic load-general cognitive ability correlation was attenuated to -.08. In this model, the correlation was the relationship for those who did not drop out by Wave 4 (mean age = 79.3 years), when participants were close in mean age to the 36-Day-Sample at the time of their cognitive tests. In conclusion, the effect sizes for these relationships all tend to be small and do not suggest that allostatic load summary scores have substantially greater clinical utility than cortisol measures for older age cognitive ability. This conclusion chimes with the previous research described in the introduction of Chapter 5; studies relating cognitive ability to allostatic load summary measures and individual allostatic load biomarkers have both tended to report small-to-moderate effect sizes.

8.4 On our allostatic load operationalisation

Our LBC1936 allostatic load operationalisation used biomarkers from the cardiovascular, inflammatory/immune and metabolic systems (including BMI, an anthropometric marker; see Juster et al., 2010, Figure 4 for a categorisation of allostatic load biomarkers). It did not include any neuroendocrine biomarkers, nor any primary mediators of the stress response (Juster et al., 2010; Seplaki et al., 2005). This is a notable limitation, because neuroendocrine markers and other primary mediators differentiate allostatic load measures from other biomarker composites based on clinical markers (Seeman et al., 2001), including metabolic syndrome summaries (Eisenmann, 2008) and the Framingham Risk Score (D'Agostino et al., 2008). Primary mediators have been found to improve prediction of mortality as well as physical and cognitive declines (Goldman, Turra, et al., 2006; Karlamangla et al., 2002; Seeman et al., 2001).

Our allostatic load operationalisation then, might also be thought of as an expanded measure of metabolic syndrome, or more generally as a measure of multisystem physiological dysregulation. Note that these limitations were part of a necessary trade-off, as cohort studies that do have allostatic load data on primary mediators lack the repeated measures and/or the other variables required to answer most of our research questions. It is extremely rare to have, as the LBC1936 has, four waves of data on the same allostatic load biomarkers, DNA methylation, telomere length, and an extensive battery of gold-standard intelligence tests.

It is worth considering what different allostatic load operationalisations imply about the importance of different biomarkers and physiological systems. The calculation methods used in most allostatic load studies, including ours, weight every biomarker equally. For example, in a study using sample quartile-based cutoffs, for all biomarkers used a biomarker value in the quartile indicating highest risk would add 1 to an individual's allostatic load score. This means that the contribution of each physiological system to the overall score will depend on the biomarkers included in the score calculation.

However, many allostatic load studies, including ours, use secondary data. In a secondary data study, the researchers must work with a previously determined biomarker set. If the researcher uses all the measured allostatic load biomarkers are when calculating allostatic load scores, the contribution of each physiological system to allostatic load scores will also be dictated by the data collection protocol. If the researcher chooses to use only a subset of the measured allostatic load biomarkers, the available weightings of each physiological system can also be limited; for example, in our study, no neuroendocrine biomarkers were collected, so the contribution of this system was fixed at zero. Future allostatic load studies with neuroendocrine biomarkers could test their hypotheses using allostatic load scores computed using different biomarker sets, to assess the effects of including neuroendocrine biomarkers and varying the weighting of different physiological systems.

Perhaps the most commonly used allostatic load operationalisation with differential weighting of biomarkers is factor analysis. This approach, however, has various problems, as outlined in Chapter 2. In Section 8.6, we will mention an upcoming test of a novel allostatic load operationalisation with differential weighting of biomarkers.

8.5 Constraints on generality

Characteristics of our studies' samples and designs place constraints on how broadly our findings can be generalised. We were interested in processes of change in older age, so we used samples that were older aged and homogeneous in age. This means that our findings may not generalise to other age groups. In particular, allostatic load may relate differently over time to biological age and epigenetic age in age-heterogeneous samples.

Results from our models that assumed missing data were not missing at random hint at another constraint on generality. The most notable difference in these models was that the baseline correlation between allostatic load and general cognitive ability was attenuated by around a third after dropout was included in the model, which suggests that the correlation was weaker in participants who remained healthy enough to participate in the study up to around age 79. There are two main aspects to this selection effect: a survivor effect and a participation effect. Around 75% of Scottish Mental Survey 1947 participants survived to age 70 (Čukić et al., 2017). By age 79, the proportion of survivors was reduced to around 55% (Čukić et al., 2017). Regarding participation, most potential participants contacted declined to participate (Deary et al., 2007). Also, generally, those who do participate in epidemiological studies tend to be more educated and have higher socioeconomic status (Galea & Tracy, 2007). An example comes from the 6-Day-Sample, which we used data from in Chapter 7. The mean difference in age 11 Terman-Merill IQ between the full sample and those who survived to age 77 was 2.4 IQ points, while the mean difference between those who survived and those who participated in the follow-up study was a substantially larger 8.8 IQ points (Johnson, Deary, & Bouchard, 2017). Both survivor and participation effects, then, made our follow-up sample less-than-representative of the nearly whole population studied in the Scottish Mental Survey 1947. Future studies of allostatic load and general cognitive ability in older age should keep in mind that this relationship may be weaker in participants willing and able to participate at older ages.

Also, the samples we used were ethnically homogeneous as all participants were white. Previous research has observed differences in allostatic load between ethnic groups (e.g., Geronimus, Hicken, Keene, & Bound, 2006), so if these differences were not mirrored on the other variables studied here, then the correlations observed may differ in a sample with a different ethnic makeup.

Our findings may also not generalise to studies using other allostatic load and cortisol calculation methods. One example is that the contemporaneous and longitudinal correlations between allostatic load and biological age would likely be smaller if a non-continuous operationalisation of allostatic load was used.

8.6 Future directions

Many interesting and worthwhile studies could follow from those presented here. Here, we will focus on three broader issues for future research: the choice of operationalisations, data size, and the length of time between measurements.

Debate remains around which allostatic load operationalisation is the most appropriate. We have preregistered a study (Crook & Booth, 2017a) that aims to clarify how closely correlated allostatic load scores from different operationalisations are, as well as how strongly different allostatic operationalisations relate to three key criterion variables: depressive symptoms, self-rated health, and mortality. The study will calculate allostatic load scores using nine calculation methods, including a novel method based on weighting biomarkers' score contribution by their association with perceived psychological stress. This new operationalisation has some parallels with Klemera-Doubal biological age, which weights biomarkers' contribution to biological age estimates based on how closely they relate to chronological age.

We will apply the nine calculation to four biomarker sets: primary mediators of allostatic load; secondary outcomes of allostatic load; a full set of available allostatic load biomarkers; and a smaller set of more commonly measured biomarkers. The correlations between scores and comparisons of criterion variable associations should help to elucidate whether a secondary outcome-based operationalisation like the one in this thesis should be expected to produce similar scores and results to another operationalisation.

It is also worth noting that results will be affected by how the other study variables are operationalised. For example, in Chapter 7, we tested associations between cortisol measures and a factor of neuroticism. However, neuroticism has various facets, including anxiety, depression, angry hostility, and impulsiveness (Costa Jr. & McCrae, 1995). These facets likely have different relationships with cortisol levels and dynamics. For example, a study of adolescents found that vulnerability and impulsiveness, but not angry hostility, were related to higher basal cortisol, while no relationships were found with other cortisol measures (Laceulle, Nederhof, van Aken, & Ormel, 2015). Hence, if we had chosen to operationalise neuroticism using its facets, our results would have differed. The thesis's results would also have differed if we operationalised other variables, such as biological age or depressive symptoms, using different indicator variables or calculation methods.

Another important issue is data size. Allostatic load is associated with psychological variables, such as cognitive ability and depressive symptoms, but with only small-to-moderate effect sizes. Also, allostatic load, cognitive ability and depressive symptoms are multifaceted, so there are complex relationships underlying the associations observed between summary variables. This means that where hypotheses are being tested, larger datasets will be required to provide strong evidence for the null or alternative hypotheses. For example, the study in Chapter 7 did not produce strong evidence for the null or the alternative hypothesis for most parameters tested because of the modest sample size. When planning data collection or acquisition, researchers tend to focus on obtaining an adequate sample size. Often, less focus is placed on the various other study design variables that can affect statistical power. In longitudinal studies, the dataset's size and shape, not just the sample size, are important: statistical power can be improved by increasing the number of repeated measures. For instance, the dual change associations tested in this thesis often had large standard errors, which would have been reduced if the same participants had been observed on more occasions (Hertzog et al., 2006).

The statistical power to detect correlated change is also affected by the growth rate reliability, which depends on the proportion of between-individual variance as well as the duration of the study, the number of measurements, and the timing of measurements (Willett, 1989). Hence, researchers can increase the power of a study to detect a slope-slope correlation by carefully planning the timing of measurements (Rast & Hofer, 2014). For example, intensive measurement burst designs early in a study can improve power in the earlier phases of a longitudinal study (Rast & Hofer, 2014). Future research using secondary data should take care to note where models have not precisely estimated a parameter of interest or hypothesis tests have not provided strong evidence for or against the focal hypothesis. Future research collecting new data should use power calculations and/or data simulation, combined with estimates from previous studies, to ensure that their studies will provide adequate answers to their research questions.

Finally, to better understand the relationships investigated in this thesis, future research should collect data with different lengths of time between measurements. In our longitudinal studies, we focused on long-term changes in older age. The LBC1936 has a typical cohort study design, where many variables are measured at each wave and the waves are a few years apart, so it was ideal for this purpose. Also, both the LBC1936 and the 36-Day-Sample enabled us to test associations with age 11 IQ, an example of the advantages of uncommon gaps between measurements. Going forward, researchers planning cohort studies should consider having some measurements weeks, days or hours apart rather than years apart. This would help to elucidate how allostatic load and its biomarkers relate to other variables at shorter timespans. For example, one study measuring blood pressure, heart rate and affect measured each variable four times each working day for two weeks (Ilies, Dimotakis, & Watson, 2010). Another measured burnout symptoms every day for two weeks and measured the cortisol awakening response and dehydroepiandrosterone-sulphate on two days, as well as the dexamethasone-suppressed cortisol awakening response on a third day (Sonnenschein et al., 2007). Importantly, these short-term bursts of repeated measures can be included within a cohort study that has longer gaps between its main measurement waves (Sliwinski, 2008). The Effects of Stress on Cognitive Aging, Physiology and Emotion project, for instance, has four 14-day measurement burst periods across a three-year period (Scott et al., 2015). Each wave includes smartphone surveys and cognitive tests throughout each day, five days where four saliva samples are delivered, and a lab visit for further tests (Scott et al., 2015). Particularly if researchers wish to use burst sampling on a subset of a larger cohort, representative sampling can be used alongside planned missingness designs to maximise statistical power for a given budget (Rhemtulla, Savalei, & Little, 2016). In summary, cohort studies with years between measurements provide only a snapshot of allostatic load levels. Designs incorporating measurement of allostatic load biomarker measurements with shorter intervals will help to elucidate how allostatic load develops and how it relates to cognitive ability and depressive symptoms within the day or the week rather than only over periods of years.

8.7 Summary

Although this thesis found allostatic load was associated with contemporaneously measured general cognitive ability and depressive symptoms, the effect sizes were small. The contemporaneous association between allostatic load and general cognitive ability in the LBC1936 was of a similar size to the average association between cortisol and general cognitive ability in the 36-Day-Sample. Hence, our results suggest that for cognitive ability and depressive symptoms, allostatic load summary scores have limited clinical utility. The generalisability of our findings is constrained by aspects of our sample and design. This includes our allostatic load operationalisation, as scores were calculated from allostatic load secondary outcome biomarkers, and no primary mediators of allostatic load were included. To better understand correlations between allostatic load scores from different operationalisations, as well as how the use of different operationalisations affects relationships with criterion variables, we preregistered and will conduct an allostatic load operationalisation comparison study. Future substantive studies should carefully consider their allostatic load operationalisation(s) and our preregistered study will assist with that. To ensure that they can answer their research questions, studies that collect new data should also ensure that they collect enough data and collect data at the timescales relevant to their interests.

Bibliography

Adam, E. K., & Kumari, M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology*, *34*(10), 1423–1436. https://doi.org/10.1016/j.psyneuen.2009.06.011

Adam, E. K., Quinn, M. E., Tavernier, R., McQuillan, M. T., Dahlke, K. A., & Gilbert, K. E. (2017). Diurnal cortisol slopes and mental and physical health outcomes: A systematic review and meta-analysis. *Psychoneuroendocrinology*, *83*, 25–41. https://doi.org/10.1016/j.psyneuen.2017.05.018

Administering IPIP measures, with a 50-item sample questionnaire. (n.d.). Retrieved from http://ipip.ori.org/New_IPIP-50-item-scale.htm

Ahrens, K. A., Rossen, L. M., & Simon, A. E. (2016). Relationship between mean leucocyte telomere length and measures of allostatic load in US reproductive-aged women, NHANES 1999-2002. *Paediatric and Perinatal Epidemiology*, *30*(4), 325–335. https://doi.org/10.1111/ppe.12277

Alberti, K. G. M. M., Eckel, R. H., Grundy, S. M., Zimmet, P. Z., Cleeman, J. I., Donato, K. A., ... Smith, S. C. (2009). Harmonizing the metabolic syndrome. *Circulation*, *120*(16), 1640–1645. https://doi.org/10.1161/circulationaha.109.192644

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., ... Chang, W. (2018). *rmarkdown: Dynamic documents for R*. Retrieved from https://CRAN.R-project.org/package=rmarkdown

Almela, M., van der Meij, L., Hidalgo, V., Villada, C., & Salvador, A. (2012). The cortisol awakening response and memory performance in older men and women. *Psychoneuroendocrinology*, *37*(12), 1929–1940. https://doi.org/10.1016/j.psyneuen.2012.04.009

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, *30*(10), 1363–1369. https://doi.org/10.1093/bioinformatics/btu049

Badrick, E., Bobak, M., Britton, A., Kirschbaum, C., Marmot, M., & Kumari, M. (2008). The relationship between alcohol consumption and cortisol secretion in an aging cohort. *The Journal of Clinical Endocrinology & Metabolism*, *93*(3), 750–757. https://doi.org/10.1210/jc.2007-0737

Badrick, E., Kirschbaum, C., & Kumari, M. (2007). The relationship between smoking status and cortisol secretion. *The Journal of Clinical Endocrinology & Metabolism*, *92*(3), 819–824. https://doi.org/10.1210/jc.2006-2155

Baguley, T., & Kaye, D. (2010). Book review. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 695–698. https://doi.org/10.1348/000711009x481027

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, *33*(7), 1123–1131. https://doi.org/10.1377/hlthaff.2014.0041

Beal, D. J. (2015). ESM 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 383–407. https://doi.org/10.1146/annurev-orgpsych-032414-111335

Beck, S., & Rakyan, V. K. (2008). The methylome: Approaches for global DNA methylation profiling. *Trends in Genetics*, *24*(5), 231–237. https://doi.org/10.1016/j.tig.2008.01.006

Belsky, D. W., Caspi, A., Houts, R., Cohen, H. J., Corcoran, D. L., Danese, A., ... Moffitt, T. E. (2015). Quantification of biological aging in young adults. *Proceedings of the National Academy of Sciences*, *112*(30), E4104–E4110. https://doi.org/10.1073/pnas.1506264112

Belsky, D. W., Moffitt, T. E., Cohen, A. A., Corcoran, D. L., Levine, M. E., Prinz, J. A., ... Caspi, A. (2018). Eleven telomere, epigenetic clock, and biomarker-composite quantifications of biological aging: Do they measure the same thing? *American Journal of Epidemiology*, *187*(6), 1220–1230. https://doi.org/10.1093/aje/kwx346

Beluche, I., Carrière, I., Ritchie, K., & Ancelin, M. L. (2010). A prospective study of diurnal cortisol and cognitive function in community-dwelling elderly people. *Psychological Medicine*, *40*(06), 1039–1049. https://doi.org/10.1017/s0033291709991103

Belzak, W. C. M., & Bauer, D. J. (2019). Interaction effects may actually be nonlinear effects in disguise: A review of the problem and potential solutions. *Addictive Behaviors*, *94*, 99–108. https://doi.org/10.1016/j.addbeh.2018.09.018

Blackburn, E. H., Epel, E. S., & Lin, J. (2015). Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science*, *350*(6265), 1193–1198. https://doi.org/10.1126/science.aab3389

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological*

Methods, *16*(3), 265–284. https://doi.org/10.1037/a0024448

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. London, UK: John Wiley & Sons.

Bollen, K. A., & Noble, M. D. (2011). Structural equation models and the quantification of behavior. *Proceedings of the National Academy of Sciences*, *108*(Supplement 3), 15639–15646. https://doi.org/10.1073/pnas.1010661108

Boonekamp, J. J., Simons, M. J. P., Hemerik, L., & Verhulst, S. (2013). Telomere length behaves as biomarker of somatic redundancy rather than biological age. *Aging Cell*, *12*(2), 330–332. https://doi.org/10.1111/acel.12050

Booth, T., Bastin, M. E., Penke, L., Maniega, S. M., Murray, C., Royle, N. A., ... Deary, I. J. (2013). Brain white matter tract integrity and cognitive abilities in community-dwelling older people: The Lothian Birth Cohort, 1936. *Neuropsychology*, *27*(5), 595–607. https://doi.org/10.1037/a0033354

Booth, T., Royle, N. A., Corley, J., Gow, A. J., C. Valdés Hernández, M. del, Maniega, S. M., ... Deary, I. J. (2015). Association of allostatic load with brain structure and cognitive ability in later life. *Neurobiology of Aging*, *36*(3), 1390–1399. https://doi.org/10.1016/j.neurobiolaging.2014.12.020

Booth, T., Starr, J. M., & Deary, I. (2013). Modeling multisystem biological risk in later life: Allostatic load in the Lothian Birth Cohort study 1936. *American Journal of Human Biology*, *25*(4), 538–543. https://doi.org/10.1002/ajhb.22406

Borsboom, D. (2017). The meaning of model equivalence: Network models, latent variables, and the theoretical space in between. https://bit.ly/2RDfv3N.

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Breitling, L. P., Saum, K.-U., Perna, L., Schöttker, B., Holleczek, B., & Brenner, H. (2016). Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clinical Epigenetics*, *8*(1). https://doi.org/10.1186/s13148-016-0186-5

Brett, C. E., & Deary, I. J. (2014). Realising health data linkage from a researcher's perspective: Following up the 6-Day Sample of the Scottish Mental Survey 1947. *Longitudinal and Life Course Studies*, *5*(3), 283–298. https://doi.org/10.14301/llcs.v5i3.266

Bryer, J., & Speerschneider, K. (2015). *likert: Functions to analyze and visualize likert type items*. Retrieved from https://CRAN.R-project.org/package=likert

Bunce, D., Batterham, P. J., Mackinnon, A. J., & Christensen, H. (2012). Depression, anxiety and cognition in community-dwelling adults aged 70 years and over. *Journal of Psychiatric Research*, *46*(12), 1662–1666. https://doi.org/10.1016/j.jpsychires.2012.08.023

Burke, H. M., Fernald, L. C., Gertler, P. J., & Adler, N. E. (2005). Depressive symptoms are associated with blunted cortisol stress responses in very low-income women. *Psychosomatic Medicine*, *67*(2), 211–216. https://doi.org/10.1097/01.psy.0000156939.89050.28

Burt, D. B., Zembar, M. J., & Niederehe, G. (1995). Depression and memory impairment: A meta-analysis of the association, its pattern, and specificity. *Psychological Bulletin*, *117*(2), 285–305.

Büchtemann, D., Luppa, M., Bramesfeld, A., & Riedel-Heller, S. (2012). Incidence of late-life depression: A systematic review. *Journal of Affective Disorders*, *142*(1-3), 172–179. https://doi.org/10.1016/j.jad.2012.05.010

Cambridge Institute of Public Health. (2016). *Changing risk behaviours and promoting cognitive health in older adults: An evidence-based resource for local authorities and commissioners*. London, UK: Public Health England.

Cassano, P., & Fava, M. (2002). Depression and public health. *Journal of Psychosomatic Research*, *53*(4), 849–857. https://doi.org/10.1016/s0022-3999(02)00304-5

Center for Disease Control and Prevention. (2017). Retrieved from https://www.cdc.gov/aging/healthybrain/index.htm

Chen, B. H., Marioni, R. E., Colicino, E., Peters, M. J., Ward-Caviness, C. K., Tsai, P.-C., ... Horvath, S. (2016). DNA methylation-based measures of biological age: Meta-analysis predicting time to death. *Aging*, *8*(9), 1844–1865. https://doi.org/10.18632/aging.101020

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chida, Y., & Steptoe, A. (2009). Cortisol awakening response and psychosocial factors: A systematic review and meta-analysis. *Biological Psychology*, *80*(3), 265–278. https://doi.org/10.1016/j.biopsycho.2008.10.004

Cho, I. H., Park, K. S., & Lim, C. J. (2010). An empirical comparative study on biological age estimation algorithms with an application of Work Ability Index (WAI). *Mechanisms of Ageing and Development*, *131*(2), 69–78. https://doi.org/10.1016/j.mad.2009.12.001

Chodosh, J., Kado, D. M., Seeman, T. E., & Karlamangla, A. S. (2007). Depressive symptoms as a predictor of cognitive decline: MacArthur Studies of Successful Aging. *The American Journal of Geriatric Psychiatry*, *15*(5), 406–415. https://doi.org/10.1097/01.jgp.0b013e31802c0c63

Chodosh, J., Miller-Martinez, D., Aneshensel, C. S., Wight, R. G., &

Karlamangla, A. S. (2010). Depressive symptoms, chronic diseases, and physical disabilities as predictors of cognitive functioning trajectories in older americans. *Journal of the American Geriatrics Society*, *58*(12), 2350–2357. https://doi.org/10.1111/j.1532-5415.2010.03171.x

Christie, J. (2011). bayesFactorCalc2.R. Retrieved from http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/bayesFactorCalc2.R

Comijs, H. C., Gerritsen, L., Penninx, B. W., Bremmer, M. A., Deeg, D. J., & Geerlings, M. I. (2010). The association between serum cortisol and cognitive decline in older persons. *The American Journal of Geriatric Psychiatry*, *18*(1), 42–50. https://doi.org/10.1097/jgp.0b013e3181b970ae

Corsentino, E. A., Sawyer, K., Sachs-Ericsson, N., & Blazer, D. G. (2009). Depressive symptoms moderate the influence of the ApolipoproteinE e4 allele on cognitive decline in a sample of community dwelling older adults. *The American Journal of Geriatric Psychiatry*, *17*(2), 155–165. https://doi.org/10.1097/jgp.0b013e31818f3a6b

Cosco, T. D., Doyle, F., Ward, M., & McGee, H. (2012). Latent structure of the Hospital Anxiety And Depression Scale: A 10-year systematic review. *Journal of Psychosomatic Research*, *72*(3), 180–184. https://doi.org/10.1016/j.jpsychores.2011.06.008

Costa Jr., P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*(1), 21–50. https://doi.org/10.1207/s15327752jpa6401_2

Cox, S. R., Bastin, M. E., Ferguson, K. J., Maniega, S. M., MacPherson, S. E., Deary, I. J., … MacLullich, A. M. (2015). Brain white matter integrity and cortisol in older men: The Lothian Birth Cohort 1936. *Neurobiology of Aging*, *36*(1), 257–264. https://doi.org/10.1016/j.neurobiolaging.2014.06.022

Cox, S. R., MacPherson, S. E., Ferguson, K. J., Royle, N. A., Maniega, S. M., C. Valdés Hernández, M. del, ... Deary, I. J. (2015). Does white matter structure or hippocampal volume mediate associations between cortisol and cognitive ageing? *Psychoneuroendocrinology*, *62*, 129–137. https://doi.org/10.1016/j.psyneuen.2015.08.005

Crawford, A., Timpson, N., Smith, G. D., & Walker, B. (2015). 19 testing causality in the association of plasma cortisol with risk of coronary heart disease: A mendelian randomisation study. *Heart*, *101*(Suppl 6), A6.4–A7. https://doi.org/10.1136/heartjnl-2015-308734.19

Crimmins, E. M., Johnston, M., Hayward, M., & Seeman, T. (2003). Age differences in allostatic load: An index of physiological dysregulation. *Experimental Gerontology*, *38*(7), 731–734. https://doi.org/10.1016/s0531-5565(03)00099-8

Crook, Z., & Booth, T. (2017a). Comparing operationalisations of allostatic load. Study preregistration. Retrieved from https://doi.org/10.17605/OSF.IO/E6GXS

Crook, Z., & Booth, T. (2017b). Considering the appropriateness of the factor analytic operationalization of allostatic load. *Psychosomatic Medicine*, *79*(1), 117–119. https://doi.org/cvc5

Crook, Z., Booth, T., Cox, S. R., Corley, J., Dykiert, D., Redmond, P., ... Deary, I. J. (2018). Apolipoprotein E genotype does not moderate the associations of depressive symptoms, neuroticism and allostatic load with cognitive ability and cognitive aging in the Lothian Birth Cohort 1936. *PLOS ONE*, *13*(2), e0192604. https://doi.org/10.1371/journal.pone.0192604

Crook, Z., Booth, T., & Deary, I. (2017a). Allostatic load, biological age, epigenetic age and telomere length: How do they relate over time in the Lothian Birth Cohort 1936? https://doi.org/10.17605/OSF.IO/WMB92.

Crook, Z., Booth, T., & Deary, I. (2017b). How do allostatic load and cognitive ability relate over time in the Lothian Birth Cohort 1936? https://doi.org/10.17605/OSF.IO/2UQJK.

Crook, Z., Booth, T., & Deary, I. (2017c). How do allostatic load and depressive symptoms relate over time in the Lothian Birth Cohort 1936? https://doi.org/10.17605/OSF.IO/BWUCT.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*(2), 121–136. https://doi.org/10.1080/15248371003699969

Čukić, I., Brett, C. E., Calvin, C. M., Batty, G. D., & Deary, I. J. (2017). Childhood IQ and survival to 79: Follow-up of 94% of the Scottish Mental Survey 1947. *Intelligence*, *63*, 45–50. https://doi.org/10.1016/j.intell.2017.05.002

D'Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: The Framingham Heart Study. *Circulation*, *117*(6), 743–753. https://doi.org/10.1161/circulationaha.107.699579

Dahl, A. K., Hassing, L. B., Fransson, E. I., Gatz, M., Reynolds, C. A., & Pedersen, N. L. (2013). Body mass index across midlife and cognitive change in late life. *International Journal of Obesity*, *37*(2), 296–302. https://doi.org/10.1038/ijo.2012.37

Danese, A., & McEwen, B. S. (2012). Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiology & Behavior*, *106*(1), 29–39. https://doi.org/10.1016/j.physbeh.2011.08.019

Dantan, E., Proust-Lima, C., Letenneur, L., & Jacqmin-Gadda, H. (2008). Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. *The International Journal of Biostatistics*, *4*(1). https://doi.org/10.2202/1557-4679.1088

Dar-Nimrod, I., Chapman, B. P., Franks, P., Robbins, J., Porsteinsson, A., Mapstone, M., & Duberstein, P. R. (2012). Personality factors moderate the associations between apolipoprotein genotype and cognitive function as well as late onset Alzheimer disease. *The American Journal of Geriatric Psychiatry*, *20*(12), 1026–1035. Retrieved from http://dx.doi.org/10.1097/JGP.0b013e318267016b

Dar-Nimrod, I., Chapman, B. P., Robbins, J. A., Porsteinsson, A., Mapstone, M., & Duberstein, P. R. (2012). Gene by neuroticism interaction and cognitive function among older adults. *International Journal of Geriatric Psychiatry*, *27*, 1147–1154. Retrieved from http://dx.doi.org/10.1002/gps.3759

Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chouraki, V., Giddaluru, S., ... Deary, I. J. (2015). Genetic contributions to variation in general cognitive function: A meta-analysis of genome-wide association studies in the CHARGE consortium (n=53 949). *Molecular Psychiatry*, *20*(2), 183–192. https://doi.org/10.1038/mp.2014.188

Davies, G., Harris, S. E., Reynolds, C. A., Payton, A., Knight, H. M., Liewald, D. C., ... Deary, I. J. (2014). A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing. *Molecular Psychiatry*, *19*, 76–87. https://doi.org/10.1038/mp.2012.159

Deary, I. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence*, *29*(5), 389–399. https://doi.org/10.1016/s0160-2896(01)00062-9

Deary, I. J., & Brett, C. E. (2015). Predicting and retrodicting intelligence between childhood and old age in the 6-Day Sample of the Scottish Mental Survey 1947. *Intelligence*, *50*, 1–9. https://doi.org/10.1016/j.intell.2015.02.002

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., ... Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, *92*(1), 135–152. https://doi.org/10.1093/bmb/ldp033

Deary, I. J., Gow, A. J., Pattie, A., & Starr, J. M. (2012). Cohort profile: The Lothian Birth Cohorts of 1921 and 1936. *International Journal of Epidemiology*, *41*(6), 1576–1584. https://doi.org/10.1093/ije/dyr197

Deary, I. J., Gow, A. J., Taylor, M. D., Corley, J., Brett, C., Wilson, V., … Starr, J. M. (2007). The Lothian Birth Cohort 1936: A study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatrics*, 7(1). https://doi.org/10.1186/1471-2318-7-28

Deary, I. J., Simonotto, E., Meyer, M., Marshall, A., Marshall, I., Goddard, N., & Wardlaw, J. M. (2004). The functional anatomy of inspection time: An event-related fMRI study. *NeuroImage*, *22*(4), 1466–1479. https://doi.org/10.1016/j.neuroimage.2004.03.047

Demanelis, K., Tong, L., & Pierce, B. L. (2019). Genetically increased telomere length and aging-related physical and cognitive traits in the UK Biobank. https://doi.org/10.1101/538124

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Basingstoke, UK: Palgrave Macmillan.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z. (n.d.). Making the most of your data with Bayes. Retrieved from http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm

Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics*, *43*(1), 49. https://doi.org/10.2307/2986113

Dipnall, J. F., Pasco, J. A., Berk, M., Williams, L. J., Dodd, S., Jacka, F. N., & Meyer, D. (2016). Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PLOS ONE*, *11*(2), e0148195. https://doi.org/10.1371/journal.pone.0148195

DiStefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20). Retrieved from http://pareonline.net/getvn.asp?v=14&n=20

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, *18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Dotson, V. M., Resnick, S. M., & Zonderman, A. B. (2008). Differential association of concurrent, baseline, and average depressive symptoms with cognitive decline in older adults. *The American Journal of Geriatric Psychiatry*, *16*(4), 318–330. https://doi.org/10.1097/jgp.0b013e3181662a9c

Duong, M. T., Bingham, B. A., Aldana, P. C., Chung, S. T., & Sumner, A. E. (2017). Variation in the calculation of allostatic load score: 21 examples from NHANES. *Journal of Racial and Ethnic Health Disparities*, *4*(3), 455–461. https://doi.org/10.1007/s40615-016-0246-8

Eatough, E., Shockley, K., & Yu, P. (2016). A review of ambulatory health data collection methods for employee experience sampling research. *Applied Psychology*, *65*(2), 322–354. https://doi.org/10.1111/apps.12068

Echouffo-Tcheugui, J. B., Conner, S. C., Himali, J. J., Maillard, P., DeCarli, C. S., Beiser, A. S., ... Seshadri, S. (2018). Circulating cortisol and cognitive and structural brain measures. *Neurology*, *91*(21), e1961–e1970. https://doi.org/10.1212/wnl.0000000000006549

Economos, A., Wright, C. B., Moon, Y. P., Rundek, T., Rabbani, L., Paik, M. C., ... Elkind, M. S. (2013). Interleukin 6 plasma concentration associates with cognitive decline: The Northern Manhattan Study. *Neuroepidemiology*, *40*(4), 253–259. https://doi.org/10.1159/000343276

Eisenmann, J. C. (2008). On the use of a continuous metabolic syndrome score in pediatric research. *Cardiovascular Diabetology*, *7*(1), 17. https://doi.org/10.1186/1475-2840-7-17

Elff, M. (2016). *memisc: Tools for management of survey data and the presentation of analysis results*. Retrieved from https://CRAN.R-project.org/package=memisc

Elias, M. F., Wolf, P. A., D'Agostino, R. B., Cobb, J., & White, L. R. (1993). Untreated blood pressure level is inversely related to cognitive functioning: The Framingham Study. *American Journal of Epidemiology*, *138*(6), 353–364. https://doi.org/10.1093/oxfordjournals.aje.a116868

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: The Guilford Press.

Ennis, G. E., Moffat, S. D., & Hertzog, C. (2016). The cortisol awakening response and cognition across the adult lifespan. *Brain and Cognition*, *105*, 66–77. https://doi.org/10.1016/j.bandc.2016.04.001

Epel, E., Lapidus, R., McEwen, B., & Brownell, K. (2001). Stress may add bite to appetite in women: A laboratory study of stress-induced cortisol and eating behavior. *Psychoneuroendocrinology*, *26*(1), 37–49. https://doi.org/10.1016/s0306-4530(00)00035-4

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167

Evans, P., Fredhoi, C., Loveday, C., Hucklebridge, F., Aitchison, E., Forte, D., & Clow, A. (2011). The diurnal cortisol cycle and cognitive performance in the healthy old. *International Journal of Psychophysiology*, *79*(3), 371–377. https://doi.org/10.1016/j.ijpsycho.2010.12.006

Evans, P., Hucklebridge, F., Loveday, C., & Clow, A. (2012). The cortisol awakening response is related to executive function in older age. *International Journal of Psychophysiology*, *84*(2), 201–204.

https://doi.org/10.1016/j.ijpsycho.2012.02.008

Fardet, L., Petersen, I., & Nazareth, I. (2012). Suicidal behavior and severe neuropsychiatric disorders following glucocorticoid therapy in primary care. *American Journal of Psychiatry*, *169*(5), 491–497. https://doi.org/10.1176/appi.ajp.2011.11071009

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology*, *4*(1), 22–36. https://doi.org/10.1027/1614-2241.4.1.22

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fonda, S. J., Bertrand, R., O’Donnell, A., Longcope, C., & McKinlay, J. B. (2005). Age, hormones, and cognitive functioning among middle-aged and elderly men: Cross-sectional evidence from the Massachusetts Male Aging Study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *60*(3), 385–390. https://doi.org/10.1093/gerona/60.3.385

Forrester, S. N., Leoutsakos, J.-M., Gallo, J. J., Thorpe, R. J., & Seeman, T. E. (2019). Association between allostatic load and health behaviours: A latent class approach. *Journal of Epidemiology and Community Health*, *73*(4), 340–345. https://doi.org/10.1136/jech-2018-211289

Franz, C. E., O’Brien, R. C., Hauger, R. L., Mendoza, S. P., Panizzon, M. S., Prom-Wormley, E., … Kremen, W. S. (2011). Cross-sectional and 35-year longitudinal assessment of salivary cortisol and cognitive functioning: The Vietnam Era Twin Study of Aging. *Psychoneuroendocrinology*, *36*(7), 1040–1052. https://doi.org/10.1016/j.psyneuen.2011.01.002

Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R.

A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology*, *52*(1), 1–10. https://doi.org/10.1007/s00127-016-1319-z

Fried, L. P., Xue, Q.-L., Cappola, A. R., Ferrucci, L., Chaves, P., Varadhan, R., ... Bandeen-Roche, K. (2009). Nonlinear multisystem physiological dysregulation associated with frailty in older women: Implications for etiology and treatment. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *64A*(10), 1049–1057. https://doi.org/10.1093/gerona/glp076

Friedman, A. B. (2014). *taRifx: Collection of utility and convenience functions.* Retrieved from https://CRAN.R-project.org/package=taRifx

Gale, C. R., Batty, G. D., Cooper, S.-A., Deary, I. J., Der, G., McEwen, B. S., & Cavanagh, J. (2015). Reaction time in adolescence, cumulative allostatic load, and symptoms of anxiety and depression in adulthood. *Psychosomatic Medicine*, *77*(5), 493–505. https://doi.org/10.1097/psy.0000000000000189

Galea, S., & Tracy, M. (2007). Participation rates in epidemiologic studies. *Annals of Epidemiology*, *17*(9), 643–653. https://doi.org/10.1016/j.annepidem.2007.03.013

Ganguli, M., Du, Y., Dodge, H. H., Ratcliff, G. G., & Chang, C.-C. H. (2006). Depressive symptoms and cognitive decline in late life. *Archives of General Psychiatry*, *63*(2), 153. https://doi.org/10.1001/archpsyc.63.2.153

Garcia-Banda, G., Chellew, K., Fornes, J., Perez, G., Servera, M., & Evans, P. (2014). Neuroticism and cortisol: Pinning down an expected effect. *International Journal of Psychophysiology*, *91*(2), 132–138. https://doi.org/10.1016/j.ijpsycho.2013.12.005

Gassen, N. C., Chrousos, G. P., Binder, E. B., & Zannas, A. S. (2017). Life stress, glucocorticoid signaling, and the aging epigenome: Implications for aging-related diseases. *Neuroscience & Biobehavioral Reviews*, 74, 356–365. https://doi.org/10.1016/j.neubiorev.2016.06.003

Geda, Y. E., Knopman, D. S., Mrazek, D. A., Jicha, G. A., Smith, G. E., Negash, S., … Rocca, W. A. (2006). Depression, apolipoprotein E genotype, and the incidence of mild cognitive impairment: A prospective cohort study. *Archives of Neurology*, *63*(3), 435–440. https://doi.org/10.1001/archneur.63.3.435

Geerlings, M. I., & Gerritsen, L. (2017). Late-life depression, hippocampal volumes, and hypothalamic-pituitary-adrenal axis regulation: A systematic review and meta-analysis. *Biological Psychiatry*, *82*(5), 339–350. https://doi.org/10.1016/j.biopsych.2016.12.032

Geoffroy, M.-C., Hertzman, C., Li, L., & Power, C. (2012). Morning salivary cortisol and cognitive function in mid-life: Evidence from a population-based birth cohort. *Psychological Medicine*, *42*(08), 1763–1773. https://doi.org/10.1017/s0033291711002704

Geronimus, A. T., Hicken, M., Keene, D., & Bound, J. (2006). "Weathering" and age patterns of allostatic load scores among Blacks and Whites in the United States. *American Journal of Public Health*, *96*(5), 826–833. https://doi.org/10.2105/ajph.2004.060749

Gerritsen, L., Comijs, H. C., Deeg, D. J., Penninx, B. W., & Geerlings, M. I. (2011). Salivary cortisol, *APOE*-e4 allele and cognitive decline in a prospective study of older persons. *Neurobiology of Aging*, *32*(9), 1615–1625. Retrieved from http://dx.doi.org/10.1016/j.neurobiolaging.2009.09.007

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57–68. https://doi.org/10.1016/j.intell.2016.01.006

Ginty, A. T., Phillips, A. C., Der, G., Deary, I. J., & Carroll, D. (2011).

Cognitive ability and simple reaction time predict cardiac reactivity in the West of Scotland Twenty-07 Study. *Psychophysiology*, *48*(7), 1022–1027. https://doi.org/10.1111/j.1469-8986.2010.01164.x

Ginty, A. T., Phillips, A. C., Roseboom, T. J., Carroll, D., & deRooij, S. R. (2012). Cardiovascular and cortisol reactions to acute psychological stress and cognitive ability in the Dutch Famine Birth Cohort Study. *Psychophysiology*, *49*(3), 391–400. https://doi.org/10.1111/j.1469-8986.2011.01316.x

Glynn, R. J., Beckett, L. A., Hebert, L. E., Morris, M. C., Scherr, P. A., & Evans, D. A. (1999). Current and remote blood pressure and cognitive decline. *JAMA*, *281*(5), 438. https://doi.org/10.1001/jama.281.5.438

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Goldman, N., Turra, C. M., Glei, D. A., Lin, Y.-H., & Weinstein, M. (2006). Physiological dysregulation and changes in health in an older population. *Experimental Gerontology*, *41*(9), 862–870. https://doi.org/10.1016/j.exger.2006.06.050

Goldman, N., Turra, C. M., Glei, D. A., Seplaki, C. L., Lin, Y.-H., & Weinstein, M. (2006). Predicting mortality from clinical and nonclinical biomarkers. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *61*(10), 1070–1074. https://doi.org/10.1093/gerona/61.10.1070

Gotlib, I. H., LeMoult, J., Colich, N. L., Foland-Ross, L. C., Hallmayer, J., Joormann, J., ... Wolkowitz, O. M. (2015). Telomere length and cortisol reactivity in children of depressed mothers. *Molecular Psychiatry*, *20*(5), 615–620. https://doi.org/10.1038/mp.2014.119

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*(1), 1–4.

https://doi.org/10.1111/j.0963-7214.2004.01301001.x

Gow, A. J., Johnson, W., Pattie, A., Brett, C. E., Roberts, B., Starr, J. M., & Deary, I. J. (2011). Stability and change in intelligence from age 11 to ages 70, 79, and 87: The Lothian Birth Cohorts of 1921 and 1936. *Psychology and Aging*, *26*(1), 232–240. https://doi.org/10.1037/a0021072

Hackett, R. A., Steptoe, A., & Kumari, M. (2014). Association of diurnal patterns in salivary cortisol with Type 2 Diabetes in the Whitehall II study. *The Journal of Clinical Endocrinology & Metabolism*, *99*(12), 4625–4631. https://doi.org/10.1210/jc.2014-2459

Hallquist, M., & Wiley, J. (2014). *MplusAutomation: Automating Mplus model estimation and interpretation*. Retrieved from https://CRAN.R-project.org/package=MplusAutomation

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., … Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, *49*(2), 359–367. https://doi.org/10.1016/j.molcel.2012.10.016

Harrell Jr, F. E., Charles Dupont, & others. (2016). *Hmisc: Harrell Miscellaneous*. Retrieved from https://CRAN.R-project.org/package=Hmisc

Harrington, K. D., Lim, Y. Y., Gould, E., & Maruff, P. (2015). Amyloid-beta and depression in healthy older adults: A systematic review. *Australian & New Zealand Journal of Psychiatry*, *49*(1), 36–46. https://doi.org/10.1177/0004867414557161

Harris, M. A., Brett, C. E., Deary, I. J., & Starr, J. M. (2016). Associations among height, body mass index and intelligence from age 11 to age 78 years. *BMC Geriatrics*, *16*(1). https://doi.org/10.1186/s12877-016-0340-0

Harris, M. A., Cox, S. R., Brett, C. E., Deary, I. J., & MacLullich, A. M.

(2017). Cognitive ability across the life course and cortisol levels in older age. *Neurobiology of Aging*, *59*, 64–71. https://doi.org/10.1016/j.neurobiolaging.2017.07.012

Hertzog, C., Lindenberger, U., Ghisletta, P., & Oertzen, T. von. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods*, *11*(3), 244–252. https://doi.org/10.1037/1082-989x.11.3.244

Hidalgo, V., Almela, M., Pulopulos, M. M., & Salvador, A. (2016). Memory performance is related to the cortisol awakening response in older people, but not to the diurnal cortisol slope. *Psychoneuroendocrinology*, *71*, 136–146. https://doi.org/f8zq9n

Holdnack, J. A. (2001). *WTAR: Wechsler Test of Adult Reading manual.* San Antonio, TX, USA: Psychological Corporation.

Hollingsworth, J. W., Hashizume, A., & Jablon, S. (1965). Correlations between tests of aging in Hiroshima subjects–an attempt to define "physiologic age". *Yale Journal of Biology and Medicine*, *38*(1), 11–26. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2591119/

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115

Horvath, S. (n.d.). *TUTORIAL for the online age calculator: Estimate DNA methylation age.* TUTORIALonlineCalculator file available at https://dnamage.genetics.ucla.edu/.

Hough, C., Morford, A., Epel, E., Lindqvist, D., Bersani, F. S., Jain, F., … Reus, V. (2017). 1005. Pre-treatment allostatic load and metabolic dysregulation predict antidepressant response in major depressive disorder. *Biological Psychiatry*, *81*(10), S406–S407. https://doi.org/10.1016/j.biopsych.2017.02.732

Howard, J. T., & Sparks, P. J. (2016). Does allostatic load calculation method matter? Evaluation of different methods and individual biomarkers functioning by race/ethnicity and educational level. *American Journal of Human Biology*, *28*(5), 627–635. https://doi.org/10.1002/ajhb.22843

Hwang, A.-C., Peng, L.-N., Wen, Y.-W., Tsai, Y.-W., Chang, L.-C., Chiou, S.-T., & Chen, L.-K. (2014). Predicting all-cause and cause-specific mortality by static and dynamic measurements of allostatic load: A 10-year population-based cohort study in Taiwan. *Journal of the American Medical Directors Association*, *15*(7), 490–496. https://doi.org/10.1016/j.jamda.2014.02.001

Ilies, R., Dimotakis, N., & Watson, D. (2010). Mood, blood pressure, and heart rate at work: An experience-sampling study. *Journal of Occupational Health Psychology*, *15*(2), 120–130. https://doi.org/10.1037/a0018350

Insel, K. C., Palmer, R. F., Stroup-Benham, C. A., Markides, K. S., & Espino, D. V. (2005). Association between change in systolic blood pressure and cognitive decline among elderly Mexican Americans: Data from the Hispanic Established Population for Epidemiology Study of the Elderly. *Experimental Aging Research*, *31*(1), 35–54. https://doi.org/10.1080/03610730590882837

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5), 640–648. https://doi.org/10.1097/ede.0b013e31818131e7

Irie, F., Fitzpatrick, A. L., Lopez, O. L., Kuller, L. H., Peila, R., Newman, A. B., & Launer, L. J. (2008). Enhanced risk for Alzheimer disease in persons with type 2 diabetes and APOE e4. *Arch Neurol*, *65*(1). https://doi.org/10.1001/archneurol.2007.29

Jee, H., & Park, J. (2017). Selection of an optimal set of biomarkers and comparative analyses of biological age estimation models in Korean females. *Archives of Gerontology and Geriatrics*, *70*, 84–91. https://doi.org/10.1016/j.archger.2017.01.005

Jia, L., Zhang, W., & Chen, X. (2017). Common methods of biological age estimation. *Clinical Interventions in Aging*, *Volume 12*, 759–772. https://doi.org/10.2147/cia.s134921

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*(1), 95–107. https://doi.org/10.1016/s0160-2896(03)00062-x

Johnson, W., Deary, I. J., & Bouchard, T. J. (2017). Have standard formulas correcting correlations for range restriction been adequately tested?: Minor sampling distribution quirks distort them. *Educational and Psychological Measurement*, 001316441773609. https://doi.org/10.1177/0013164417736092

Juster, R.-P., Marin, M.-F., Sindi, S., Nair, N. P. V., Ng, Y. K., Pruessner, J. C., & Lupien, S. J. (2011). Allostatic load associations to acute, 3-year and 6-year prospective depressive symptoms in healthy older adults. *Physiology & Behavior*, *104*(2), 360–364. https://doi.org/10.1016/j.physbeh.2011.02.027

Juster, R.-P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews*, *35*(1), 2–16. https://doi.org/10.1016/j.neubiorev.2009.10.002

Jylhävä, J., Pedersen, N. L., & Hägg, S. (2017). Biological age predictors. *EBioMedicine*, *21*, 29–36. https://doi.org/10.1016/j.ebiom.2017.03.046

Kantarci, K., Lowe, V., Przybelski, S. A., Weigand, S. D., Senjem, M. L., Ivnik, R. J., ... Jack, C. R. (2012). APOE modifies the association between aβ load and cognition in cognitively normal older adults. *Neurology*, *78*(4), 232–240. https://doi.org/10.1212/wnl.0b013e31824365ab

Karlamangla, A. S., Miller-Martinez, D., Lachman, M. E., Tun, P. A., Koretz, B. K., & Seeman, T. E. (2014). Biological correlates of adult cognition: Midlife in the United States (MIDUS). *Neurobiology of Aging*, *35*(2), 387–394. https://doi.org/10.1016/j.neurobiolaging.2013.07.028

Karlamangla, A. S., Singer, B. H., McEwen, B. S., Rowe, J. W., & Seeman, T. E. (2002). Allostatic load as a predictor of functional decline. *Journal of Clinical Epidemiology*, *55*(7), 696–710. https://doi.org/10.1016/s0895-4356(02)00399-2

Karlsson, I. K., Bennet, A. M., Ploner, A., Andersson, T. M.-L., Reynolds, C. A., Gatz, M., & Pedersen, N. L. (2015). Apolipoprotein E e4 genotype and the temporal relationship between depression and dementia. *Neurobiology of Aging*, *36*(4), 1751–1756. https://doi.org/10.1016/j.neurobiolaging.2015.01.008

Kase, C. S., Wolf, P. A., Kelly-Hayes, M., Kannel, W. B., Beiser, A., & D'Agostino, R. B. (1998). Intellectual decline after stroke: The Framingham Study. *Stroke*, *29*(4), 805–812. https://doi.org/10.1161/01.str.29.4.805

Keller, M. C. (2014). Gene-by-environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. *Biological Psychiatry*, *75*(1), 18–24. Retrieved from http://dx.doi.org/10.1016/j.biopsych.2013.09.006

Kim, J.-M., Stewart, R., Kim, S.-Y., Kim, S.-W., Bae, K.-Y., Yang, S.-J., ... Yoon, J.-S. (2011). Synergistic associations of depression and apolipoprotein E genotype with incidence of dementia. *International Journal of Geriatic Psychiatry*, *26*(9), 893–898. https://doi.org/10.1002/gps.2621

Kirchner, H., Shaheen, F., Kalscheuer, H., Schmid, S., Oster, H., & Lehnert, H. (2017). The telomeric complex and metabolic disease. *Genes*, *8*(7), 176. https://doi.org/10.3390/genes8070176

Kirschbaum, C., & Hellhammer, D. H. (1989). Salivary cortisol in psychobiological research: An overview. *Neuropsychobiology*, *22*(3), 150–169. https://doi.org/10.1159/000118611

Klemera, P., & Doubal, S. (2006). A new approach to the concept and computation of biological age. *Mechanisms of Ageing and Development*, *127*(3), 240–248. https://doi.org/10.1016/j.mad.2005.10.004

Knorr, U., Vinberg, M., Kessing, L. V., & Wetterslev, J. (2010). Salivary cortisol in depressed patients versus control persons: A systematic review and meta-analysis. *Psychoneuroendocrinology*, *35*(9), 1275–1286. https://doi.org/10.1016/j.psyneuen.2010.04.001

Kobrosly, R. W., Seplaki, C. L., Cory-Slechta, D. A., Moynihan, J., & van Wijngaarden, E. (2013). Multisystem physiological dysfunction is associated with depressive symptoms in a population-based sample of older adults. *International Journal of Geriatric Psychiatry*, *28*(7), 718–727. https://doi.org/10.1002/gps.3878

Kobrosly, R. W., van Wijngaarden, E., Seplaki, C. L., Cory-Slechta, D. A., & Moynihan, J. (2014). Depressive symptoms are associated with allostatic load among community-dwelling older adults. *Physiology & Behavior*, *123*, 223–230. https://doi.org/10.1016/j.physbeh.2013.10.014

Koenen, K. C., Lyons, M. J., Goldberg, J., Simpson, J., Williams, W. M., Toomey, R., ... Tsuang, M. T. (2003). A high risk twin study of combat-related PTSD comorbidity. *Twin Research*, *6*(3), 218–226. https://doi.org/10.1375/136905203765693870

Konstabel, K., Lönnqvist, J.-E., Leikas, S., Velázquez, R. G., Qin, H., Verkasalo, M., & Walkowitz, G. (2017). Measuring single constructs by single items: Constructing an even shorter version of the “Short five” personality inventory. *PLOS ONE*, *12*(8), e0182714. https://doi.org/10.1371/journal.pone.0182714

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking ”big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, *136*(5), 768–821. https://doi.org/10.1037/a0020327

Köhler, S., van Boxtel, M. P. J., van Os, J., Thomas, A. J., O'Brien, J. T., Jolles, J., ... Allardyce, J. (2010). Depressive symptoms and cognitive decline in community-dwelling older adults. *Journal of the American Geriatrics Society*, *58*(5), 873–879. https://doi.org/10.1111/j.1532-5415.2010.02807.x

Kruis, J., & Maris, G. (2016). Three representations of the ising model. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep34175

Kuningas, M., Rijk, R. H. de, Westendorp, R. G. J., Jolles, J., Slagboom, P. E., & van Heemst, D. (2007). Mental performance in old age dependent on cortisol and genetic variance in the mineralocorticoid and glucocorticoid receptors. *Neuropsychopharmacology*, *32*(6), 1295–1301. https://doi.org/10.1038/sj.npp.1301260

Kuznetsova, K. A., Maniega, S. M., Ritchie, S. J., Cox, S. R., Storkey, A. J., Starr, J. M., ... Bastin, M. E. (2016). Brain white matter structure and information processing speed in healthy older age. *Brain Structure and Function*, *221*(6), 3223–3235. https://doi.org/10.1007/s00429-015-1097-5

Laceulle, O. M., Nederhof, E., van Aken, M. A. G., & Ormel, J. (2015). Adolescent personality: Associations with basal, awakening, and stress-induced cortisol responses. *Journal of Personality*, *83*(3), 262–273. https://doi.org/10.1111/jopy.12101

Lam, L. L., Emberly, E., Fraser, H. B., Neumann, S. M., Chen, E., Miller, G. E., & Kobor, M. S. (2012). Factors underlying variable DNA methylation in a human community cohort. *Proceedings of the National Academy of Sciences*, *109*(Supplement_2), 17253–17260. https://doi.org/10.1073/pnas.1121249109

Lamers, F., Vogelzangs, N., Merikangas, K. R., Jonge, P. de, Beekman, A. T. F., & Penninx, B. W. J. H. (2013). Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. *Molecular Psychiatry*, *18*(6), 692–699. https://doi.org/10.1038/mp.2012.144

Law, R., Evans, P., Thorn, L., Hucklebridge, F., & Clow, A. (2015). The cortisol awakening response predicts same morning executive function: Results from a 50-day case study. *Psychoneuroendocrinology*, *61*, 68. https://doi.org/10.1016/j.psyneuen.2015.07.576

Lee, B. K., Glass, T. A., McAtee, M. J., Wand, G. S., Bandeen-Roche, K., Bolla, K. I., & Schwartz, B. S. (2007). Associations of salivary cortisol with cognitive function in the Baltimore Memory Study. *Arch Gen Psychiatry*, *64*(7), 810–818. https://doi.org/10.1001/archpsyc.64.7.810

Lee, B. K., Glass, T. A., Wand, G. S., McAtee, M. J., Bandeen-Roche, K., Bolla, K. I., & Schwartz, B. S. (2008). Apolipoprotein E genotype, cortisol, and cognitive function in community-dwelling older adults. *American Journal of Psychiatry*, *165*(11), 1456–1464. https://doi.org/10.1176/appi.ajp.2008.07091532

Levine, M., & Crimmins, E. (2014). Evidence of accelerated aging among African Americans and its implications for mortality. *Social Science & Medicine*, *118*, 27–32. https://doi.org/10.1016/j.socscimed.2014.07.022

Levine, M. E. (2013). Modeling the rate of senescence: Can estimated biological age predict mortality more accurately than chronological age? *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *68*(6), 667–674. https://doi.org/10.1093/gerona/gls233

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, J., & Fischer, J. S. (2004). *Neuropsychological assessment* (5th ed.). Oxford, UK: Oxford University Press.

Li, G., Cherrier, M. M., Tsuang, D. W., Petrie, E. C., Colasurdo, E. A., Craft, S., ... Wilkinson, C. W. (2006). Salivary cortisol and memory function in human aging. *Neurobiology of Aging*, *27*(11), 1705–1714. https://doi.org/10.1016/j.neurobiolaging.2005.09.031

Lim, Y. Y., Ellis, K. A., Ames, D., Darby, D., Harrington, K., Martins, R. N., ... Maruff, P. (2013). Aβ amyloid, cognition, and APOE genotype in healthy older adults. *Alzheimer's & Dementia*, *9*(5), 538–545. https://doi.org/10.1016/j.jalz.2012.07.004

Lopresti, A. L., Maker, G. L., Hood, S. D., & Drummond, P. D. (2014). A review of peripheral biomarkers in major depression: The potential of inflammatory and oxidative stress biomarkers. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *48*, 102–111. https://doi.org/10.1016/j.pnpbp.2013.09.017

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, *153*(6), 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039

Luchetti, M., Terracciano, A., Stephan, Y., & Sutin, A. R. (2016). Personality and cognitive decline in older adults: Data from a longitudinal sample and meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *71*(4), 591–601. https://doi.org/10.1093/geronb/gbu184

Luciano, M., Pujals, A. M. F., Marioni, R. E., Campbell, A., Hayward, C., MacIntyre, D. J., ... Deary, I. J. (2015). Current versus lifetime depression, *APOE* variation, and their interaction on cognitive performance in younger and older adults. *Psychosomatic Medicine*, *77*(5), 480–492. Retrieved from http://dx.doi.org/10.1097/PSY.0000000000000190

Lupien, S. J., Gillin, C. J., & Hauger, R. L. (1999). Working memory is more sensitive than declarative memory to the acute effects of corticosteroids: A doseresponse study in humans. *Behavioral Neuroscience*, *113*(3), 420–430. https://doi.org/10.1037/0735-7044.113.3.420

Luppa, M., Sikorski, C., Luck, T., Ehreke, L., Konnopka, A., Wiese, B., ... Riedel-Heller, S. (2012). Age- and gender-specific prevalence of depression in latest-life systematic review and meta-analysis. *Journal of Affective Disorders*, *136*(3), 212–221. https://doi.org/10.1016/j.jad.2010.11.033

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40. https://doi.org/10.1037/1082-989x.7.1.19

MacLullich, A. M., Deary, I. J., Starr, J. M., Ferguson, K. J., Wardlaw, J. M., & Seckl, J. R. (2005). Plasma cortisol levels, brain volumes and cognition in healthy elderly men. *Psychoneuroendocrinology*, *30*(5), 505–515. https://doi.org/10.1016/j.psyneuen.2004.12.005

Marazziti, D., Rutigliano, G., Baroni, S., Landi, P., & Dell'Osso, L. (2014). Metabolic syndrome and major depression. *CNS Spectrums*, *19*(04), 293–304. https://doi.org/10.1017/s1092852913000667

Marcus, M., Yasamy, M. T., van Ommeren, M., Chisholm, D., & Saxena, S. (2012). Depression: A global public health concern. Retrieved from http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf

Marioni, R. E., Harris, S. E., Shah, S., McRae, A. F., Zglinicki, T. von, Martin-Ruiz, C., ... Deary, I. J. (2016). The epigenetic clock and telomere length are independently associated with chronological age and mortality. *International Journal of Epidemiology*, *45*(2), 424–432. https://doi.org/10.1093/ije/dyw041

Marioni, R. E., Shah, S., McRae, A. F., Chen, B. H., Colicino, E., Harris, S. E., ... Deary, I. J. (2015). DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biology*, *16*(1), 25. https://doi.org/10.1186/s13059-015-0584-6

Martin-Ruiz, C., Saretzki, G., Petrie, J., Ladhoff, J., Jeyapalan, J., Wei, W., … Zglinicki, T. von. (2004). Stochastic variation in telomere shortening rate causes heterogeneity of human fibroblast replicative life span. *Journal of Biological Chemistry*, *279*(17), 17826–17833. https://doi.org/10.1074/jbc.m311980200

Mather, K. A., Jorm, A. F., Parslow, R. A., & Christensen, H. (2011). Is telomere length a biomarker of aging? A review. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *66A*(2), 202–213. https://doi.org/10.1093/gerona/glq180

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, *113*(1), 181–190. https://doi.org/10.1037/0033-2909.113.1.181

McCormick, C., Lewis, E., Somley, B., & Kahan, T. (2007). Individual differences in cortisol levels and performance on a test of executive function in men and women. *Physiology & Behavior*, *91*(1), 87–94. https://doi.org/10.1016/j.physbeh.2007.01.020

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine*, *338*(3), 171–179. https://doi.org/10.1056/nejm199801153380307

McEwen, B. S. (2005). Stressed or stressed out: What is the difference? *Journal of Psychiatry & Neuroscience*, *30*(5), 315–318. Retrieved from http://jpn.ca/wp-content/uploads/2014/04/30-5-315.pdf

McEwen, B. S. (2006). Protective and damaging effects of stress mediators: Central role of the brain. *Dialogues in Clinical Neuroscience*, *8*(4), 367–381. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181832/

McEwen, B. S., & Gianaros, P. J. (2011). Stress- and allostasis-induced brain plasticity. *Annual Review of Medicine*, *62*(1), 431–445. https://doi.org/10.1146/annurev-med-052209-100430

McEwen, B. S., & Rasgon, N. L. (2018). The brain and body on stress. In J. J. Strain & M. Blumenfield (Eds.), *Depression as a systemic illness*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/med/9780190603342.003.0002

McEwen, B. S., & Seeman, T. (1999). Protective and damaging effects of mediators of stress: Elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences*, *896*(1), 30–47. https://doi.org/10.1111/j.1749-6632.1999.tb08103.x

McEwen, B. S., & Wingfield, J. C. (2007). Allostasis and allostatic load. In G. Fink (Ed.), *Encyclopedia of stress* (2nd ed., Vol. 1, pp. 135–141). New York, NY, USA: Academic Press.

McNally, R. J. (2016). Can network analysis transform psychopathology? *Behaviour Research and Therapy*, *86*, 95–104. https://doi.org/10.1016/j.brat.2016.06.006

Meier-Ewert, H. K., Ridker, P. M., Rifai, N., Price, N., Dinges, D. F., & Mullington, J. M. (2001). Absence of diurnal variation of C-reactive protein concentrations in healthy human subjects. *Clinical Chemistry*, *47*(3), 426–430. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11238292

Metti, A. L., Yaffe, K., Boudreau, R. M., Simonsick, E. M., Carnahan, R. M., Satterfield, S., ... Cauley, J. A. (2014). Trajectories of inflammatory markers and cognitive decline over 10 years. *Neurobiology of Aging*, *35*(12), 2785–2790. https://doi.org/10.1016/j.neurobiolaging.2014.05.030

Millard, L. A. C., Davies, N. M., Timpson, N. J., Tilling, K., Flach, P. A., & Smith, G. D. (2015). MR-PheWAS: Hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. *Scientific Reports*, *5*(1). https://doi.org/10.1038/srep16645

Miller, R., & Plessow, F. (2013). Transformation techniques for cross-sectional and longitudinal endocrine data: Application to salivary cortisol concentrations. *Psychoneuroendocrinology*, *38*(6), 941–946. https://doi.org/10.1016/j.psyneuen.2012.09.013

Mitchell, A. J. (2013). The mini-mental state examination (MMSE): An update on its diagnostic validity for cognitive disorders. In *Cognitive screening instruments* (pp. 15–46). Springer London. https://doi.org/10.1007/978-1-4471-2452-8_2

Mooijaart, S. P., Sattar, N., Trompet, S., Lucke, J., Stott, D. J., Ford, I., ... Craen and, A. J. M. de. (2013). Circulating interleukin-6 concentration and cognitive decline in old age: The PROSPER study. *Journal of Internal Medicine*, *274*(1), 77–85. https://doi.org/10.1111/joim.12052

Morgan, G., Hodge, K., Wells, K., & Watkins, M. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via monte carlo simulations. *Journal of Intelligence*, *3*(1), 2–20. https://doi.org/10.3390/jintelligence3010002

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407–422. https://doi.org/10.1016/j.intell.2013.06.004

Murri, M. B., Pariante, C., Mondelli, V., Masotti, M., Atti, A. R., Mellacqua, Z., ... Amore, M. (2014). HPA axis and aging in depression: Systematic review and meta-analysis. *Psychoneuroendocrinology*, *41*, 46–62. https://doi.org/10.1016/j.psyneuen.2013.12.004

Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR$*$D antidepressant trial. *Psychological Methods*, *16*(1), 17–33. https://doi.org/10.1037/a0022634

Muthén, L. K., & Muthén, B. O. (n.d.). *Mplus user's guide. Seventh edition.*

Los Angeles, CA: Muthén & Muthén.

Müezzinler, A., Zaineddin, A. K., & Brenner, H. (2013). A systematic review of leukocyte telomere length and age in adults. *Ageing Research Reviews*, *12*(2), 509–519. https://doi.org/10.1016/j.arr.2013.01.003

Nater, U. M., Hoppmann, C., & Klumb, P. L. (2010). Neuroticism and conscientiousness are associated with cortisol diurnal profiles in adults—role of positive and negative affect. *Psychoneuroendocrinology*, *35*(10), 1573–1577. https://doi.org/10.1016/j.psyneuen.2010.02.017

National Health Service. (2016). Clinical depression - symptoms. Retrieved from https://www.nhs.uk/conditions/clinical-depression/symptoms/

National Institute for Health and Care Excellence. (2017). Type 2 diabetes: Prevention in people at high risk. https://www.nice.org.uk/guidance/ph38/chapter/Glossary.

Nelson, H. E., & Willison, J. (1991). *National Adult Reading Test (NART)*. Windsor: Nfer-Nelson.

Newell-Price, J., Bertagna, X., Grossman, A. B., & Nieman, L. K. (2006). Cushing's syndrome. *The Lancet*, *367*(9522), 1605–1617. https://doi.org/10.1016/s0140-6736(06)68699-6

Nicolson, N. A. (2008). Measurement of cortisol. In L. J. Luecken & L. C. Gallo (Eds.), *Handbook of physiological research methods in health psychology* (pp. 37–74). SAGE Publications. https://doi.org/10.4135/9781412976244.n3

Niti, M., Yap, K.-B., Kua, E.-H., & Ng, T.-P. (2009). APOE-e4, depressive symptoms, and cognitive decline in Chinese older adults: Singapore Longitudinal Aging Studies. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *64A*(2), 306–311. https://doi.org/10.1093/gerona/gln013

Norton, S., Cosco, T., Doyle, F., Done, J., & Sacker, A. (2013).

The Hospital Anxiety and Depression Scale: A meta confirmatory factor analysis. *Journal of Psychosomatic Research*, *74*(1), 74–81. https://doi.org/10.1016/j.jpsychores.2012.10.010

Ottino-González, J., Jurado, M. A., Garcı⊠a-Garcı⊠a, I., Segura, B., Marqués-Iturria, I., Sender-Palacios, M. J., ... Garolera, M. (2018). Allostatic load and disordered white matter microstructure in overweight adults. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-34219-8

Pacella, M. L., Hruska, B., & Delahanty, D. L. (2013). The physical health consequences of PTSD and PTSD symptoms: A meta-analytic review. *Journal of Anxiety Disorders*, *27*(1), 33–46. https://doi.org/10.1016/j.janxdis.2012.08.004

Passos, J. F., Saretzki, G., & von Zglinicki, T. (2007). DNA damage in telomeres and mitochondria during cellular senescence: Is there a connection? *Nucleic Acids Research*, *35*(22), 7505–7513. https://doi.org/10.1093/nar/gkm893

Paterniti, S., Verdier-Taillefer, M.-H., Dufouil, C., & Alpérovitch, A. (2002). Depressive symptoms and cognitive decline in elderly people: Longitudinal study. *The British Journal of Psychiatry*, *181*(5), 406–410. https://doi.org/10.1192/bjp.181.5.406

Pena, E. A., & Slate, E. H. (2014). *gvlma: Global validation of linear models assumptions*. Retrieved from https://CRAN.R-project.org/package=gvlma

Pendlebury, S. T., & Rothwell, P. M. (2009). Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: A systematic review and meta-analysis. *The Lancet Neurology*, *8*(11), 1006–1018. https://doi.org/10.1016/s1474-4422(09)70236-4

Penninx, B. W. J. H. (2017). Depression and its somatic consequences: Allostatic load as the connecting link. *European Psychiatry*, *41*, S62.

https://doi.org/10.1016/j.eurpsy.2017.01.055

Picard, M., Juster, R.-P., & McEwen, B. S. (2014). Mitochondrial allostatic load puts the 'gluc' back in glucocorticoids. *Nature Reviews Endocrinology*, *10*(5), 303–310. https://doi.org/10.1038/nrendo.2014.22

Picard, M., & McEwen, B. S. (2018). Psychological stress and mitochondria. *Psychosomatic Medicine*, *80*(2), 126–140. https://doi.org/10.1097/psy.0000000000000544

Picard, M., Prather, A. A., Puterman, E., Cuillerier, A., Coccia, M., Aschbacher, K., ... Epel, E. S. (2018). A mitochondrial health index sensitive to mood and caregiving stress. *Biological Psychiatry*, *84*(1), 9–17. https://doi.org/10.1016/j.biopsych.2018.01.012

Picard, M., Zhang, J., Hancock, S., Derbeneva, O., Golhar, R., Golik, P., ... Wallace, D. C. (2014). Progressive increase in mtDNA 3243A>G heteroplasmy causes abrupt transcriptional reprogramming. *Proceedings of the National Academy of Sciences*, *111*(38), E4033–E4042. https://doi.org/10.1073/pnas.1414028111

Pickering, T. G. (1990). The clinical significance of diurnal blood pressure variations. Dippers and nondippers. *Circulation*, *81*(2), 700–702. https://doi.org/10.1161/01.cir.81.2.700

Pietrzak, R. H., Laws, S. M., Lim, Y. Y., Bender, S. J., Porter, T., Doecke, J., ... Maruff, P. (2016). Plasma cortisol, brain amyloid-β, and cognitive decline in preclinical Alzheimer's disease: A 6-year prospective cohort study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2016.08.006

Portella, M. J., Harmer, C. J., Flint, J., Cowen, P., & Goodwin, G. M. (2005). Enhanced early morning salivary cortisol in neuroticism. *American Journal of Psychiatry*, *162*(4), 807–809. https://doi.org/10.1176/appi.ajp.162.4.807

Porter, R. J., Barnett, N. A., Idey, A., McGuckin, E. A., & O'Brien, J. T.

(2002). Effects of hydrocortisone administration on cognitive function in the elderly. *Journal of Psychopharmacology*, *16*(1), 65–71. https://doi.org/10.1177/026988110201600106

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*, *28*(7), 916–931. https://doi.org/10.1016/s0306-4530(02)00108-7

Qin, S., Hermans, E. J., van Marle, H. J. F., Luo, J., & Fernández, G. (2009). Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. *Biological Psychiatry*, *66*(1), 25–32. https://doi.org/10.1016/j.biopsych.2009.03.006

Qiu, W. Q., Zhu, H., Dean, M., Liu, Z., Vu, L., Fan, G., … Au, R. (2015). Amyloid-associated depression and ApoE4 allele: Longitudinal follow-up for the development of Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *31*(3), 316–322. https://doi.org/10.1002/gps.4339

Quervain, D. J.-F. de, Roozendaal, B., Nitsch, R. M., McGaugh, J. L., & Hock, C. (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nature Neuroscience*, *3*(4), 313–314. https://doi.org/10.1038/73873

Rajan, K. B., Wilson, R. S., Skarupski, K. A., Leon, C. F. M. de, & Evans, D. A. (2014). Gene-behavior interaction of depressive symptoms and the apolipoprotein E e4 allele on cognitive decline. *Psychosomatic Medicine*, *76*(2), 101–108. https://doi.org/10.1097/psy.0000000000000029

Rast, P., & Hofer, S. M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: Simulation results based on actual longitudinal studies. *Psychological Methods*, *19*(1), 133–154. https://doi.org/10.1037/a0034524

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3: The Standard Progressive Matrices*. Oxford, UK: Oxford Psychologists Press.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

R Core Team. (2018a). *foreign: Read data stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...* Retrieved from https://CRAN.R-project.org/package=foreign

R Core Team. (2018b). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, *14*(2), 95–101. https://doi.org/10.1111/j.0963-7214.2005.00342.x

Revelle, W. (2016). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from http://CRAN.R-project.org/package=psych

Revelle, W. (2018). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Révész, D., Verhoeven, J. E., Milaneschi, Y., Geus, E. J. de, Wolkowitz, O. M., & Penninx, B. W. (2014). Dysregulated physiological stress systems and accelerated cellular aging. *Neurobiology of Aging*, *35*(6), 1422–1430. https://doi.org/10.1016/j.neurobiolaging.2013.12.027

Rhemtulla, M., Savalei, V., & Little, T. D. (2016). On the asymptotic relative efficiency of planned missingness designs. *Psychometrika*, *81*(1), 60–89. https://doi.org/10.1007/s11336-014-9422-0

Riese, H., Rijsdijk, F. V., Rosmalen, J. G. M., Snieder, H., & Ormel, J. (2009). Neuroticism and morning cortisol secretion: Both heritable, but no shared genetic influences. *Journal of Personality*, *77*(5), 1561–1576. https://doi.org/10.1111/j.1467-6494.2009.00592.x

Riethman, H. (2008). Human telomere structure and biology. *Annual Review of Genomics and Human Genetics*, *9*(1), 1–19. https://doi.org/10.1146/annurev.genom.8.021506.172017

Ritchie, S. J., Tucker-Drob, E. M., Cox, S. R., Dickie, D. A., C. Valdés Hernández, M. del, Corley, J., ... Deary, I. J. (2017). Risk and protective factors for structural brain ageing in the eighth decade of life. *Brain Structure and Function*, *222*(8), 3477–3490. https://doi.org/10.1007/s00429-017-1414-2

Robertson, T., Beveridge, G., & Bromley, C. (2017). Allostatic load as a predictor of all-cause and cause-specific mortality in the general population: Evidence from the Scottish Health Survey. *PLOS ONE*, *12*(8), e0183297. https://doi.org/10.1371/journal.pone.0183297

Rode, L., Nordestgaard, B. G., & Bojesen, S. E. (2015). Peripheral blood leukocyte telomere length and mortality among 64 637 individuals from the general population. *JNCI: Journal of the National Cancer Institute*, *107*(6). https://doi.org/10.1093/jnci/djv074

Rode, L., Nordestgaard, B. G., Weischer, M., & Bojesen, S. E. (2014). Increased body mass index, elevated C-reactive protein, and short telomere length. *The Journal of Clinical Endocrinology & Metabolism*, *99*(9), E1671–E1675. https://doi.org/10.1210/jc.2014-1161

Rosmond, R. (2003). Stress induced disturbances of the HPA axis: A pathway to Type 2 diabetes? *Medical Science Monitor*, *9*(2), RA35–RA39. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12601304

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.

18637/jss.v048.i02

RStudio Team. (2016). *RStudio: Integrated development environment for R*. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Russell, E., Koren, G., Rieder, M., & van Uum, S. (2012). Hair cortisol as a biological marker of chronic stress: Current status, future directions and unanswered questions. *Psychoneuroendocrinology*, *37*(5), 589–601. https://doi.org/10.1016/j.psyneuen.2011.09.009

Sachs-Ericsson, N., Joiner, T., Plant, E. A., & Blazer, D. G. (2005). The influence of depression on cognitive decline in community-dwelling elderly persons. *The American Journal of Geriatric Psychiatry*, *13*(5), 402–408. https://doi.org/10.1097/00019442-200505000-00009

Sahin, E., Colla, S., Liesa, M., Moslehi, J., Müller, F. L., Guo, M., ... DePinho, R. A. (2011). Telomere dysfunction induces metabolic and mitochondrial compromise. *Nature*, *470*(7334), 359–365. https://doi.org/10.1038/nature09787

Salthouse, T. A. (2014). Correlates of cognitive change. *Journal of Experimental Psychology: General*, *143*(3), 1026–1048. https://doi.org/10.1037/a0034847

Sanders, J. L., & Newman, A. B. (2013). Telomere length in epidemiology: A biomarker of aging, age-related disease, both, or neither? *Epidemiologic Reviews*, *35*(1), 112–131. https://doi.org/10.1093/epirev/mxs008

Savransky, A., Chiappelli, J., Rowland, L. M., Wisner, K., Shukla, D. K., Kochunov, P., & Hong, L. E. (2017). Fornix structural connectivity and allostatic load. *Psychosomatic Medicine*, *79*(7), 770–776. https://doi.org/10.1097/psy.0000000000000487

Scharoun-Lee, M., Adair, L. S., Kaufman, J. S., & Gordon-Larsen, P. (2009). Obesity, race/ethnicity and the multiple dimensions of socioeconomic status during the transition to adulthood: A factor analysis approach. *Social Science & Medicine*, *68*(4), 708–716. https://doi.org/10.1016/j.socscimed.2008.12.009

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*(2), 23–74. Retrieved from http://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf

Schmidt, M. (1996). *Rey Auditory Verbal Learning Test (RAVLT): A handbook*. Los Angeles, CA: Western Psychological Services.

Scott, S. B., Graham-Engeland, J. E., Engeland, C. G., Smyth, J. M., Almeida, D. M., Katz, M. J., ... Sliwinski, M. J. (2015). The Effects of Stress on Cognitive Aging, Physiology and Emotion (ESCAPE) project. *BMC Psychiatry*, *15*(1). https://doi.org/10.1186/s12888-015-0497-7

Scottish Council for Research in Education. (1933). *The intelligence of Scottish children: A national survey of an age-group*. London, UK: University of London Press.

Scottish Council for Research in Education. (1949). *The trend of Scottish intelligence: A comparison of the 1947 and 1932 surveys of the intelligence of eleven-year-old pupils*. London, UK: University of London Press.

Seeman, T. E., McEwen, B. S., Rowe, J. W., & Singer, B. H. (2001). Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. *Proceedings of the National Academy of Sciences*, *98*(8), 4770–4775. https://doi.org/10.1073/pnas.081072698

Seeman, T. E., Singer, B. H., Rowe, J. W., Horwitz, R. I., & McEwen, B. S. (1997). Price of adaptation—allostatic load and its health consequences. *Archives of Internal Medicine*, *157*(19), 2259. https://doi.org/10.1001/archinte.1997.00440400111013

Seeman, T., Gruenewald, T., Karlamangla, A., Sidney, S., Liu, K., Mcewen, B., & Schwartz, J. (2010). Modeling multisystem biological risk in young adults: The Coronary Artery Risk Development in Young Adults Study. *American Journal of Human Biology*, *22*(4), 463–472. https://doi.org/10.1002/ajhb.21018

Seeman, T., Merkin, S. S., Crimmins, E., Koretz, B., Charette, S., & Karlamangla, A. (2008). Education, income and ethnic differences in cumulative biological risk profiles in a national sample of US adults: NHANES III (1988–1994). *Social Science & Medicine*, *66*(1), 72–87. https://doi.org/10.1016/j.socscimed.2007.08.027

Seplaki, C. L., Goldman, N., Glei, D., & Weinstein, M. (2005). A comparative analysis of measurement approaches for physiological dysregulation in an older population. *Experimental Gerontology*, *40*(5), 438–449. https://doi.org/10.1016/j.exger.2005.03.002

Seplaki, C. L., Goldman, N., Weinstein, M., & Lin, Y.-H. (2006). Measurement of cumulative physiological dysregulation in an older population. *Demography*, *43*(1), 165–183. https://doi.org/10.1353/dem.2006.0009

Shalev, I., Entringer, S., Wadhwa, P. D., Wolkowitz, O. M., Puterman, E., Lin, J., & Epel, E. S. (2013). Stress and telomere biology: A lifespan perspective. *Psychoneuroendocrinology*, *38*(9), 1835–1842. https://doi.org/10.1016/j.psyneuen.2013.03.010

Shiels, P. G., Stenvinkel, P., Kooman, J. P., & McGuinness, D. (2017). Circulating markers of ageing and allostatic load: A slow train coming. *Practical Laboratory Medicine*, 7, 49–54. https://doi.org/10.1016/j.plabm.2016.04.002

Simons, M. J. (2015). Questioning causal involvement of telomeres in aging. *Ageing Research Reviews*, *24*, 191–196. https://doi.org/10.1016/j.arr.2015.08.002

Singh-Manoux, A., Dugravot, A., Elbaz, A., Shipley, M., Kivimaki, M., & Kumari, M. (2014). No evidence of a longitudinal association between diurnal cortisol patterns and cognition. *Neurobiology of Aging*, *35*(10), 2239–2245. Retrieved from http://dx.doi.org/10.1016/j.neurobiolaging.2014.03.015

Sliwinski, M. J. (2008). Measurement-burst designs for social health research. *Social and Personality Psychology Compass*, *2*(1), 245–261. https://doi.org/10.1111/j.1751-9004.2007.00043.x

Smith, A. (1982). *Symbol Digit Modalities Test: Manual*. Los Angeles, CA: Western Psychological Services.

Sonnenschein, M., Mommersteeg, P. M. C., Houtveen, J. H., Sorbi, M. J., Schaufeli, W. B., & van Doornen, L. J. P. (2007). Exhaustion and endocrine functioning in clinical burnout: An in-depth study using the experience sampling method. *Biological Psychology*, *75*(2), 176–184. https://doi.org/10.1016/j.biopsycho.2007.02.001

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195–231. https://doi.org/10.1037/0033-295x.99.2.195

Stalder, T., Kirschbaum, C., Kudielka, B. M., Adam, E. K., Pruessner, J. C., Wüst, S., ... Clow, A. (2016). Assessment of the cortisol awakening response: Expert consensus guidelines. *Psychoneuroendocrinology*, *63*, 414–432. Retrieved from http://dx.doi.org/10.1016/j.psyneuen.2015.10.010

Starkman, M. N., Giordani, B., Gebarski, S. S., Berent, S., Schork, M., & Schteingart, D. E. (1999). Decrease in cortisol reverses human hippocampal atrophy following treatment of Cushing's disease. *Biological Psychiatry*, *46*(12), 1595–1602. https://doi.org/10.1016/s0006-3223(99)00203-6

Stawski, R. S., Almeida, D. M., Lachman, M. E., Tun, P. A., Rosnick, C. B., & Seeman, T. (2011). Associations between cognitive function and naturally occurring daily cortisol during middle adulthood: Timing is everything. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *66B*(Supplement 1), i71–i81. https://doi.org/10.1093/geronb/gbq094

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2013). Methodological artifacts in dimensionality assessment of the Hospital Anxiety and Depression Scale (HADS). *Journal of Psychosomatic Research*, *74*(2), 116–121. https://doi.org/10.1016/j.jpsychores.2012.11.012

Sun, X., Steffens, D. C., Au, R., Folstein, M., Summergrad, P., Yee, J., … Qiu, W. Q. (2008). Amyloid-associated depression. *Archives of General Psychiatry*, *65*(5), 542. https://doi.org/10.1001/archpsyc.65.5.542

Taylor, A. M., Pattie, A., & Deary, I. J. (2018). Cohort profile update: The Lothian Birth Cohorts of 1921 and 1936. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyy022

Thoresen, M. (2019). Spurious interaction as a result of categorization. *BMC Medical Research Methodology*, *19*(1). https://doi.org/10.1186/s12874-019-0667-2

Thorndike, R. L. (1987). Stability of factor loadings. *Personality and Individual Differences*, *8*(4), 585–586. https://doi.org/10.1016/0191-8869(87)90224-8

Tiemensma, J., Biermasz, N. R., Middelkoop, H. A. M., van der Mast, R. C., Romijn, J. A., & Pereira, A. M. (2010). Increased prevalence of psychopathology and maladaptive personality traits after long-term cure of Cushing's disease. *The Journal of Clinical Endocrinology & Metabolism*, *95*(10), E129–E141. https://doi.org/10.1210/jc.2010-0512

Tiemensma, J., Kokshoorn, N. E., Biermasz, N. R., Keijser, B.-J. S. A., Wassenaar, M. J. E., Middelkoop, H. A. M., … Romijn, J. A. (2010). Subtle cognitive impairments in patients with long-term cure of Cushing's disease. *The Journal of Clinical Endocrinology & Metabolism*, *95*(6), 2699–2714. https://doi.org/10.1210/jc.2009-2032

Tobin, C. (2018). *ggthemr: Themes for ggplot2*. Retrieved from https://github.com/cttobin/ggthemr

Toledo, J. B., Toledo, E., Weiner, M. W., Jack, C. R., Jagust, W., Lee, V. M.-Y., ... Trojanowski, J. Q. (2012). Cardiovascular risk factors, cortisol, and amyloid-β deposition in Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's & Dementia*, *8*(6), 483–489. https://doi.org/10.1016/j.jalz.2011.08.008

Tomas, J. M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 84–98. https://doi.org/10.1080/10705519909540120

Tomiyama, A. J., O'Donovan, A., Lin, J., Puterman, E., Lazaro, A., Chan, J., ... Epel, E. (2012). Does cellular aging relate to patterns of allostasis? *Physiology & Behavior*, *106*(1), 40–45. https://doi.org/10.1016/j.physbeh.2011.11.016

Treadway, M. T., & Leonard, C. V. (2016). Isolating biomarkers for symptomatic states: Considering symptom–substrate chronometry. *Molecular Psychiatry*, *21*(9), 1180–1187. https://doi.org/10.1038/mp.2016.83

Vaessen, T., Kasanova, Z., Hernaus, D., Lataster, J., Collip, D., van Nierop, M., & Myin-Germeys, I. (2018). Overall cortisol, diurnal slope, and stress reactivity in psychosis: An experience sampling approach. *Psychoneuroendocrinology*, *96*, 61–68. https://doi.org/10.1016/j.psyneuen.2018.06.007

Valkanova, V., Ebmeier, K. P., & Allan, C. L. (2013). CRP, IL-6 and depression: A systematic review and meta-analysis of longitudinal studies. *Journal of Affective Disorders*, *150*(3), 736–744. https://doi.org/10.1016/j.jad.2013.06.004

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. Retrieved from https://www.jstatsoft.org/v45/i03/

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295x.113.4.842

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

van Ockenburg, S. L., Booij, S. H., Riese, H., Rosmalen, J. G. M., & Janssens, K. A. M. (2015). How to assess stress biomarkers for idiographic research? *Psychoneuroendocrinology*, *62*, 189–199. https://doi.org/10.1016/j.psyneuen.2015.08.002

Wallace, M., Harville, E., Theall, K., Webber, L., Chen, W., & Berenson, G. (2013). Preconception biomarkers of allostatic load and racial disparities in adverse birth outcomes: The Bogalusa Heart Study. *Paediatric and Perinatal Epidemiology*, *27*(6), 587–597. https://doi.org/10.1111/ppe.12091

Wechsler, D. (1998a). *WAIS-III UK administration and scoring manual.* London, UK: Psychological Corporation.

Wechsler, D. (1998b). *WMS-III UK administration and scoring manual.* London, UK: Psychological Corporation.

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLOS ONE*, *12*(4), e0174944. https://doi.org/10.1371/journal.pone.0174944

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29. Retrieved from http://www.jstatsoft.org/v40/i01/

Wickham, H., Hester, J., & Chang, W. (2018). *devtools: Tools to make developing R packages easier*. Retrieved from https://CRAN.R-project.org/package=devtools

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x

Wiley, J. F., Gruenewald, T. L., Karlamangla, A. S., & Seeman, T. E. (2016). Modeling multisystem physiological dysregulation. *Psychosomatic Medicine*, *78*(3), 290–301. https://doi.org/10.1097/psy.0000000000000288

Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement*, *49*(3), 587–602. https://doi.org/10.1177/001316448904900309

Wilson, R. S., Barnes, L. L., Leon, C. F. M. de, Aggarwal, N. T., Schneider, J. S., Bach, J., … Bennett, D. A. (2002). Depressive symptoms, cognitive decline, and risk of AD in older persons. *Neurology*, *59*(3), 364–370. https://doi.org/10.1212/wnl.59.3.364

Wilson, R. S., Leon, C. F. M. de, Bennett, D. A., Bienias, J. L., & Evans, D. A. (2004). Depressive symptoms and cognitive decline in a community population of older persons. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(1), 126–129. Retrieved from http://europepmc.org/abstract/med/14707321

World Health Organization. (1998). Stress and the nervous system: Summary. Geneva, Switzerland: Author. Retrieved from http://apps.

who.int/iris/bitstream/10665/64430/1/WHO_RPS_98.1.pdf

World Health Organization. (2015). World report on ageing and health. Geneva, Switzerland: Author. Retrieved from http://apps.who.int/iris/bitstream/10665/64430/1/WHO_RPS_98.1.pdf

Xie, Y. (2018a). *bookdown: Authoring books and technical documents with R markdown*. Retrieved from https://github.com/rstudio/bookdown

Xie, Y. (2018b). *knitr: A general-purpose package for dynamic report generation in R*. Retrieved from https://yihui.name/knitr/

Yaffe, K. (2007). Metabolic syndrome and cognitive disorders. *Alzheimer Disease & Associated Disorders*, *21*(2), 167–171. https://doi.org/10.1097/wad.0b013e318065bfd6

Young, A. H., Sahakian, B. J., Robbins, T. W., & Cowen, P. J. (1999). The effects of chronic administration of hydrocortisone on cognitive function in normal male volunteers. *Psychopharmacology*, *145*(3), 260–266. https://doi.org/10.1007/s002130051057

Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *64*(2), 113–128. https://doi.org/10.1007/bf02294531

Zahodne, L. B., Stern, Y., & Manly, J. J. (2014). Depressive symptoms precede memory decline, but not vice versa, in non-demented older adults. *Journal of the American Geriatrics Society*, *62*(1), 130–134. https://doi.org/10.1111/jgs.12600

Zalli, A., Carvalho, L. A., Lin, J., Hamer, M., Erusalimsky, J. D., Blackburn, E. H., & Steptoe, A. (2014). Shorter telomeres with high telomerase activity are associated with raised allostatic load and impoverished psychosocial resources. *Proceedings of the National Academy of Sciences*, *111*(12), 4519–4524. https://doi.org/10.1073/pnas.1322145111

Zannas, A. S., Arloth, J., Carrillo-Roa, T., Iurato, S., Röh, S., Ressler, K. J.,

… Mehta, D. (2015). Lifetime stress accelerates epigenetic aging in an urban, African American cohort: Relevance of glucocorticoid signaling. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0828-5

Zhang, W.-G., Zhu, S.-Y., Bai, X.-J., Zhao, D.-L., Jiang, S.-M., Li, J., … Chen, X.-M. (2014). Select aging biomarkers based on telomere length and chronological age to build a biological age equation. *AGE*, *36*(3). https://doi.org/10.1007/s11357-014-9639-y

Zhu, H. (2018). *kableExtra: Construct complex table with 'kable' and pipe syntax*. Retrieved from https://CRAN.R-project.org/package=kableExtra

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x

Appendix

This appendix contains the published version of the letter in Chapter 2. The reference for this article is:

Crook, Z., & Booth, T. (2017). Considering the appropriateness of the factor analytic operationalization of allostatic load. *Psychosomatic Medicine*, 79(1), 117–119. https://doi.org/cvc5

MacArthur studies of successful aging. Arch Intern Med 1997; 157:2259–68.

7. McEwen BS, Stellar E. Stress and the individual. Mechanisms leading to disease. Arch Intern Med 1993;153:2093–101.
8. Meimaridou E, Hughes CR, Kowalczyk J, Guasti L, Chapple JP, King PJ, Chan LF, Clark AJ, Metherell LA. Familial glucocorticoid deficiency: new genes and mechanisms. Mol Cell Endocrinol 2013;371:195–200.
9. Nicholls DG, Fergusson SJ. Bioenergetics. 4 ed. Cambridge, MA: Academic Press; 2013.
10. Picard M, Turnbull DM. Linking the metabolic state and mitochondrial DNA in chronic disease, health, and aging. Diabetes 2013;62:672–8.
11. Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, Kim SJ, Mehta H, Hevener AL, de Cabo R, Cohen P. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. Cell Metab 2015;21:443–54.
12. Zhang Q, Raoof M, Chen Y, Sumi Y, Sursal T, Junger W, Brohi K, Itagaki K, Hauser CJ. Circulating mitochondrial DAMPs cause inflammatory responses to injury. Nature 2010;464:104–7.
13. Lu B, Kwan K, Levine YA, Olofsson PS, Yang H, Li J, Joshi S, Wang H, Andersson U, Chavan SS, Tracey KJ. Alpha 7 nicotinic acetylcholine receptor signaling inhibits inflammasome activation by preventing mitochondrial DNA release. Mol Med 2014.
14. Picard M, Juster RP, McEwen BS. Mitochondrial allostatic load puts the 'gluc' back in glucocorticoids. Nat Rev Endocrinol 2014;10:303–10.
15. McEwen BS. Protective and damaging effects of stress mediators: central role of the brain. Dialogues Clin Neurosci 2006;8:367–81.
16. Juster RP, Moskowitz DS, Lavoie J, D'Antono B. Sex-specific interaction effects of age, occupational status, and workplace stress on psychiatric symptoms and allostatic load among healthy Montreal workers. Stress 2013;16:616–29.
17. Picard M. Pathways to aging: the mitochondrion at the intersection of biological and psychosocial sciences. J Aging Res 2011;2011:814096.
18. Picard M, McManus MJ, Gray JD, Nasca C, Moffat C, Kopinski PK, Seifert EL, McEwen BS, Wallace DC. Mitochondrial functions modulate neuroendocrine, metabolic, inflammatory, and transcriptional responses to acute psychological stress. Proc Natl Acad Sci U S A 2015;112:E6614–23.

OPEN

Considering the Appropriateness of the Factor Analytic Operationalization of Allostatic Load

In a recent issue of *Psychosomatic Medicine*, Wiley et al. (1) made a valuable contribution to the discussion of the optimal measurement of allostatic load (AL). In the most comprehensive factor analytic investigation of AL to date, they found that a bifactor model with a general AL factor and seven physiological system factors fits better than a higher-order model in which the seven system factors loaded on the general AL factor. Similar models have been applied by the author (T.B.) and others to operationalize AL (2–4). Here, we consider the primary theoretical assumptions underlying latent variable modeling, argue that the construct of AL is inconsistent with these assumptions, and propose alternate operationalizations of AL.

UNDERLYING CONSTRUCT (COMMON CAUSE)

A latent variable model is estimated based on the patterns of covariance in a set of variables. By including an AL general factor in a latent variable model, researchers are positing that an underlying construct is the common cause of the observed covariation in all of the modeled biological measures. Although the theoretical relation of the common cause or construct to the original variables differs in bifactor versus higher-order models, in either case, we must ask: What could this common factor be? Wiley et al. stated that the AL factor "[captures] the notion that there is an underlying process influencing multiple physiological systems" ((1): p. 4). However, the observation of a general factor estimated from inter-individual summary statistics (i.e., covariances) says little about what this process may actually be.

INDEPENDENCE CONDITIONAL ON THE LATENT TRAIT

A primary assumption of latent variable models is that once the effect of the latent factors has been accounted for, the measured variables—in this case, the biological measures—are independent. This is unlikely to be the case with AL measures. Levels of different biomarkers are linked causally to each other, rather than only through the common cause latent variable(s). For example, body mass index (BMI) has previously been used as a metabolic system AL biological measure (e.g., (2,5)). However, Mendelian randomization studies have found that increased BMI has a causal effect on levels of other metabolic biological measures as well as levels of AL biomarkers used to represent other physiological systems, such as blood pressure and inflammation (e.g., (6)). Thus, it is most likely that the biomarkers are not conditionally independent but are instead dynamically related in complex networks. Such networks can produce observed correlations between variables that have no common cause (7).

INTERCHANGEABILITY OF INDICATORS

A further assumption of the latent variable model is that the definition of the latent variable does not change when different sets of indicators are used (8). This holds because the indicators are affected by, but do not affect, the latent variable. Another key finding of Wiley et al. was that fitting models in which the biological measures from each of the 7 physiological systems were excluded caused no large changes in AL factor loadings (1). This method provides only a weak test of interchangeability. The stability of general intelligence factor loadings has long been a research focus for intelligence researchers, so AL researchers may benefit from applying their approaches to this issue (e.g., (9,10)). For example, researchers could compute and correlate AL scores from different nonoverlapping multisystem sets of biological measures (8). The existence of diverse causal links between AL biological measures from different physiological systems suggests to us that the nature of what relates the biomarkers may change depending on which measures are included in the model. We predict that more thorough, more powerful tests of the stability of AL factor loadings will find that it does not hold.

FORMATIVE VERSUS REFLECTIVE INDICATORS

In the common factor model, the biological measures are reflective indicators, that is, they are manifested by a common cause latent variable. However, to the extent that the model assumptions are violated (previously discussed), the factor model is not appropriate. Thus, it may instead be profitable to consider the biological measures as formative indicators, that is, as variables that define the construct (8). This way of thinking about how the biological measures relate to AL is consistent with any number of weighted or sum scores. It is also consistent with AL theory, in that more severe, more widespread physiological dysregulation will relate to higher AL scores.

Alternatively, the associations between AL biological measures could be modeled using each measure individually, without the need for any single latent or observed summary. This could be done with network analysis, which has been used beneficially by researchers studying symptom networks in mental disorders (11). Allostatic load indicators can also be modeled separately without consideration of their associations. Consistent with this approach, *Psychosomatic Medicine* typically provides data of separate biological measures when articles report about complex phenomena such as AL and metabolic syndrome.

Aside from any issues regarding model assumptions, two further points warrant comment about the models presented by Wiley et al.

IMPROVED MODEL FIT FOR BIFACTOR APPROACH

The complex causal links between biological measures from different physiological systems also help to explain why the bifactor AL model fits better than the hierarchical AL model. The hierarchical model imposes "proportionality constraints" ((12): p. 115): the ratio of the AL general factor loadings to the system factor loadings is constrained to equality within the biological measures of each physiological system. Considering the diverse causal links between different AL biomarkers, both within and across systems, these proportionality constraints are likely to be violated. Furthermore, it has been shown that when the true model contains "unmodelled complexity" ((13): p. 407) in the form of small correlated residuals and cross-loadings, or even modeled complexity in the form of correlated residuals across factors, fit indices and criteria may be biased in favor of the bifactor model. Consequently, the better fit of the bifactor model follows from AL theory and research, as well as from methodological findings, for reasons other than those Wiley et al. (1) focused on.

VARIANCE EXPLAINED BY PHYSIOLOGICAL DYSREGULATION FACTORS

Statistically, a desirable property of a general factor is that it accounts for most variance in the constituent indicator variables. In the study by Wiley et al., the AL factor explained only approximately 11% of the variance in the AL biological measures. Some of the physiological system-specific factors were also weak. For example, the hypothalamic-pituitary-adrenal axis and inflammation factors explained only approximately 9% and 16% of the variance in their respective biological measures. Note that weak factor saturation of physiological dysregulation factors has also been an issue in other samples (2,3).

PROPERTIES OF OPTIMAL SCORES FOR AL

Ideally, AL scores should be: 1) calculated using biological measures from various physiological systems; 2) consistently calculated across samples; and (3) closely related to criterion variables. Those who desire scores that are rooted in AL theory would prefer the AL scoring method that produces the scores most closely related to chronic/repeated perceived stress. For a pragmatist, the focus may not be on investigating how different physiological dysregulation scores relate to prior perceived stress but rather on finding the scores that most strongly predict important health outcomes such as cardiovascular disease and death. It may also be advantageous to have scores that explicitly represent the accumulation of the effects of repeated environmental challenges.

Our theoretical and methodological concerns with the factor analytic operationalization of AL suggest to us that factor scores will not prove to be the optimal AL scoring method. We therefore believe that further research is required to determine the optimal operationalization(s) of AL.

ZANDER CROOK, MSc
Department of Psychology
The University of Edinburgh, Edinburgh
United Kingdom

TOM BOOTH, PhD
Department of Psychology
The University of Edinburgh, Edinburgh
United Kingdom

Source of Funding and Conflicts of Interest: ZC is supported by a UK Economic and Social Research Council Advanced Quantitative Methods PhD award. Neither author reports a conflict of interest.

REFERENCES

1. Wiley JF, Gruenewald TL, Karlamangla AS, Seeman TE. Modeling multisystem physiological dysregulation. Psychosom Med 2016;78:290–301.
2. Booth T, Starr JM, Deary I. Modeling multisystem biological risk in later life: allostatic load in the Lothian Birth Cohort study 1936. Am J Hum Biol 2013;25: 538–43.
3. Seeman T, Gruenewald T, Karlamangla A, Sidney S, Liu K, McEwen B, Schwartz J. Modeling multisystem biological risk in young adults: the Coronary Artery Risk Development in Young Adults study. Am J Hum Biol 2010;22: 463–72.
4. Howard JT, Sparks PJ. Does allostatic load calculation method matter? Evaluation of different methods and individual biomarkers functioning by race/ethnicity and educational level. Am J Hum Biol 2016;28:627–35. http://dx.doi.org/10.1002/ajhb.22843.
5. Crimmins EM, Johnston M, Hayward M, Seeman T. Age differences in allostatic load: an index of physiological dysregulation. Exp Gerontol 2003;38:731–4.
6. Millard LA, Davies NM, Timpson NJ, Tilling K, Flach PA, Smith G. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. Sci Rep 2015;5: 16645. http://dx.doi.org/10.1038/srep16645.
7. van der Maas HLJ, Dolan CV, Grasman RPPP, Wicherts JM, Huizenga HM, Raijmakers MEJ. A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. Psychol Rev 2006;113:842–61.
8. Bollen KA, Bauldry S. Three Cs in measurement models: causal indicators, composite indicators, and covariates. Psychol Methods 2011;16:265–84.
9. Johnson W, Bouchard TJ Jr., Krueger RF, McGue M, Gottesman II. Just one *g*: consistent results from three test batteries. Intelligence 2004;32:95–107.
10. Thorndike RL. Stability of factor loadings. Pers Individ Dif 1987;8:585–6.
11. Borsboom D, Cramer AOJ. Network analysis: an integrative approach to the structure of psychopathology. Annu Rev Clin Psychol 2013;9:91–121.
12. Yung Y-F, Thissen D, McLeod LD. On the relationship between the higher-order factor model and the hierarchical factor model. Psychometrika 1999;64:113–28.
13. Murray AL, Johnson W. The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. Intelligence 2013;41: 407–22.

The Authors Reply: Pursuing the Optimal Operationalization of Allostatic Load

In their commentary on our paper (1), Crook and Booth (2) raise important questions regarding the theoretical and methodological appropriateness of applying factor analysis to model allostatic load (AL). They argue that factor scores are not the "optimal" measure of AL and conclude that more research is needed.

From a methodological perspective, Crook and Booth argue that the poor fit of the hierarchical model may be due to the proportionality constraint inherent in its structure. Whether the proportionality constraint or other features of the model are the sources, it does not change the result that the hierarchical model provided an inferior fit to the data than did the bifactor model. Considering the bifactor model, Crook and Booth suggested a stronger test of exchangeability by computing and correlating AL scores from nonoverlapping biomarkers (2). However, unlike intelligence research, studies on biomarkers tend to have relatively few indicators as a consequence of feasibility factors (e.g., available blood sample volume, participant burden, and cost). Separating already limited biomarker panels into two nonoverlapping sets will be too few to estimate the bifactor model in many studies. As advances in multiplexing reduce barriers to assaying numerous biomarkers (3), we look forward to studies that address this question.

Crook and Booth also note that, on average, the general factor explains approximately 11% of the variance in the biomarkers (2). In psychometric studies of carefully designed scales, 11% may be considered a small amount of variance, but this is expected for biomarkers for several reasons. First, except for the heart rate variability measures, each biomarker is distinct (e.g., interleukin 6 and tumor necrosis factor α are separate analytes with unique roles in the immune system and inflammatory processes). Second, biomarkers have circadian rhythms, which introduce variability due to the timing of assessment. Third, the timeframe reflected in each biomarker varies. For example, glycosylated hemoglobin approximately indexes the previous three months, overnight urinary epinephrine and norepinephrine index approximately 12 hours, and blood pressure is comparatively momentary. Fourth, the general factor accounted for variance over and above the covariates age and sex. Therefore, we expected the general factor to account for a modest amount of the variance in biomarkers. Furthermore, considered over and above the effects of age, 11% of the variance is not necessarily trivial. Finally, the overall model (i.e., general + system factors and covariates) accounted for an average of 55% of the variance in biomarkers, rising to 60% when excluding soluble intracellular adhesion molecule 1 and low-density lipoprotein. If researchers believe