Latent variable modelling of personality-health associations: Measures, models and extensions

Gareth Hagger-Johnson, MA, M.Sc.

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

to

The School of Philosophy Psychology and Language Sciences

University of Edinburgh

21st August 2008

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Gareth Hagger-Johnson

Abstract

Functional health status, morbidity and mortality are determined partly by health behaviours (World Health Organization, 2002), which have determinants of their own. Personality traits, such as Conscientiousness, have a strong association with health behaviours (Bogg & Roberts, 2004). There is a less consistent and generally weaker association between traits and health outcomes (e.g. Neuroticism and mortality). The central problem in this thesis is how to measure, model, maximize, and extend trait-health associations. Conceptual issues associated with modelling traits and health are discussed in chapter one. The next three chapters concern such measurement issues about: personality traits (chapter two), health behaviours (chapter three) and health outcomes, with particular reference to functional health status (chapter four). These chapters are followed by a move to modelling (chapter five), with particular reference to the generalized latent variable modelling (LVM) framework (Muthén & Muthén, 1998–2007). The HAPPLE study is introduced (chapter six) which is used to model associations between Conscientiousness and health criteria within the LVM framework (chapter seven). Moving beyond self-reported outcomes, which are a mono-method approach, the role of multiple health behaviours in predicting cardiovascular mortality is considered (chapter eight). In a third section, cortisol is introduced, which is a biomarker of stress reactivity. The diurnal profile of cortisol output is described (chapter nine). Latent growth curve modelling is used to illustrate its association with Neuroticism, in a sample of student volunteers (chapter 10). Taken together, the results highlight the need for a general

framework of modelling techniques, in personality-health research. I conclude that biopsychosocial models with excellent explanatory power, which are still parsimonious, can be achieved with LVM and its extensions. However, trait researchers will need to state more clearly the intended destinations of their work in order to attract contributions from, and share knowledge with, other disciplines.

Word count: 62,500

Acknowledgements

I would like to thank Martha Whiteman, Anthony Coxon and Ian Mason for their supervision. For various sources of advice, information, support and encouragement, I would like to thank Edward Boniface, Martin Corley, Angela Clow, Ian Deary, Marshall Dozier, Paul Dudgeon, Andy Fugard, Catherine Fredhoi, Lynn Hagger, Harry Hagger-Johnson, Megan Hagger-Johnson, Barry Johnson, John Johnson, Corey McMillan, Erik Meijer Jeremy Miles, Paul Morris, Bengt Muthén, Linda Muthén, Susan Shenkin, Darren Shickle, Richard Street, Andrew Wawrzyniak, Robert West, Carol Wollaston.

Contents

List of Tables

List of Figures

List of Abbreviations

A	Agreeableness
A Level	Advanced Level
A1	Trust
A2	Straightforwardness
A3	Altruism
A4	Compliance
A5	Modesty
A6	Tendermindedness
AC	Accident Control
AFI	Approximate Fit Indices
AIC	Akaike Information Criterion
AUC	Area Under Curve
BIC	Bayesian Information Criterion
BP	SF-36 Bodily Pain
C	Conscientiousness
C1	Competence
C2	Order
C3	Dutifulness
C4	Achievement Striving
C5	Self-Discipline
C6	Deliberation
CAR	Cortisol Awakening Response

CASCOT ..... Computer Assisted Structured COding Tool
CFA ..... Confirmatory Factor Analysis
CFI ..... Comparative Fit Index
CFS ..... Chronic Fatigue Syndrome
CHD ..... Coronary Heart Disease
CVD ..... Cardiovascular Disease
Dex ..... Dexamethasone
DP ..... Diurnal Cortisol Profile
DSM-IV-TR ..... Diagnostic and Statistical Manual of Mental Disorders, 4th edition, text revision
E ..... Extraversion
E1 ..... Warmth
E2 ..... Gregariousness
E3 ..... Assertiveness
E4 ..... Activity
E5 ..... Excitement-Seeking
E6 ..... Positive Emotions
EFA ..... Exploratory Factor Analysis
ELISA ..... Enzyme-Linked ImmunoSorbent Assay
EV ..... SF-36 Energy Vitality
GCSE ..... General Certificate in Secondary Education
GH ..... SF-36 General Health
HALS ..... Health and Lifestyle Survey
HAPPLE ..... Health and Personality Processes: Links Explored
HBC ..... Health Behaviour Checklist
HPA ..... Hypothalamic-Pituitary-Adrenal
HTML ..... Hypertext Markup Language
IMR ..... Internet Mediated Research
IPIP ..... International Personality Item Pool
IPIP NEO ..... IPIP representation of the NEO-PI-R
LCM ..... Latent Curve Modelling

LVM ..... Latent Variable Modelling
MCS ..... SF-36 Mental Component Summary
ME ..... SF-36 Mental Health
ML ..... Maximum Likelihood
MLM ..... Multi-level modelling
MTMM ..... Multitrait-Multimethod Matrix
N ..... Neuroticism
N1 ..... Anxiety
N2 ..... Angry Hostility
N3 ..... Depression
N4 ..... Self-Consciousness
N5 ..... Impulsiveness
N6 ..... Vulnerability
NEO-FFI ..... NEO Five-Factor Inventory
NEO-PI-R ..... Revised NEO Personality Inventory
NHS ..... National Health Service
NSB ..... Non-Specific Binding
O ..... Openness to Experience
O1 ..... Openness to Fantasy
O2 ..... Openness to Aesthetics
O3 ..... Openness to Feelings
O4 ..... Openness to Actions
O5 ..... Openness to Ideas
O6 ..... Openness to Values
OLS ..... Ordinary Least Squares
PCS ..... SF-36 Physical Component Summary
PF ..... SF-36 Physical Functioning
PHP ..... Personal Home Page
PTSD ..... Post Traumatic Stress Disorder
RE ..... SF-36 Role Emotional
RMSEA ..... Root Mean Square Error of Approximation

RP ..... SF-36 Role Physical

RPM ..... Rotations per minute

SEM ..... Structural Equation Modelling

SES ..... Socio-Economic Status

SF ..... SF-36 Social Functioning

SF-36 ..... SF-36 Health Outcome Survey

SOC ..... Standard Occupational Classification

SQL ..... Structured Query Language

SR ..... Substance Risk

SRMR ..... Standardized Root Mean Residual

SSL ..... Secure Sockets Layer

SSRI ..... Selective Serotonin Reuptake Inhibitor

TLI ..... Tucker-Lewis Index

TMB ..... Tetramethylbenzidine

TR ..... Traffic Risk

TSST ..... Trier Social Stress Test

WLS ..... Weighted Least Squares

WME ..... Wellness Maintenance and Enhancement

CHAPTER 1

Introduction

> *If personality traits do influence health, then this is one of the prime reasons to measure personality traits in medical settings. However, there are difficulties in establishing the true nature of the relationship between personality and health, including measurement, the distinction between subjectively reported symptoms and objective signs of illness, and the direction of causation.*
> (Matthews, Deary, & Whiteman, 2003, p. 273)

1.1 Introduction

Personality traits are sometimes found to influence health (e.g., Huovinen, Kaprio, & Koskenvuo, 2001; Wilson et al., 2005), sometimes they are not (e.g., Huppert & Whittington, 1995; Nakaya et al., 2003; Schapiro et al., 2003). Results partly depend on the statistical model chosen by the researcher to examine the data (Wiebe & Fortenberry, 2007). These observations inform my central question, which runs throughout the thesis, of the importance of statistical models in exploring and understanding relationships between personality traits and health. Does model choice mask, or reveal, relationships between traits and health (Wiebe & Fortenberry, 2007, p. 150)? A key solution, which also runs throughout the thesis, highlights the need for single generalized framework for personality-health models.

1.1.1 Definitions of personality traits and health

Health has been defined broadly as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (World Health Organization, 1946, cited in Bowling, 1997, p. 4). Many other definitions of health are possible, including those emphasizing functional health status, which covers "the implications of disease and treatment in terms of what people are able to do and how they feel" (Ware, Snow, Kosisnki, & Gandek, 1993, p. 9:1). Health may also refer to health behaviours (e.g. cigarette smoking, alcohol use), which are usually studied because they raise the risk of disease and death (Belloc, 1973; Doll, Peto, Boreham, & Sutherland, 2004; Duffy, 1995; Jarvis & Wardle, 2006). Of the top ten risk factors globally, three are health behaviours (unsafe sex, tobacco consumption, alcohol consumption) two are strongly associated with behaviour (high blood pressure and obesity) and the remaining three are partly behaviour, partly public health issues (underweight, sanitation and hygiene; iron deficiency; indoor smoke from solid fuels, World Health Organization, 2002, p. 7). Personality traits refer to "those characteristics of the person that account for consistent patterns of thinking, feeling and behaving" (Pervin, Cervone, & John, 2004, p. 6). Since people's consistent patterns of behaviour affect what happens to them, the relationship between traits and health has long interested researchers (Matthews, Deary, & Whiteman, 2003).

1.1.2 Constructs

A construct is "an informed, scientific idea developed or constructed to describe behaviour. We can't see, hear, or touch constructs, but we can infer their existence from overt behaviour" (R. K. Cohen & Swerdlik, 1998, p. 17). It may, however, be possible to measure their effects (Miles & Shevlin, 2000, p. 203). In personality-health research, both personality traits and health are constructs. Constructs are variables that cannot be measured directly, but the researcher assumes that they exist and can be measured indirectly, or imperfectly, with some error. The task of the researcher is to minimize this error as much as possible

(R. B. Kline, 2005, p. 96), before attempting to show how constructs are related to each other in a statistical model (Cliff, 1983). The purpose of models is to reveal an association between two or more variables, and communicate their relationship to a specified audience (Tabachnick & Fidell, 2000), "in a simple way so that it is understandable and interpretable" (Muthén & Muthén, 1998–2007, p. 1). In modelling parlance, this is called the "association" between two variables. Maximizing the size of the association between personality and health is a key aim in the field (Mershon & Gorsuch, 1988; Paunonen, 1998; Paunonen & Ashton, 2001; Paunonen, 2003; Wasylkiw & Fekken, 2002).

1.1.3 Choice of constructs

No model can contain all of the variables that connect personality with health. Good models are parsimonious (efficient), which mean that they do not contain too much information (Crawley, 2007). There are several different measures of traits, of health behaviours, and of health outcomes. There are many more different strategies for testing their relationships. The broad umbrella term "methodology" describes the process by which researchers choose what to measure, how to measure it, and how to model it: "how we will go about studying any phenomena" (Silverman, 2001, p. 4). Methodology concerns study of research methods themselves (Langdridge, 2003, p. 258), which is not the same as methods, which are the specific techniques used to do research (Silverman, 2001, p. 4). Constructs usually undergo a process called validation before they are put into models and granted the status of variables (Borsboom, Mellenbergh, & Heerden, 2003; Cliff, 1983). Validation demonstrates that measures actually reflect what they are supposed to measure, and that the measure is reliable (P. Kline, 2000). Model choice, then, intimately depends on methodology.

1.1.4 A proliferation of modelling techniques

Unfortunately, I argue, the proliferation of approaches to modelling (T. Smith, 2006) has limited the potential for personality-health research to (1) develop

clear, unambiguous and testable measures and hypotheses (Wiebe & Fortenberry, 2007); and appropriate explanations (T. Smith, 2006); (2) communicate its findings to neighbouring disciplines (R. Hogan, 2005; Krueger, Caspi, & Moffitt, 2000) of the research. Commentators in the health sciences have been skeptical about the field (e.g. Stansfeld, 2002a), but measures of traits are needed in medical and public health settings if personality-health research is to progress (Krueger et al., 2000; Matthews, Deary, & Whiteman, 2003). I return to these points in the concluding chapter.

1.1.5 Similarities between methods

It turns out that specific statistical techniques have a great deal in common (Muthén & Muthén, 1998–2007), and it while it is customary to emphasize the particular unique features of a technique (e.g., Tabachnick & Fidell, 2000), it may be more fruitful in the long run to emphasize the similarities (Kirkwood & Sterne, 2001; McDonald, 1985). It is possible, I will argue, to do both. This thesis contains a number of techniques, some of which are rarely used by personality-health researchers. However, they are all "special cases" (Muthén, 2002, p. 83) of a larger framework of models. This framework is called the generalized latent variable modelling framework, has been articulated by Muthén (2002). It is introduced in chapter five.

1.1.6 The field

The thesis belongs in the field of personality-health research, which is part of two sub-disciplines of psychology:

Differential. Differential psychology is the study of individual differences in personality (H. J. Eysenck & Eysenck, 1985), mental ability (Deary, 2000), and other traits (P. Kline, 2000), using psychometric methodology.

Health. Health psychology is the psychological approach to understanding health behaviours and health outcomes (Ogden, 2007), which are defined in chapters three and four.

In order to connect traits with "unambiguous outcomes (e.g., mortality, objectively diagnosed disease)" (T. Smith, 2006, p. 228), personality-health researchers must study biological variables, psychological constructs, and health behaviours. This means that they must use methods from several other disciplines, beyond differential and health psychology. Personality-health research is "necessarily inter-disciplinary" (Evans, Hucklebridge, Clow, Hucklebridge, & Angela, 2000), not least because measuring health is an inter-disciplinary enterprise.

The role of constructs in differential and health psychology

The relationship between differential and health psychology is not straightforward, and is often researched for reasons other than intrinsic interest in personality and health. Differential psychologists have focused on fewer, more tightly defined constructs, such as intelligence or personality traits (Deary, Clyde, & Frier, 1997; Deary, 2000). In order to demonstrate that these constructs are valid, they need to show that they are related to other variables, outside (external to) the construct (Deary & Hettema, 1993). If measures of Conscientiousness or mental ability influence health, then this provides evidence to support their existence. According to Bowers (1987), health was often used in personality research purely for this reason. In contrast, health psychologists have tended to resist higher order constructs (e.g. ability, personality) and developed a larger number of "newer" constructs which have proliferated in the literature (Deary, Clyde, & Frier, 1997). Many of these could be criticized for falling foul of the jangle fallacy (Deary, 2000, p. 111), the tendency to give broader constructs such as Neuroticism different names (e.g. trait anxiety, trait negative affectivity, trait negative emotionality). It could be argued that differential psychologists favour "fidelity" (a smaller number of narrowly defined constructs) and health psychologists favour "bandwidth" (comprehensive coverage of a larger set of constructs; Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995; Briggs & Cheek, 1986; J. Hogan & Roberts, 1996a). This is undoubtedly an over-simplified characterization, but it nonetheless highlights some of the difficulties that inter-disciplinary researchers face.

1.1.7 Assumptions

A field of enquiry may have a set, or more than one set, of implicit assumptions. Implicit assumptions are shared understandings between researchers, that may not be articulated in research reports and publications. It was argued that sciences operated under a single set of assumptions, called a paradigm (Kuhn, 1996). When a paradigm could not explain a phenomenon, a period of rapid and divergent change would occur, called a paradigm shift. Researchers would begin to "describe and interpret them [phenomena] in different ways" (Kuhn, 1996, p. 17). Once a new paradigm had been agreed, a period of calm would allow researchers to focus on explanation. However, later commentators demonstrated that several sets of theories can be studied simultaneously. The "research program" was possible to evaluate because multiple working hypotheses could be evaluated according to empirical research. In the early stages of a research program, there may be a proliferation of ideas and terms that appear in the literature. It is important to articulate assumptions made, and to use empirical research to compare the relative merits of different theories. It is normal to have several sets of competing theories. Empirical research "is the final arbiter among competing research programs" (Gholson & Barker, 1985, p.758). This account best characterizes the status quo in personality-health research. There are several competing theories of personality and health, and a fruitful exchange of ideas between research programs. It is widely recognized that more than one theory or hypothesis can be considered simultaneously, so that the discipline is not restricted by a narrow focus (Chamberlin, 1965). Researchers do not operate in a single paradigm, even if they are influenced by a historical research tradition. Rapid development is encouraged when a study can demonstrate that a new account is conceptually possible, opening up the potential for empirical support, even if that study does not finalize the explanation (Gholson & Barker, 1985). Implicit assumptions, then, can be questioned by empirical evidence.

Assumptions in differential and health psychology. Differential psychologists tend to assume two things. First, if people differ, they must differ by certain degrees

(H. J. Eysenck, 1992). Second, "if something does not exist, then one cannot measure it" (Borsboom, 2005, p. 150). It therefore follows that if a construct does exist, by certain degrees, then it can be measured. Third, constructs have "causal primacy" (Matthews, Deary, & Whiteman, 2003, p. 6). This means that traits are normally considered as independent variables, variables that influence health, not vice versa. Health psychology is characterized by a "move away from a simple linear model of health" in which a physical variable causes a disease (Ogden, 2007, p. 3). The move is towards an assumption that biological, psychological and social factors are all involved in health behaviour and health outcomes: this is the biopsychosocial model (Adler & Matthews, 1994). Differential and health psychology, then, combines health psychology and differential psychology, to study the relationship between individual differences and health: personality-health associations.

1.2 Five research questions

The research I undertook for this thesis was designed to address several gaps in the personality-health literature, and answer a number of specific questions. Korotkov and Hannah (2004) noted that the field was "plagued by several theoretical and methodological concerns" (p. 188). The background to some of these concerns will be provided in the next few chapters. The following five research questions were generated to reflect these gaps:

Bandwidth. Can traits and facets (narrower constructs than traits) be included in the same model as broad and narrow measures of health? This would address the bandwidth-fidelity dilemma. Paunonen (1998); Paunonen and Ashton (2001); Paunonen (2003) argued that addressing the neglect of personality facets was a clear priority for personality-health research.

Health behaviours. Several researchers have included a limited range of health behaviours in personality-health models (see chapter three). Can including a larger number of behaviours (R. R. Vickers, Conway, & Hervig, 1990) offer any additional insights? What is the validity of a general health behaviour con-

struct? How many dimensions are needed to summarize the covariance of multiple health behaviours?

Outcomes. Given that education and health behaviours have been identified as mechanisms that could influence health (Hampson, Goldberg, Vogt, & Dubanoski, 2007), can broader constructs, such as socio-economic status and multiple health behaviours, explain more variation in health outcomes? Does the association extend to mental health outcomes, in addition to physical ones?

Mechanism. Cortisol, a biomarker for stress reactivity (Clow, Thorn, Evans, & Hucklebridge, 2004), has a large literature supporting its relationship to health status (e.g., McEwen et al., 1997; McEwan, 1998). It is a possible mechanism that links traits, such as Neuroticism, with illness. However, evidence to support this association has been inconsistent. How should cortisol be measured and modelled, so that correlations can be discovered and maximized (Pruessner et al., 1997)? What is the appropriate methodological gold standard for cortisol measurement, that can be adopted by trait researchers? Does the association between Neuroticism and cortisol exist at all?

Biomarker. A secondary question, which is described in chapter 10, concerns whether self reports of health are associated with cortisol. If they are, this is a substantively important finding. It would provide some preliminary evidence that cortisol is a biomarker for functional health status, as well as a biomarker for stress (Clow et al., 2004)

Maximization. An unresolved issue in cortisol-trait research is how to aggregate the data and model in such a way that not only detects an association, but maximizes the size of that association (Pruessner et al., 1997). This is particularly important for cortisol-trait research.

For each of these problems, I argue that the solution is a combined measurement and modelling strategy. Finding measures and models that are appropriate for personality-health data is the departure point for the thesis. The thesis is about more than maximizing correlations between personality and health, and the problems outlined above. It is also about the intimate relationship between

measures, methods and results. Later chapters emphasize the need for a systematic, rather than idiosyncratic, approach to modelling personality-health associations. In the end, it is a call to look toward what R. Hogan (2005) called the "consumers" of trait research, thinking about model audience. One major purpose of a model, after all, is to communicate the model (Muthén & Muthén, 1998–2007), to a specific audience (Tabachnick & Fidell, 2000).

1.3 Structure of the thesis

Introduction. This introductory chapter has introduced the field of personality-health research, located with the sub discipline of differential and health psychology. A line of enquiry to help address several gaps in the field is now proposed:

Personality traits. Chapter two describes the big five model of personality traits, and the instruments chosen to measure them (Goldberg, 1999; Costa & McCrae, 1992a). The rationale for using the big five, a popular and well validated model of personality is defended. The big five can be measured using long or short personality questionnaires. Long questionnaires provide both traits and narrow "facets". This is important because a key debate in the personality-health literature concerns whether measuring traits, but not facets, compromises specificity in favour of breadth. This chapter provides the foundation for understanding how I chose to measure traits and facets in my research.

Health behaviours. The purpose of chapter three is to consider health behaviour measures, emphasizing that they can be measured alone or in combination. Multiple health behaviours can be modelled as dimensions. For example, I made the decision to measure health behaviours using the Health Behaviour Checklist (HBC, R. R. Vickers et al., 1990) and to take separate measures of cigarette smoking and alcohol use.

Health outcomes. Chapter four introduces the SF-36 Health Outcome Survey as a validated measure of functional health status. The SF-36 summary scores are operationalized as health outcomes, meaning that they are downstream from variables that predict health. This should not imply that health outcomes cannot influence upstream variables, only that they are suitable as dependent variables in models of personality-health.

Latent variables. Chapter five is an introduction to the generalized latent variable modelling framework (Muthén, 2002). This framework is the foundation for subsequent models presented in the thesis.

HAPPLE study. Chapter six describes the Health and Personality Processes: Links Explored (HAPPLE) Study. The study was designed to capture a wide range of demographic variables, particularly socio-economic status (derived from measures of education, occupational status, and area based deprivation scores) and a wide range of health behaviours. It describes the aim, methods, and choice of internet mediated data collection.

Conscientiousness. In chapter seven, the association between Conscientiousness and health criteria is examined. Socio-economic status and multiple health behaviours are evaluated as potential mediators of the association.

The big four. Chapter eight revisits the modelling of multiple health behaviours, testing the predictive validity of the big four health behaviours in relation to cardiovascular mortality. The Health and Lifestyle Survey (HALS) data are used to model health behaviour covariance and survival.

Cortisol. Chapter nine introduces cortisol, a stress hormone. Cortisol provides an objective biological measure, and is a candidate mechanism for explaining why traits, particularly Neuroticism, could be related to illness (Pruessner et al., 1997).

Growth modelling. Chapter 10 presents the results from the cortisol study, and illustrates how the generalized latent variable modelling framework can detect, and maximize, associations with personality traits. It illustrates how latent growth curve modelling can incorporate time (hours during the day). It addresses an important gap in the literature, which was the lack of an adequate way

to aggregate cortisol data efficiently (Pruessner et al., 1997). Results showed that Neuroticism was associated with cortisol output across the day, and this was accounted for by variance from the N2 Angry Hostility facet. Higher N2 scores were associated with greater cortisol output. This is the same facet that has very recently been linked to genetic markers for hypothalamic-pituitary-adrenal (HPA) axis activity (Wasserman et al., 2007). Furthermore, self reported general health status, as measured by the General Health scale of the SF-36, was associated with cortisol output. Higher GH scores were associated with lowered cortisol output. This provides preliminary evidence that cortisol could act as a biomarker for general functional health status.

Conclusion. In the concluding chapter, I defend the line of enquiry taken in the research. It makes the case for latent variable modelling (LVM), future extensions, and the importance of engaging consumers of personality-health research, so that the possibilities of collaboration and communicating with audiences outside the field can be made part of the modelling process.

The conclusion will summarize the main findings and discuss implications and potential applications. Specifically, I propose that: (1) the generalized latent variable modelling framework should be used to communicate models in a common modelling parlance; (2) models of personality-health should contain indicators of socio-economic status (e.g. education), personality facets, multiple health behaviours and gender. These are substantively important variables; (3) there are at least five ways in which personality-health associations have applied utility.

Part I

Measurement models

CHAPTER 2

Measuring personality traits

> *[T]he Big Five is a reasonably agreed upon system and for now the best working hypothesis* (de Raad & Goldberg, 2002, p. 18).

2.1 Introduction

The first three chapters of this thesis are about measurement. This chapter is concerned with personality traits. It is divided into four sections. The first section introduces the big five model of personality traits. The second describes reliability and validity, with particular reference to the NEO-PI-R and IPIP NEO instruments. The third considers the big five measurement models in more detail, and the facets that underlie each trait. Finally, psychobiological approaches to traits are considered and their potential role as explanatory mechanisms or biomarkers. The central claim in this chapter is that the big five model is sufficiently validated as a descriptive tool. Therefore, researchers can adopt big five measures of traits and use them in structural models, where traits are linked to health behaviours and health outcomes.

2.2 Personality theory

Personality traits are only one part of personality. Other aspects of the person, for example, include: character, biography, reputation, memory, emotion and psychodynamic processes (McCrae & Costa, 1996; Funder & Funder, 2001). The McCrae and Costa (1996) theory of personality covers basic tendencies, characteristic adaptations, objective biography, self-concept, external influences plus dynamic processes which are the interaction between these elements. Personality traits are included under basic tendencies alongside genetics, physical characteristics, cognitive capacities, physiological drives and focal vulnerabilities. Under basic tendencies, McCrae and Costa (1996, p. 72) listed individuality (all adults can be characterized by their differential standing on a series of personality traits that influence patterns of thoughts, feelings and behaviours, origin (personality traits are endogenous basic tendencies); development (traits develop through childhood and reach mature form in adulthood, thereafter they are stable in cognitively intact individuals, structure (traits are organized hierarchically from narrow and specific to broad and general dispositions, the Big Five traits represent the highest level of this hierarchy. The "arrows" are processes, connecting these different elements bring their "model to life" (p. 77) and include information processing, coping and defence, volition, regulation of emotions, interpersonal processes, and identity formation. Traits represent one aspect of this personality theory.

2.2.1 Defining personality traits

Formally defined, personality traits refer to "those characteristics of the person that account for consistent patterns of thinking, feeling and behaving" (Pervin et al., 2004, p. 6). Typically, trait researchers are interested in situational consistency in inter-individual differences across large samples of the population (Cervone & Pervin, 2007) — the stable portion throughout time that is consistent across many situations (Funder & Funder, 2001). The key word here is consistency — the stable portion of the person. Stability is useful to the trait researchers because

it means it can be measured and assumed to exert an influence on health related variables over time. In contrast, moods are "states" which cannot be assumed to be ever-present. Like descriptions in general, descriptions of traits involve their most salient features. Traits are a nomothetic, not idiographic, approach because they describe traits relative to others (nomothetic), rather than focusing on an individual person (idiographic). Idiographic approaches assume that no consistencies or generalized statements can be made across people (Matthews, Deary, & Whiteman, 2003, p. 6). The term "idiothetic" can also be used, which refers to the measurement of a person's traits, but without applying norms to them (Paaunonen & Jackson, 1986). Norms are scores on psychometric measures that have been standardized with reference to a "reference source" (R. K. Cohen & Swerdlik, 1998, p. 117), often a large representative sample of the general population.

2.2.2 *History of trait description*

People have long tried to describe individual differences in personality. As early as ancient Greece and ancient China, people tried to formulate theories to "systematize diversity" (Van Gestel & Van Broeckhoven, 2003) they saw in people's personality traits. Modern studies of personality included Freud's work and the influential Authoritarian Personality, written by Adorno and the Frankfurt school (J. L. Martin, 2001). Freud developed his theory of the id, ego and superego. Relics of Freudian constructs (e.g. "ego") still survive in the terminology of the DSM-IV-TR (American-Psychiatric-Association, 1994). These constructs are generally not used by trait researchers today. Personologists have retained an interest in personality types derived from these works, such as hystericals, obsessionals and narcissists (e.g. Young-Bruehl, 2003).

History of trait measurement

Allport was perhaps the "founding father" (Winter, 1997, p. 723) of the modern period of trait research. Since that time, the popularity of personality trait

research has come in waves (R. Hogan, 2005). In terms of traits, there was a peak after World War II, then a dip during the "faking controversy" of the 1960s (R. Hogan, 2005). In this period, it was claimed that traits lacked validity because people could fake their responses to personality questionnaires (R. Hogan, 2005). After the critique that situations were better predictors of behaviour than traits, in *Personality and Assessment* from Mischel (1968), trait research almost drowned — and had certainly hit rock-bottom (see also Gould, 1996). Trait research resurfaced in a third wave, powered by the tide of meaningful associations with "real life" outcomes such as health (criterion validity) and strong psychobiological evidence (Matthews, Deary, & Whiteman, 2003; R. Hogan, 2005; Hough & Oswald, 2005). In fact, Bowers (1987) argued that health in effect became an external criterion to validate and anchor the existence of personality traits: "health and health-related behaviors were implicitly being construed as an external criterion that helped to anchor current concepts of personality" (p. 344). There are still criticisms of trait research (see below), but there has been more than ten years of buoyant theoretical and applied research on traits (Hough & Oswald, 2005; Matthews, Deary, & Whiteman, 2003). Associations with health criteria have played a large part in this growth.

2.2.3 The two areas of trait research

Trait research can be divided into two distinct, but clearly interlocking, areas (H. Eysenck, 2006). The descriptive or taxonomic area has two approaches (methods), and is concerned with description of traits (e.g. the lexical approach, and the questionnaire approach, as in the NEO-PI-R inventory), beginning with Allport's work. The biological program is concerned with causal explanations for traits, with reference to behavioural genetics and biological variables. Both areas use personality questionnaire to measure traits. Both approaches refer to the notion of hierarchies. H. Eysenck (2006) explained how in his theory of personality, single behaviours and cognitions lie at the bottom. Next, "habitual acts or cognitions" such as unpunctuality, summarize these behaviours and cognitions. Next, are traits, which H. Eysenck (2006) defines as "significant intercorrelations

between different habitual behaviors" (p. 244). At the top are higher order factors or dimensions of personality. The task of finding explanations for traits is not the aim of this thesis. However, it is worth noting that biological explanations for traits are equally important for both the descriptive approach to traits (Goldberg, 1999; Costa & McCrae, 1992a), and the biologically informed theories of (H. Eysenck, 2006). They are not mutually exclusive.

2.2.4 Personality adjectives

Adjectives in natural language are used to describe people's traits. Ashton and Lee (2005a) note that this was recognized long ago, but was formalized into a hypothesis more recently. Formally stated, the lexical hypothesis is that "[i]mportant phenotypic attributes become encoded in the natural language" (Saucier & Goldberg, 2001, p. 848). This process occurs over time, so that etymology, the study of how words enter the language, is a resource for trait researchers (Piedmont & Aycock, 2007). It provides a methodology for measuring "when constructs became lexically formalized" (p. 1062). Phenotypic attributes are observable, describable traits. It therefore follows that researchers can find full coverage of the important traits, "by referring to the personality lexicon", that is, trait terms in the dictionary (Ashton & Lee, 2005a, p. 6). It does not follow that trait terms are explanatory — their function is purely descriptive (Saucier & Goldberg, 2001). People can rate themselves and others on lists of adjectives according to how strongly they agree it describes their traits, for example, on a Likert scale from 1 (strongly agree) to 5 (strongly disagree). These ratings are also called "responses" or "endorsements". Therefore, it should be possible to analyse responses adjectives in the language and reduce them to "a relatively small set of roughly independent axes along which people differ in typical behavioral tendencies" (Ashton & Lee, 2005a, p. 7). This process was conducted, and researchers concluded that five dimensions are necessary to capture and describe the main dimensions of variation. They are the same dimensions that were found by researchers using the questionnaire method, which supports the notion that five dimensions are appropriate (de Raad & Goldberg, 2002). In the next section, I describe the test

development process. This applies both to the adjective and the questionnaire approach to personality trait measurement. In summary, the big five model has become the dominant paradigm in trait psychology, an "organizer in fields of experience and research" (de Raad & Goldberg, 2002, p. 5). The history of this model is now considered.

2.2.5 History of the NEO-PI-R and the IPIP

The NEO-FFI is a personality inventory which distils four decades of research into five constructs (see Costa & McCrae, 1992a): Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A) and Conscientiousness (C). These traits emerged when analyzing four major data sets (Costa, McCrae, & Dye, 1991). The NEO began with N and E, with O added in 1976-8. The original NEO was published in 1985 and contained traits and facets for N, E and O; with traits only for A and C (Costa et al., 1991). The NEO-PI-R is a longer version of the NEO-FFI which contains traits and facets for all five factors (Costa & McCrae, 1992a). Two forms of the NEO-PI-R are available. For S uses self-reports, and form R uses ratings made by other people (raters). Correlations between scores on the two forms provide evidence for inter-rater reliability (described below). The NEO inventories can be described as palimpsests — multi-layered records (Soanes & Stevenson, 2005) of several instruments that have been reinforced, added to, and changed over the years. The NEO-PI-R as it stands today, retains traces of its earlier forms. The International Personality Item Pool (IPIP), since 1999, has provided freely available personality items that correspond to the NEO and other inventories (Goldberg et al., 2006). It contains big five scales in various lengths (300, 100, 120 and 50 items). Goldberg's aim was to increase the pace at which trait research could proceed, by reducing the costs involved. The correlation between the IPIP NEO and the NEO-PI-R scales is high, providing evidence that it measures the same constructs.

2.2.6 The development of facets

Personality facets are more specific, narrow measures of traits that combine to make up the five traits in the hierarchical big five model (de Raad & Goldberg, 2002). Personality facets are defined by personality items, and personality traits are defined by personality facets in turn. Facets are narrower, more detailed descriptions of traits, lower down in the hierarchy. Extending the Austin and Deary (2002) analogy, if traits were centimeters on "rulers", then facets would be millimeters. The relative consensus for the Big Five does not yet extend to the lower order facets (Roberts & Bogg, 2004). Costa et al. (1991) argued that the adjective approach was inappropriate for facets, because psychologically important facets are not likely to be encoded as adjectives, in the same way as traits, by lay people. They argued that the best basis for naming and describing facets was the psychological literature. Crucially, the facets identified for each domain were "not because each is naturally divisible into six parts, but because at least six dimensions were suggested by the literature, and more than six scales would tax the user's ability to learn and remember the facets" (p. 888). Therefore, traits were developed from the "top down" using factor analysis, and facets were developed from the "bottom up", selected to synthesize literature and appeal to users. This is a controversial strategy for the development of a measurement model, because these criteria are not driven by empirical research findings (Van Gestel & Van Broeckhoven, 2003). However, several researchers have argued that having a reference point (the big five) outweighs the current limitations of facets (de Raad & Goldberg, 2002). They can still be used in research settings, and there is no reason why they cannot be further improved in the future.

2.3 Reliability and validity

The first stage in developing a measure of personality traits is to have a construct. A construct is "an informed, scientific idea developed or constructed to describe behaviour. We can't see, hear, or touch constructs, but we can infer their existence

from overt behaviour" (R. K. Cohen & Swerdlik, 1998, p. 17). Constructs are theoretical properties — researchers cannot measure personality traits directly, but they can hypothesize that they exist. In order to measure them, they need to take several indirect measures, usually several personality *items*. On the assumption that a construct exists, the task of the researcher is to create a scale that can measure it. According to classical test theory, the dominant paradigm in psychometrics, the measure of the construct should be reliable and valid (P. Kline, 2000). The NEO-PI-R and IPIP are now used to illustrate the scale construction process, with particular reference to reliability and validity.

2.3.1 Reliability

In classical test theory, a score obtained in response to an item consists of variance from the construct, and error variance (P. Kline, 2000). The aim is to minimize error variance (P. Kline, 2000), by extracting the variance that items share. The less error a scale has, the more reliable it is (P. Kline, 2000). Reliability has four forms, all of which are evaluated using correlations:

Inter-rater. If the traits of a person are rated by two people, a reliable instrument will produce a high correlation between the two ratings. For the NEO-PI-R (Costa & McCrae, 1992a) significant inter-rater reliability (self and spouse) coefficients are .54 (N), .44 (O), .35 (C). Peer and self comparison studies resulted in coefficients of .30, .40, .38., .31, .34 (N, E, O, A and C). Peer and peer comparison resulted in .43, .42, .45., .49 and .22 (N, E, O, A and C).

Test-retest. A reliable scale will correlate highly between measurement time points, assuming that the construct it measures is stable (P. Kline, 2000). This component of reliability refers to the "consistency from one measurement to another" (Cronbach & Shavelson, 2004, p. 394). Test-retest stability coefficients in the NEO-PI-R are .79, .79, .80, .75 and .83 (N, E, O, A and C, respectively Costa & McCrae, 1992a). To my knowledge, the test-retest stability of the IPIP NEO has not been tested (see chapter six, where I present test-retest data for this instrument).

Parallel forms. If two scales were developed in the same way, to measure the same construct, then they should correlate highly. Correcting for differences in internal consistency between the IPIP NEO and the NEO-PI-R, the correlation between each inventory for the facets is .90 to .98 (N), .91 to .99 (E), .90 to .99 (O), .86 to .97 (A), .87 to .99 (C). This provides evidence of parallel forms reliability for the IPIP NEO (Goldberg, 1999). This is not the same as convergent validity, described below.

Internal consistency. This is the form of reliability that has been adopted most widely, perhaps because it is necessary for other forms of reliability and validity. It concerns the internal properties of a scale (P. Kline, 2000). It is seen by many researchers "as the correlation of an instrument with itself" (Cronbach & Shavelson, 2004, p. 394). Generally speaking, a value of .80 is considered internally consistent (P. Kline, 2000). However, researchers have claimed that a criteria of .70 (e.g., Ware et al., 1993), or even .60 (e.g., Phillips, Douglas, Burns, & Mark, 2005), can be considered acceptable. There are four ways to calculate internal consistency:

Inter-item. The average inter-item correlation is the mean correlation, taking every pair of correlations between items (Briggs & Cheek, 1986; R. K. Cohen & Swerdlik, 1998; Diamantopoulos, Adamantios, Siguaw, & Judy, 2006).

Item-total. To calculate the average item-total correlation, a score is created from the sum of the items, and the correlations with this score are included in the calculation (R. K. Cohen & Swerdlik, 1998; P. Kline, 2000).

Split half. Here, a set of items is split into two, and the scores between these splits are correlated. A high correlation "two scores are obtained from a single testing by scoring separately the odd-numbered items and the even-numbered items" (Cronbach & Shavelson, 2004, p. 395).

Cronbach alpha. This is the same as split half reliability, but for the average of all random possibilities of splits. The formula allows an estimate to be calculated, so this is not done by hand. The estimate is not perfect, but this

should not be interpreted as meaning that alpha is always an underestimate of reliability (Cronbach & Shavelson, 2004). In the NEO-PI-R, internal consistency coefficients are .92, .89, .87, .86 and .90 (N, E, O, A and C respectively). These show that the measures are internally consistent. For facets, internal consistencies range from .68 to .78 (N), .63 to .77 (E), .58 to .80 (O), .56 to .79 (A), .62 to .75 (C). These values are lower, but may reflect the smaller number of items used to measure each facet. The internal consistencies of the IPIP NEO facets range from .77 to .88 (N), .64 to .81 (E), .61 to .84 (O), .61 to .84 (A), .67 to .80 (C) (Goldberg, 1999).

In summary, reliability is the relationship of a test to itself, its internal characteristics (P. Kline, 2000). However, it also refers to the consistency in its measures. Extending the analogy from Austin and Deary (2002), that traits are like rulers, researchers would not expect a ruler to perform with less consistently depending on the situation or object being measured. Similarly, the reliability of a scale should, ideally, not change as a function of the internal characteristics of the test itself.

2.3.2 Validity

A valid scale measures what it claims to measure (P. Kline, 2000). There are three components to validity:

Face. Face validity means that the scale appears to measure what it claims to measure (Cronbach & Shavelson, 2004). This is assessed subjectively by the researcher (R. K. Cohen & Swerdlik, 1998).

Content. Content validity refers to the content coverage of the construct a measure is designed to assess. For example, a measure that assessed only part of the Neuroticism construct (e.g. Immoderation) would have poor content validity.

Criterion. Criterion validity is concerned with whether the test gives results in agreement with other measures of the same thing. There are two main types of criterion validity used by trait researchers: concurrent and predictive.

Concurrent. Concurrent validity is shown by a high correlation between two tests, which are taken at the same time. It is essentially the same as parallel forms reliability. The problem with both is that the "other" test is assumed to be the gold standard scale that actually measures the construct. There may not be a suitable alternate form which you can use. This is often the reason for developing a new test.

Predictive. Predictive validity means "refers to ability of a test to predict a relevant criterion" (Cronbach & Shavelson, 2004, p. 36). Typically, this is assessed by looking for a correlation between the scale and another variable. Taking the example of Conscientiousness, it has been shown to predict many different health behaviours (Bogg & Roberts, 2004; Booth-Kewley & Vickers, 1994; Korotkov & Hannah, 2004; Wasylkiw & Fekken, 2002).

The term *incremental* validity is used to refer to the additional predictive validity that a test offers over and above an existing test, already in use.

Construct. Construct validity is the ultimate aim in scale development, and research into the construct more generally. It comprises convergent validity and discriminant validity, both of which are assessed as a set of correlations. However, these correlations alone cannot provide construct validity. They must connect sensibly with wider substantive theory about the construct (P. Kline, 2000). This is shown by the *pattern* of correlations with other constructs:

Convergent. If a test has convergent validity, it should correlate with tests that measure the same construct, but also "related constructs" (R. K. Cohen & Swerdlik, 1998, p. 201). For example, the NEO-PI-R measure of N correlates with the Eysenck Personality Inventory (H. Eysenck & Eysenck, 1964) measure of N (Costa & McCrae, 1992a). Therefore, it

has convergent validity. This is difference from parallel forms reliability because the test need not be developed at the same time, in the same way, or cover exactly the same content domain.

Discriminant. If a test has discriminant validity, it should not correlate with constructs that are not theoretically related to the construct the test is designed to measure. For example, a measure of emotional intelligence should not correlate with a measure of anxiety. This does not mean that negative correlations support discriminant validity. Negative correlations would indicate the convergent validity of a test that reflects the construct, perhaps reverse scored. It is the absence of a correlation, or a low correlation, that supports discriminant validity.

Ultimate. An additional form of predictive validity called "ultimate validity" has been proposed (O'Toole & Stankov, 1992). This is particularly relevant to personality-health research. Ultimate validity means that a scale predicts mortality, the ultimate health outcome. Ultimate validity has been demonstrated for C, which predicts mortality (e.g. Friedman et al., 1993, 1995; Weiss & Costa, 2005), and the Self Discipline facet in particular (Weiss & Costa, 2005). The reason for this association is not fully understood. It may be explained by healthy behaviours, such as adherence to treatment regimes. However, the fact that the trait measure predicts mortality is sufficient for demonstrating so-called ultimate validity.

The nomological network. Over time, construct validity is supported by what is termed a nomological network. Originally, this term was used to describe "the interlocking system of laws which constitute a theory as a nomological network" (Cronbach & Meehl, 1955, p. 290). However, trait research has not reached the status of laws (H. J. Eysenck, 1997). The term has since been used, mostly by Eysenck, to refer to the body of literature and experiments (rather than laws) supporting the construct validity of traits, such as N and E (e.g. H. Eysenck, 2006). There are three remaining points worth noting about reliability and validity, both of which are important for the models that I present later in this thesis.

First, constructs should be reliable and validated *before* they are put into models (R. B. Kline, 2005, p. 58). For this reason, the chapters concerned with measurement of traits, health behaviours, and self-reports of health (two, three and four), focus on reliability and validity. Second, reliability is necessary, but not sufficient, for validity (P. Kline, 2000). In a sense, reliability is the bread of a test, and validity is the butter. Third, measurement error will always be present, but has different implications for reliability and validity. As Cliff (1983) put it, "the variable at least partly measures something different from what we think it does" (p. 121) and the reliability problem as "the variable partly doesn't measure anything at all" (p. 121).

The Multitrait-Multimethod Matrix (MTMM). It is clear that to assess divergent validity, more than one trait should be studied. However, more than one method should also be studied. The reliability and validity of methods is often overlooked by researchers, who tend to focus on the traits themselves. Shared method variance can artificially inflate validity coefficients, because two measures can correlate simply because they are both self-reports, for example. The MTMM matrix is a matrix of intercorrelations that arise when a construct is assessed by several traits and several methods (Campbell & Fiske, 1959). Assessment of this matrix provides internal consistency, inter-rate reliability, convergent validity and discriminant validity. It provides a useful method that can be used to illustrate construct validity. For example, high correlations between self and rater reports of personality traits, suggest good inter-rater reliability. Campbell and Fiske (1959) stress that "[w]hile agreement between several methods is desirable, convergence between two is a satisfactory minimal requirement" (p. 103). It must be possible to demonstrate that correlations do not arise simply because the two measures share method variance.

2.3.3 Limitations to the classical test theory approach to scale construction.

The scale construction approach, described above, is not a perfect method for measuring traits. There are a number of limitations in the classical test theory

approach. First, it is easy to produce internally consistent scales that are misleadingly high (.9 or higher) "by writing items which are more or less paraphrases of each other" (p. 160). For this reason, moderate inter-item correlations in the .2 to .4 range are desirable (Briggs & Cheek, 1986; Diamantopoulos et al., 2006), which produce internal consistency coefficients around .8. Second, there are influences, beyond the traits themselves, that could consistently distort people's responses to items. Response sets are a term used to describe how people try to present a good impression (impression management), or respond in a socially desirable way (P. Kline, 2000, p. 159). Faking items in order to present a favourable impression, perhaps for an employer in occupational testing settings, can distort the meaning of scale scores. Third, the level of trait in a respondent is not the only variable in a scale. The level of trait that each item assesses (in ability testing, its difficulty), can change the likelihood that someone with a low or high level of that trait will respond (Embretson & Reise, 2000). Item response theory was developed in order to model the individual differences of people, and the properties of items, to address this. It can also be used to check for item bias, where the likelihood of responding to an item changes in different groups of the population (Zumbo, 1999). Fourth, some researchers argue that factor analysis and latent variables are not ideally suited to certain constructs. For example, socioeconomic status is reflected by variables such as income, education, occupation, and area-based deprivation. It is not a cause of these variables, which would be implied in the factor analytic tradition (Diamantopoulos, Adamantios, Siguaw, & Judy, 2006; Howell, Breivik, & Wilcox, 2007). Fifth, it is also important to note that there are strong critics of the factor analytic approach to scale construction in general (Butt, 2004; Gould, 1996; Swann & Seyle, 2005; Preacher & MacCallum, 2003) and of the big five specifically (Block, 1995; H. J. Eysenck, 1992), who argue that trait researchers have relied too heavily on factor analysis in scale construction. Notwithstanding these criticisms, the measurement model for the big five is reliable and well validated instrument. Therefore, it is appropriate to use in research settings. This does not mean that more advanced techniques cannot be used to improve it in the future.

2.4 The measurement model for the big five

Having established the reliability and validity for the NEO-PI-R, the focus now turns to the facets and traits themselves, in the big five hierarchy. In this section, the traits and facets of the NEO-PI-R and IPIP NEO are organized in the measurement model for these inventories, are described. Later chapters will refer to these traits and facets. In the subheadings that follow, the left hand term refers to the IPIP NEO label, and the right hand term refers to the NEO-PI-R label.

2.4.1 Neuroticism (N)

Most inventories retrieved in factor analysis studies find Neuroticism (hereafter, N) as a principal component of variation. N does not refer to a form of psychopathology, but is a relic of the term "neurosis", popular in early 20th Century personality research. It is a non-clinical pervasive tendency to experience negative affective states (Costa & McCrae, 1992a; Watson & Pennebaker, 1989). High N scorers are more likely than low N scorers to experience fear, sadness, embarrassment, anger, guilt and disgust, in many situations. They are prone to all of these states, but the trait itself is a not a "state measure" (state measures are internally reliable but do not have to have test-retest stability). Of the Big Five, N is probably the most "internal". That is, it is not interpersonal (like A) and is more difficult for other people to recognize or observe. In the NEO manual (Costa & McCrae, 1992a, p.50), N has the lowest peer/self correlation of the Big Five measures.

Characteristics of high N scorers

An individual scoring high on N will spend longer completing a questionnaire (because they ruminate over responses), and is more likely to go to the doctor with unexplained symptoms (Costa & McCrae, 1987). They have increased risk of clinical conditions such as anxiety, depression (Costa & McCrae, 1992a) and personality disorder (Austin & Deary, 2002; Matthews, Deary, & Whiteman,

2003). Physical health is also related to N (Daruna, 1996; Denollet, Sys, & Brutsaert, 1995; Drossman et al., 2000; Huovinen et al., 2001; Maier & Smith, 1999; Wilson et al., 2005). N is stable across the life span but declines slightly with age, and is higher in females (McCrae et al., 2002, 1999; McCrae, Costa, Ostendorf, et al., 2000). The causes of N are not known, but serotonin is clearly involved and genetic markers show linkages with anxiety related traits (Kuo et al., 2007; Takano et al., 2007; Van Gestel & Van Broeckhoven, 2003). There is some preliminary evidence that administration of selective serotonin reuptake inhibitors (SSRIs) can lower N scores (Andrews, Parker, & Barrett, 1998; Bagby, Levitan, Kennedy, Levitt, & Joffe, 1999; De Fruyt, Van Leeuwen, Bagby, Rolland, & Rouillon, 2006; Du, Bakish, Ravindran, & Hrdina, 2002; Knutson et al., 1998), although the reasons for this are not known. It could represent a reduction in state related anxiety and depression, which share variance with N.

Neuroticism and health. N has been associated with objectively measured health outcomes (Daruna, 1996; Denollet et al., 1995; Drossman et al., 2000; Huovinen et al., 2001; Maier & Smith, 1999; Wilson et al., 2005). Stress hormones are one potential mechanisms to explain these associations, because N plays a central role in stress reactivity (Bolger & Schilling, 1991; Kendler, Kuhn, & Prescott, 2004; Van Os & Jones, 1999). It sometimes associated with mortality (Korten et al., 1999; Weiss & Costa, 2005), although here the findings are inconsistent. One study has found that the association is removed when adjusted for confounding factors (Shipley, Weiss, Der, Taylor, & Deary, 2007) and several studies have found no association (Almada et al., 1991; Friedman et al., 1993; Huppert & Whittington, 1995; Iwasa et al., 2008; Maier & Smith, 1999). The mixed findings may reflect differences in sampling, methodology and modelling techniques used. However, some authors have argued the association is spurious, because it is detecting subjective health complaints — as a nuisance factor, tapping into "psychologically important but organically spurious variance in physical symptom measures" (Watson & Pennebaker, 1989, p. 248). Costa and McCrae (1987)

agreed that separating "the medical from the psychological effects of Neuroticism is a delicate operation, but one which should benefit both medicine and psychology — and ultimately the public's health" (p. 300).

Facets of N

Anxiety / Anxiety (N1). High scorers on N1 do not necessarily have clinical levels of anxiety but are likely to say that they feel apprehensive, fearful, prone to worry, nervous, tense, jittery. Costa and McCrae (1992a) explain that this scale does not measure phobias but perhaps fears or "free-floating" anxiety (p. 16). Low N1 scorers would be calm, relaxed and tend not to ruminate on the possibility of negative outcomes. Indeed, rumination has recently been proposed as a possible mediator between N and depression.

Anger / Angry Hostility (N2). N2 describes anger and related states. High scorers might describe frustrations and bitterness. N2 reflects experience of anger (anger in) not expression of anger (anger out). Low N2 scorers are easy going and slow to become angry (Costa & McCrae, 1992a).

Depression / Depression (N3). Like N1, N3 is not a clinical scale but correlates with clinical measures of depression (Endler, Denisoff, & Rutherford, 1998). A more appropriate term might be dysphoria, used by some researchers to mean non-clinical depression or generalized depressive affects. Feelings and emotions endorsed by N3 scorers are guilt, sadness, hopelessness, discouragement, dejection, but low scorers are not necessarily cheerful (cheerful is part of E).

Self-Consciousness / Self-Consciousness (N4). N4 scorers endorse items describing shame, embarrassment, being uncomfortable around others, sensitive to ridicule, feelings of inferiority, shyness and social anxiety. N4 is not clinical social anxiety disorder. N4 partly accounts for the correlations which often appear in studies between N and E, but N4 refers to public rather than private self consciousness (Costa & McCrae, 1992a). Low N4 scorers are not necessarily poised or sociable, just less distracted by awareness of others.

Immoderation / Impulsiveness (N5). N5 increases the likelihood of cravings or urges such as food, cigarettes and possessions. These are emotional "crutches" or support systems, later regretted. It is tempting to compare N5 with Impulsiveness (Imp) or Sensation-Seeking (discussed below as part of Zuckerman's model). N5 is different because it refers to spontaneity. Comparability with Impulsiveness (Imp) or Sensation-Seeking (SS) is tempting, but these are not the same as N5. They refer to spontaneity, risk-taking or rapid decision time, depending on the researcher, whereas N5 describes the resistance of urges.

Vulnerability / Vulnerability (N6). High N6 scorers are more likely to feel dependent, hopeless and panic in emergency situations. Vulnerability refers to general vulnerability to stress, the tendency to believe that perceived demands outweigh perceived resources. It can therefore be used as a self-report measure for perceived stress.

2.4.2 Extraversion (E)

Extraverts are assertive, talkative, active, cheerful and look for excitement, stimulation and upbeat situations. High Extraversion (E) scorers do particularly well in enterprizing occupations and as salespeople (Costa & McCrae, 1992a). This trait, Costa and McCrae (1992a) explain, is a construct that broke mental sets (happy-unhappy, friendly-hostile, outgoing-shy), previously thought to be inseparable. This is because introverts are not necessarily unhappy, hostile or shy. They may simply enjoy being alone. They may be reserved, independent and even paced. Costa and McCrae (1992a) urge researchers to think of introversion as being the absence of E, rather than its polar opposite. They argue that the "introspection" construct is similar, but the reflective aspects of introspection tend toward O, at least in the NEO. In his theory of cortical arousal, H. J. Eysenck and Eysenck (1985) proposed that E scorers have lower brain stimulation (cortical arousal) and seek stimulation externally to compensate. E is a risk factor for mortality in some data sets. For example, in the Terman study of gifted children (Friedman et al., 1993, 1995; L. R. Martin et al., 2002), in which cheerful

children died younger. This may occur because it correlates with some forms of risky health behaviours, such as smoking (Patton, Barnes, & Murray, 1993) and unprotected sexual intercourse (Hoyle, Fejfar, & Miller, 2000).

Facets of E

Friendliness / Warmth (E1). Traits can be plotted in a graph, called the interpersonal space, using a technique called multidimensional scaling. Their visual proximity then corresponds to their similarity, or correlation. In the interpersonal space, E1 is very close to A. However, A describes cordiality and heartiness, whereas E1 describes people who are affectionate, friendly and easily form attachments (Costa & McCrae, 1992a). Low scorers are formal, reserved and distant but are not necessarily hostile/uncompassionate. This is a facet about interpersonal intimacy.

Gregariousness / Gregariousness (E2). E2 refers to a preference for others' company, with low E2 scorers preferring their own company.

Assertiveness / Assertiveness (E3). E3 describes dominant, talkative, socially ascendant people who speak without hesitation. These people often become leaders (Costa & McCrae, 1992a). Assertiveness has the potential to be targeted in interventions, as in the Penn Resiliency Program for children (Gillham et al., 2006). This intervention was designed to increase assertiveness, in addition to other interpersonal skills. Low scorers on E3 stay in the background and let others talk or make decisions.

Activity / Activity (E4). High scorers on E4 have a rapid tempo, vigorous movement, a sense of energy and a need to keep busy. Low scorers are leisurely/relaxed but not necessarily lazy (Costa & McCrae, 1992a).

Excitement-Seeking / Excitement-Seeking (E5). E5 is close to Sensation Seeking in the interpersonal space, and describes the tendency to crave stimulation, bright colours and noises. Low scorers do not crave these stimulations. Sensation

Seeking is a widely researched trait in the personality literature (Donohew et al., 2000; Egan, Charlesworth, Richardson, Blair, & McMurran, 2001; Freres, Gillham, Reivich, & Shatte, 2002; Palmgreen, Donohew, Lorch, Hoyle, & Stephenson, 2001).

Cheerfulness / Positive Emotions (E6). Joy, happiness, love, excitement, laughing easily and often, cheerful and optimistic states are more likely to be experienced by high E6 scorers. This is the facet most correlated with happiness. Low scorers are not unhappy automatically, but tend to be less exuberant and high spirited (Costa & McCrae, 1992a).

2.4.3 Openness to Experience (O)

Openness (O) is the least consistent description of traits across different inventories. For example, in the 100-item IPIP it is called "Intellect" (Goldberg, 1999), referring more to intellectual interests rather than mental abilities. Costa and McCrae (1992a) characterize O as O *to experience*. Each of the facets tell us how open respondents are to different *kinds* of experiences: fantasy, feelings, actions, ideas, values. They explain that unconventional ideas and experimentalism may seem odd to others, but experimentation with political, ethical and social ideas does not imply a lack of principles. The individual "may apply his or her evolving value system as conscientiously as a traditionalist does" (Costa & McCrae, 1992a, p. 15). Molecular genetic research has shown that O may have genetic components, particularly for the facets Aesthetics, Actions, and Values (Tochigi et al., 2006). O correlates with creativity but is not, at least in the NEO inventory, a creative ability. In its broadest sense, O refers to curiosity about inner and outer worlds (Costa & McCrae, 1992a). The consequences are that high scorers are keen to experience many things, positive and negative, internally and externally. This might have adaptive or harmful consequences, depending on the situations. Descriptions of O are perhaps inconsistent in the literature because O describes many different things (Costa & McCrae, 1992a).

Facets of O

Imagination / O to Fantasy (O1). This facet refers to vivid imagery, daydreaming and inner worlds. O1 scorers probably have rich and creative lives. Low scorers are more prosaic and focus on the task at hand.

Artistic Interests / O to Aesthetics (O2). Those open to aesthetics like art and beauty, poetry, music. They do not inevitably have talents for these things, just an interest. However, interest may bring knowledge, and therefore skills.

Emotionality / O to Feelings (O3). O3 means being receptive to one's own feelings. This is not the same as being receptive to other's feelings, which would involve notions of emotional intelligence, or perhaps A. High O scorers see emotions as an important part of life, and experience deeper and more differentiated emotional states (Costa & McCrae, 1992a). These feelings might be happy or unhappy — the facet involves *experience*, rather than the type, of feelings.

Adventurousness / O to Actions (O4). O4 allows its high scorers to try new things, go to new places, eat unusual food, try different hobbies. Low scorers stick to the "tried and true" - but perhaps have tried very little to begin with.

Intellect / O to Ideas (O5). High O5 scorers have intellectual curiosity, open minds, try new and unusual ideas, have philosophical discussions, and try brain teasers. They are not necessarily more intelligent but it can carry encourage intellectual potential. This might happen symbiotically, because these new ideas will bring new knowledge. Low scorers are not curious and focus resources on a narrow range of topics (Costa & McCrae, 1992a).

Liberalism / O to Values (O6). Being closed to values describes dogmatism. Honor, traditionalism, conservatism (but not necessarily voting behavior) contrast with individuals who like to reexamine social, political and religious values. It may be a defense mechanism in some form but not defensive as a trait (Costa & McCrae, 1992a).

2.4.4 Agreeableness (A)

Like E, Agreeableness (A) describes interpersonal qualities. These tendencies might involve "moving toward" people or "moving away" (see Horney, 1993; Costa & McCrae, 1992a). High A scorers are sympathetic, altruistic, give and expect help to/from others, and are more popular (moving toward). Disagreeable people are antagonistic, skeptical and competitive (moving away). Disagreeableness is not necessarily maladaptive. Taking an example from Costa and McCrae (1992a), there are several examples of situations where disagreeableness would be an advantage. For example, in the courtroom and the science lab — where skeptical, critical and scientific thinking are required. Low A is related to type A personality, hostility and to cigarette smoking (M. Whiteman, Fowkes, Deary, & Lee, 1997; M. C. Whiteman, Fowkes, & Deary, 1997). Hostility is associated with increased risk of coronary heart disease (CHD), although the size of the association is small (Myrtek, 2001). A correlates negatively with risky health behaviours, especially traffic related risks such as speeding or not wearing a seat belt (Booth-Kewley & Vickers, 1994), probably because of its proximity to hostility in the interpersonal space (Acton & Revelle, 2004).

Facets of A

Trust / Trust (A1). Trust (A1) describes honest and good intentions. Low A1 scores describe cynicism and skepticism, and attributing benevolent intent to others. Again, skepticism can be an adaptive trait in certain contexts.

Morality / Straightforwardness (A2). High A2 scorers are frank, direct and sincere, whereas low scorers are likely to use flattery, deception, crafty, indirect and guarded communication. These qualities may be adaptive in certain situations, such as flirting and seduction. Low scorers would probably describe high A2 scorers as naïve (Costa & McCrae, 1992a, p. 17-18). Historically, A2 was discussed by moral philosophers and overlaps with constructs such as Machiavellianism (a popular construct in the 1970s) and self-monitoring (Costa et al., 1991).

Altruism / Altruism (A3). Altruism (A3) describes concern for others' welfare. Generosity and help contrast with the self centeredness to the low A3 scorer. This facet was historically discussed in social psychology, and covers a kind of "mundane courtesy and consideration" (Costa et al., 1991).

Cooperation / Compliance (A4). Compliance (A4) describes deferring to others, forgiving and forgetting, being meek and mild. This facet is about anger expression. It is an important construct in social psychology, psychiatry and ethology (Costa et al., 1991).

Modesty / Modesty (A5). A5 describes people who are humble, have humility, and who are self-effacing. They are not necessarily low on self-esteem. Neither are they preoccupied with themselves. The opposite of A5 is arrogance — describing conceited, superior and narcissistic traits (Costa & McCrae, 1992a).

Sympathy / Tendermindedness (A6). Low A6 scorers characterize hardheaded people who are not guided by feelings of sympathy. They might describe themselves as rational, using cold logic. They are not easily "moved to pity" (Costa & McCrae, 1992a, p.18)

2.4.5 Conscientiousness (C)

Measures of Conscientiousness (C) describe individual differences in impulse control. Purpose, will and determination allow a high C scorer to manage desires and resist temptation (Costa & McCrae, 1992a; Costa et al., 1991). The causes of C are not known, but this trait has strong developmental aspects. It may reflect the internalization of shared or non-shared environments — parenting style, or cultural guidelines for behaviour. C is necessary for planning, organizing and carrying out tasks (but does not describe mental abilities). In applied settings, C is an important trait. It correlates with task performance, contextual performance and with "adaptive social functioning" (defined as career success, work involvement, marital/relationship stability, health behaviours and longevity). This does not

necessarily mean that people higher in C follow society's rules. They are more likely to engage in politically orientated work such as volunteering, attending political rallies and demonstrations (Roberts, Chernyshenko, Stark, & Goldberg, 2005). Great musicians and athletes need C in order to train effectively (Costa & McCrae, 1992a). Low scorers do not necessarily lack principles, but they may apply them less exactingly (Costa & McCrae, 1992a). C appears to be particularly important for the prediction of health behaviours (Bogg & Roberts, 2004).

A recent addition to the big five. As a relatively recent addition to the big five, C has received less research (c.f. Friedman et al., 1993, 1995). Compared to the facets of N, which have clinical significance (e.g. anxiety and depression), the facets of C are not fully validated. In several of the leading personality-health data sets (e.g., Gow, Whiteman, Pattie, & Deary, 2005; Hampson, Andrews, Barckley, Lichtenstein, & Lee, 2006; L. R. Martin et al., 2002; M. C. Whiteman, Bedford, Grant, Fowkes, & Deary, 2001), general measures of C were recorded, rather than traits *and* facets (c.f. Goldberg, 1999). These measures do not have complete coverage of C (Roberts et al., 2005). Roberts et al. (2005) considered the convergent, discriminant and incremental validities of seven inventories which had C items. Proactive C involves achievement and commitment to work, inhibitive C involves moral scrupulousness and cautiousness. They concluded that C seems to describe two components: proactive and inhibitive, which are similar to the "directions" described by Costa et al. (1991). However, further research is needed in order to clarify this model. It is likely that within the next few years, attention will switch to confirmatory factor analysis rather than exploratory factor analysis, to clarify the number and nature of C facets (Roberts et al., 2005). One way to contribute to the validation of C facets is to include them in research, find out if they have any specific association with health criteria and share the findings with the research community.

Problems with C

There is a lack of agreement on the lower order taxonomy of C, partly due to its recent arrival in the literature. Further down in the hierarchy, there is less agreement. Roberts et al. (2005) found that researchers were using different facets of C in their work. These included variously: achievement, dependability, impulse control, order, morality, persistence, order, industriousness, religiousness and decisiveness.

Facets of C

Self-Efficacy / Competence (C1). Competence (C1) describes capable, sensible, prudent, effective traits that allow people to be well prepared for life. It is related to Self Esteem (SE) and internal locus of control. Low scorers are often unprepared. It is noteworthy that self reported mental ability correlates with C1. This means that people scoring high on C1 believe themselves to have higher mental ability (Costa et al., 1991).

Orderliness / Order (C2). Order (C2) increases the likelihood of having a neat, tidy, well organized environment, where things are kept in their proper places. Too much C3 is similar in content to compulsive personality disorder. Low scorers find it difficult to be organized.

Dutifulness / Dutifulness (C3). Dutifulness (C3) is the facet most related to ethical principles and moral obligations. If you are undependable and unreliable, you are probably a low C1 scorer. This is related to standards of conduct and Freudian concept's of ego strength, but not necessarily to moral reasoning or the origin/sophistication of moral principles (Costa & McCrae, 1992a). Developmental studies, showing whether C3 reflects internalization of the rules of authority figures, would be informative.

Achievement-Striving / Achievement Striving (C4). Achievement-Striving (C4) describes high aspirations, excellence, working hard to achieve goals, diligence,

purpose, sense of direction in life (Costa & McCrae, 1992a). Low scorers are lackadaisical, lazy and aimless. This does not, however, mean that they are unhappy. Excessive C4 is related to Type A personality, because it describes excessive striving for achievement (Costa et al., 1991).

Self-Discipline / Self-Discipline (C5). Self Discipline (C5) is needed to complete tasks. There are always distractions and the potential for boredom, but C5 protects against distractions and procrastination. Empirically, Impulsiveness is distinct from C5 but clearly they share some of the same hallmarks. To clarify the difference between them, Costa et al. (1991) explained that low C5 scorers cannot make themselves do what they *do* want to do (in the long run). High Impulsiveness scorers cannot resist doing what they *do not* want to do. This is subtle, but important, distinction. Self control is also different to C5. C5 is not the same as self control, which is related to N. C5 is more proactive, concerning perseverance unappealing tasks (Costa et al., 1991).

Cautiousness / Deliberation (C6). Deliberation (C6) involves thinking carefully before acting, being cautious and deliberate. Low scorers are hasty, but the ability to make spontaneous and snap decisions is useful in many circumstances.

2.4.6 Summary: Personality facets of the big five

In summary, the six facets of the Big Five in the NEO and the IPIP-NEO provide additional information for each trait. They have convergent and divergent validity (Costa & McCrae, 1992a, p. 47), and their own criterion validity. The big five model provides "a common psychometric tongue" (Saucier & Goldberg, 2002, p. 13). However, the validity of both traits and facets can benefit from further research: "the Big Five factors are far from definitive, and the derived assessment instruments deserve constant attention and an open eye for new facets and features to be included, in the model as well as in its assessment" (de Raad & Goldberg, 2002, p. 18). Therefore, although I have chosen to use the NEO-PI-R

and IPIP NEO in this thesis, this should not imply that they are perfect instruments, nor that their are no critics of the big five model. I turn now to some of the criticisms that have been levied at the big five model, and at trait research more generally.

2.4.7 Critics of the Big Five

Block (1995) is a strong critic of the big five model. He argues that the although the big five are claimed to be unrelated, they are frequently correlated. Taking just one example, a correlation of -.40 was observed between N and E in a British study (Egan, Deary, & Austin, 2000). This is a large correlation, given that the two dimensions should be uncorrelated. Other researchers are not critical of the big five *per se*, but argue that other traits should be added. For example, honesty or humility (Lee, Ogunfowora, & Ashton, 2005; Ashton & Lee, 2005b), attractiveness (valence) and religiosity or spirituality (Goldberg, 1998), masculinity and femininity (Paunonen & Jackson, 2000) are not included in the five factor model, but are clearly important sources of individual differences. However, there is less supporting evidence for these constructs as personality traits, when compared to the nomological network surrounding the big five, and there are problems in finding a sufficient number adjectives or items for some of these constructs (Paunonen & Jackson, 2000). Although more traits could be added (Goldberg, 1998; Paunonen & Jackson, 2000), there is at least consensus about what the first five are (Gow et al., 2005). H. J. Eysenck (1992) argued that the big five lacked a strong nomological network. However, Psychoticism, his own construct, can be represented as low A and low C (McCrae & Costa, 1985). Therefore, the big five can incorporate other nomological networks.

Critics of trait research

There are many critics of trait research more generally. A complete review is beyond the scope of this chapter, but it is worth noting why these criticisms are not relevant in this context:

Correlation size. It has been argued that correlations between traits and behaviour are not very high (Mischel, 1968; Gould, 1996), and this would support the view that situations, rather than traits, are better predictors of behaviour. In personality-health research, it is health criteria, and biomarkers (discussed below), that are correlated with traits. Whether the correlation is with behaviour or health criteria, a small to moderate size of association (effect size), is important. Very often, the correlations observed are larger than recognized risk factors for health (Bogg & Roberts, 2004).

Consequences. Commentators have argued that individual differences research fails to appreciate political factors and the potential for objectionable applications of its findings. Referring to mental ability measures, Deary (2000) has responded that "they may be put to humanitarian or misanthropic purposes" (p. 25). Separating research findings from the possible ways in which they might be used is delicate matter for all researchers, with no easy solution. However, individual differences is a broad field and many personality-health researchers are committed to understanding health inequalities. Increasingly, personality-health researchers are including social determinants of health in their models (see chapter seven).

Tautology. Critics have argued that traits are similar to the behaviour they claim to predict, and are therefore tautological or circular explanations of personality (Butt, 2004). The solution is to look for validity external to traits themselves (Deary & Hettema, 1993).

Utility. Finally, some people argue that traits have no applied utility. This scepticism has also occurred in disciplines aligned with personality-health research, such as psychosomatic medicine. One response was that the effect size between psychological factors and health "is so low it has as yet no practical meaning for prevention and prediction purposes" in medical settings (Myrtek, 2001, p. 245). This response is moot — the effect sizes are comparable to other risk factors for illness (Bogg & Roberts, 2004), and are larger in several cases (Du et al., 2002; Piedmont, 2001). Another argument is that trait-health associations have "too much face validity" (Stansfeld,

2002b, p. 113), meaning that researchers are too quick to accept personality as a predictor of health outcomes, without the supporting evidence base. These criticisms are particularly important for personality-health research, where engaging with what R. Hogan (2005) called the "consumers" (R. Hogan, 2005) of trait research is important. I return to this issue in the concluding chapter.

2.5 Psychobiology, mechanisms and biomarkers

Biological variables play an important role in the validation of the big five, and in theories of personality and health. The aim of this chapter is not to review theories about the psychological basis of traits, because this is a vast research area (H. J. Eysenck & Eysenck, 1985; H. Eysenck, 2006; Matthews, Deary, & Whiteman, 2003; Stelmack, 2004). The range of measures studied includes: electroencephalograms (EEGs, a measure designed to reflect a construct called cortical arousal), electrodermal response, stimulated salivation, hormones, the immune system, neurotransmitters, sedation threshold, pupil dilation, body types, cortisol and blood groups (see, e.g. H. Eysenck, 2006). Every biological system, from the brain to the blood, could potentially be correlated with personality traits. This clearly covers a very large range of variables. H. Eysenck (2006) argued that the field could be characterized as a set of "anomalies, failures to replicate, and areas with insufficient data" (p. 270). Matthews, Deary, and Whiteman (2003) observed that the biological literature has suffered from an over-emphasis on Eysenck's arousal construct, a theory that E was related to a need for stimulation. More recently, the stress moderation model of personality-health (T. Smith, 2006) has provided a theoretical framework for how stress processes might influence health. This suggests that there might be more substantive and replicable findings, if more research is conducted on the correlation between the endocrine (hormone) system, and traits relevant to stress perception, such as N. This will become important focal point for the cortisol research described later in this thesis.

2.5.1 Biomarkers

Biological variables play an important role in trait research, as biomarkers. The term biomarker refers to: "A substance used as an indicator of the presence of material of biological origin, of a specific organism, or a physiological condition or process; spec. a diagnostic indicator of (predisposition to) a medical condition" (Soanes & Stevenson, 2005). Synonymous with "marker", the term in psychobiology has been used to describe a biological measure that is thought to reflect an underlying biological process, or latent variable, such as stress (Dowd & Goldman, 2006), or aging (Bekaert, De Meyer, & Van Oostveldt, 2005). One measure could be a biomarker for more than one process. For example, the stress hormone cortisol, depending on how it is measured, could be a biomarker for stress (Clow et al., 2004)or for disease[1]. If the biomarkers of traits are discovered, then the genes and environments causing those biomarkers can be more easily investigated (Van Gestel & Van Broeckhoven, 2003). Finding reliable and replicable biomarkers of traits is an important first step, before biological explanations for traits are sought (Matthews, Deary, & Whiteman, 2003). A correlation between a trait and a biomarker alone is not an explanation, because causality cannot be inferred from a correlation (Bhopal, 2002). However, correlations do have ultimate causes (Gottfredson & Deary, 2004), and finding out why correlations occur is part of understanding how an association might work (Abelson, 1995).

2.5.2 Biomarkers and descriptive inventories

It is not the case that biomarkers are more relevant for trait inventories that were developed using psychobiological theories. Two popular inventories that emphasize biological theories, are the Tridimensional Personality Questionnaire (TPQ, Cloninger, Przybeck, & Svrakic, 1991), the Eysenck Personality Inventory (EPI, H. Eysenck & Eysenck, 1964) and the Eysenck Personality Questionnaire (EPQ, H. J. Eysenck & Eysenck, 1975). Descriptive and taxonomic inventories (e.g. Goldberg, 1999; Costa & McCrae, 1992a), whether developed using either

[1] http://tinyurl.com/39atln

questionnaire and lexical methods, benefit equally from validation in the form of correlations with biomarkers. The main point is that biomarkers are evidence of external, predictive validity (Deary & Hettema, 1993). There is no necessary relationship between an inventory and the potential for biological associations, not the size of those association, because "Personality inventories are not personality theories" (Matthews, Deary, & Whiteman, 2003, p. 25). Arguably, biomarkers that correlate with descriptive inventories are *more* useful to trait theories. They help to validate the descriptive inventories and lexical hypothesis, which otherwise might be accused of lacking theoretical grounding (e.g. Block, 1995). Eysenck said that "Even a bad map is better than no map at all' (H. Eysenck, 2006, p. xi) but argued that researchers needed to make their models more realistic (p. xii) with reference to biology. Authors of descriptive inventories do often believe that their traits are biologically underpinned (Matthews, Deary, & Whiteman, 2003), but are cautious to state that biomarkers are putative (Costa & McCrae, 1992a). Biomarkers are a first step toward psychobiological explanations for traits (H. Eysenck, 2006).

Description and explanation Description is not the same as explanation (T. Smith, 2006). The big five is a measurement model, describing how facets and traits are organized into a descriptive hierarchy. Explanations of traits, ultimately, require structural models. Structural models consider how facets and traits are related other variables, sometimes called "explanatory variables", which would include biomarkers, as well as genes, environments, and social processes. The question "what are traits" is different to "how many traits are there"? Why, rather than whether, individuals differ, is important to consider because this may ultimately help explain why traits are related to health and illness. The descriptive theories, then, currently provide us with an initial "map" of traits, but the map may not correspond to the biological explanations. The big five is like a bridge between two separate islands. Without it, descriptive theories would make little connection with biological theories. It therefore allows collaboration between

two separate strands of research. The big five is a useful benchmarking, or reference point, tool (de Raad & Goldberg, 2002; T. Smith, 2006), widely used and able to describe the traits from most other inventories. A key advantage of the NEO-PI-R is that it has two additional traits (A and C) and six facets for all five traits. It is therefore a good compromise between fully describing and beginning to explain traits with reference to psychobiological variables.

2.5.3 Genetics

A review of current genetic evidence for traits is beyond the scope of this thesis. However, it is worth noting that genetic studies are providing clues about the types of biomarkers that could be related to traits, particularly N. Behavioral genetics considers how traits and explained by genes and environments (W. Johnson & Krueger, 2005). The term "genetic" does not imply that genes are more important, but the best way to determine the contribution of environment is to determine the contribution of genes — the remainder is environment. This is different from molecular genetics, which shows how individual genes may be involved in traits (Van Gestel & Van Broeckhoven, 2003). Van Gestel and Van Broeckhoven (2003) argued that there was a "gold rush fever" to find specific genes for personality, when the molecular genetic technology became available. However, is it only traits related to anxiety, negative affects and avoidance that seem to have been consistently replicated in the literature (Ebstein, 2006; Munafo et al., 2003). Since these are part of the N construct, there are strong reasons to suspect that biomarkers for N are most related to genetic and psychobiological variables. This is because specific genes are related to neurotransmitters such as serotonin (Ebstein, 2006; Takano et al., 2007) and stress related hormones such as cortisol (Wasserman et al., 2007). Several gene regions appear to be involved in N (see, for example, Van Gestel & Van Broeckhoven, 2003).

2.5.4 Genetics and facets

A problem for genetic trait studies is the existence of personality facets (Van Gestel & Van Broeckhoven, 2003). Facets were not developed empirically, using factor analysis, but to represent comprehensively summarize the proliferation of trait terms in the literature. Facets, "although constructed to be homogenous on a psychological level, represent a sum of traits with a different biological background" (Van Gestel & Van Broeckhoven, 2003, p. 845). Furthermore, the methods used in molecular genetic studies assume that quantitative traits are normally distributed and "Scores on facets are mostly far from normally distributed" (p. 845). A further problem with genetic studies for facets is the lack of uniformity in the phenotype and in modeling. The choice of personality inventory, for example, is "not based on rational grounds, but on emotional or personal preferences...such as degree of familiarity with a certain questionnaire or its designer" (p. 846). The big five model may help provide such uniformity. However, the fact that facets such as Anxiety and Depression are organized into N at the phenotype level, this does not imply that they are related at the genotype level (Van Gestel & Van Broeckhoven, 2003). For example, anxiety and depression might result from N, and psychobiological processes, rather comprise part of the N construct (Takano et al., 2007). Traits and facets may have different degrees of causal primacy.

2.5.5 Conclusion

This chapter has introduced the big five model of personality traits, and two instruments designed to measure them: the NEO-PI-R and the IPIP NEO. The reliability and validity for these measures has been described. The central claim of this chapter is that the measurement model for the big five is sufficiently validated, so that it can provide a benchmark across many different research studies (de Raad & Goldberg, 2002). For this reason, it would not be appropriate to conduct exploratory factor analysis on data obtained from these instruments, in an

attempt to find new traits or variations on the big five traits. This would be counterproductive, because it would not allow the research to be compared with the wider nomological network, where trait researchers use the big five to speak in "a common psychometric tongue" (Saucier & Goldberg, 2002, p. 13). This is particularly important for the cortisol research described later in this thesis, where difficulty emerges in trying to compare correlations with traits across a large, inter-disciplinary literature.

CHAPTER 3

Measuring Health Behaviours

The first step in influencing health behaviours in any group is to understand why people make the choices that they do (Home Office, 2004).

3.1 Introduction

This is the second of three introductory chapters concerning measurement in personality-health models. Health behaviours are the focus of this chapter. Health behaviours could mediate the association between traits and health outcomes. That is, they could a candidate mechanism that transmits the effect of traits on health, via health behaviours (Hampson, Goldberg, et al., 2007). Traits could influence health behaviours, which influence health in turn. This is a key assumption made in the health behaviour model of personality and health (T. Smith, 2006; Wiebe & Fortenberry, 2007). In the first section, I introduce health behaviours, and their relationship to health outcomes such as morbidity and mortality. I make particular reference to the "big four" health behaviours (Pronk, Peek, & Goldstein, 2004), which are smoking, alcohol, activity and diet. The next section describes how health multiple behaviours can be measured. Unlike personality traits (chapter two) and health outcomes (chapter three), there is no widely agreed measurement model for multiple health behaviours. The Health Behaviour Checklist (HBC, R. R. Vickers et al., 1990) has been adopted by some

trait researchers (e.g., Booth-Kewley & Vickers, 1994; Wasylkiw & Fekken, 2002), but is not very popular. Finally, I conclude by arguing that there is a clear need to address the limitations of the HBC, by combining this instrument with additional measures of other health behaviours, such as cigarette smoking and alcohol use.

3.2 What are health behaviours?

Health behaviours account for an large proportion of worldwide disease (World Health Organization, 2002). Much of the variation in disease susceptibility is genetic or environmentally determined, but a good proportion of it is behavior, and behaviour can be changed (Brunner & Marmot, 2006). Healthy behaviours can lower the risk of premature mortality and of morbidity. Healthy behaviours are defined as (a) behaviours that reduce the risk of overtaxing the body's adaptive capacity; (b) behaviour that involve reducing risk-taking; (c) behaviours that should help prevent the onset of illness; (d) behaviours that could improve health rather than merely prevent illness (R. R. Vickers et al., 1990). In the literature, these are usually described as "health behaviours", with unhealthy behaviours described as "risky health behaviours". However, "health behaviours" can refer to both, without specifying the direction. Pronk et al. (2004) define the "big four" (p.5) unhealthy behaviours, as: smoking, alcohol consumption, inactivity, and poor diet.

3.2.1 Measuring health behaviours

Health behaviours are constructs, because they cannot normally be measured directly. Ideally, health behaviours would be measured by taking a combination of biometric measures and self reports. For example, self reported smoking status could be verified using saliva, tested to confirm the presence of nicotine (salivary cotinine, Caraballo, Giovino, Pechacek, & Mowery, 2001). Unfortunately, these methods are expensive and intrusive. Most researchers rely on self reported health behaviours. These share many of the same strengths and weaknesses as other self reports (e.g. personality traits). For example, they have the advantage

of being cheap, easy to administer, and can be validated against other measures. They have the disadvantage of being error-prone, with the possibility of response sets (e.g. socially desirable responding) occurring, and the potential for memory problems as people try to recall their health behaviours. It should not be assumed that biometric measures are necessarily better than self reports, because both have sources of error (P. R. Martin, 1998). Later in the chapter, different methods for measuring health behaviours will be considered.

3.2.2 The big four

The big four contribute substantially to the leading causes of death in developed countries. They appear in the policy documents of governments and health organizations worldwide (Murray & Lopez, 2006) and are the targets for intervention, in primary care and health promotion settings (Home Office, 2004; Kypri & McAnally, 2005). Modifying health behaviours can increase life expectancy (Doll et al., 2004) and lower the risks of disease. The relationship between health behaviours and mortality is causal and cumulative (Belloc & Breslow, 1972; B. M. Brock, Haefner, & Noble, 1988; Wingard, Berkman, & Brand, 1982). This means that a higher number of unhealthy behaviours increases the likelihood of poor health status. Additively scoring health behaviours increases the size of the association with morbidity (Metzner, Carman, & House, 1983; Segovia, Bartlett, & Edwards, 1991) and mortality (Breslow & Enstrom, 1980a). There is evidence that multiple unhealthy behaviours can have interactive or synergistic effects on risk of disease (Luoto, Prättälä, Uutela, & Puska, 1998; Meng, Maskarinec, Lee, & Kolonel, 1999), because their combined risk is greater than when the scores are added together (Laaksonen, Prattala, Helasoja, Uutela, & Lahelma, 2003). Two of the key diseases link health behaviours, particularly the big four, to mortality, are cancer and cardiovascular disease.

3.2.3 The big four and mortality

The big four are associated with all cause and cause specific mortality (Bonow, 2002; Mokdad, Marks, Stroup, & Gerberding, 2004; Pronk et al., 2004). They account for up to 50% of causes of death in developed countries (Mokdad et al., 2004). There are many other causes of death and disease, including genetics, environment and external causes such as accidents. It is important to note that health behaviours are not the only determinants of health and illness. However, the big four are strongly related to increased risk of death. This is because they are associated with diseases of the heart or blood vessels (Bonow, 2002). Cardiovascular disease, or CVD, is a general term of these diseases. The term coronary heart disease (CHD) is used to describe cardiovascular diseases specifically of the heart. There are increased risks of cardiovascular diseases associated with smoking (Ambrose & Barua, 2004), alcohol (O'Keefe, Bybee, & Lavie, 2007), lack of exercise (Hamer, 2006) and poor diet (Kuller, 2006). The big four health behaviours are also associated with risk of cancer. Cancer is the second most common cause of death. Each of the big four increases the likelihood of developing cancer: smoking (Doll et al., 2004), alcohol (Kloner & Rezkalla, 2007), lack of exercise (Warburton, Nicol, & Bredin, 2006) and diet (Holmes, 2006). Many of these associations can be interpreted as causal, because they meet the epidemiological criteria for judgement (see Lilienfeld & Stolley, 1994). The criteria for judgement can be interpreted as the nomological network of evidence supporting a causal association. The term originates from epidemiology, where associations are assessed for their: (a) consistency, (b) strength, (c) specificity, (d) temporal relationship, (e) coherence. Bhopal (2002) adds (d) experimental evidence and (e) biological plausibility to this list.

Smoking

Cigarette and tobacco smoking are the most "avoidable cause of death and disability in developed countries" (R. Edwards, 2004, p. 217). The association between cigarette smoking and increased risk of mortality, is a well established as-

sociation (Home Office, 2004; Doll et al., 2004; World Health Organization, 2002). However, smoking is not evenly distributed throughout the population. Those with lower SES are more likely to smoke, and to smoke more (Jarvis & Wardle, 2006). Other variables linked with increased likelihood of smoking include male sex[1], lower social class (Jarvis & Wardle, 2006), ethnic minority status except Chinese (Home Office, 1999)[2], younger age, non-heterosexual orientation (H. Ryan, Wortley, Easton, Pederson, & Greenwood, 2001), lower social support (Fisher, 1997) and personality traits such as high hostility (M. Whiteman, Deary, Lee, & Fowkes, 1997) or low self esteem (Glendinning, 2002). Smoking is strongly associated with socio-economic classification, being far more common among those in routine and manual occupational groups than those in managerial and professional groups. The longitudinal aspects of smoking status are not well understood. In a recent longitudinal study of personality and health, childhood Extraversion, Agreeableness and educational attainment predicted adult smoking (Hampson, Goldberg, Vogt, & Dubanoski, 2006).

Self reports and smoking. There is evidence that self-reports of smoking are a reliable indicator of actual smoking in population-based surveys, at least in medical settings (Caraballo et al., 2001; Caraballo, Giovino, & Pechacek, 2004). This validation technique is based on analysis of salivary cotinine, an objective measure of nicotine exposure. However, the error associated with self reports of smoking is not randomly distributed throughout the population. Certain groups, such as pregnant women, or those with a disease related to smoking, are more likely to report non-smoking status if they do smoke (Caraballo et al., 2001, 2004).

[1]Prevalence in 2005 was highest in 20-24 category according to results from the 2005 General Household Survey `http://www.ons.gov.uk/ghs`

[2]The odds of smoking are higher for women in social classes I to IV, but not V. This is an example of effect modification, because smoking status predicts smoking more strongly for women than for me. However, smoking is more prevalent for men, overall. In 2003, 28% of men smoked, compared to 25% of women. This dropped to 28% for men and 25% for women in 2004(Home Office, 2004).

Mechanisms linking smoking with mortality

As discussed above, smoking reduces life expectancy (Doll et al., 2004), and causes disability (R. Edwards, 2004), but what are the mechanisms? Tobacco smoke does not cause cancer because of nicotine — it is the toxins such as benzopyrene release by any burnt material that is the mechanism linking smoking to cancer. Oxidation produces as epoxide which binds to DNA and distorts it. Therefore, DNA damage is the cause of cancer (Paz-Elizur et al., 2003). Smokers often report that smoking is beneficial for stress levels, but this is probably due to a combination of the mood stabilizing effects of nicotine and trying to avoid the negative effects of nicotine withdrawal (Jarvis & Wardle, 2006) — that is, a return to baseline levels of mood. However, when Neuroticism is controlled for, there is a positive relationship between nicotine and good mood (Kalman, 2002). Interventions designed to stop people from smoking are generally unsuccessful. Most interventions have a success rate of just 10 to 25% after six to 12 months (Lipkus, Barefoot, Williams, & Siegler, 1994). This is because nicotine is a highly addictive drug, rewarding the smoker with dopamine, which creates the subjective experience of pleasure (Dani, Ji, & Zhou, 2001).

Social patterning of smoking. Smoking is socially patterned, despite being an individual behaviour (Jarvis & Wardle, 2006). Smoking prevalence is graded by occupational class, and by education, showing that health behaviours are just a matter of individual responsibility but occur in a "broad social context" (p. 240). These authors argue that individual level explanations fail to explain why disadvantaged people are drawn to these behaviours. Nicotine is powerfully addictive, and the poor may be more drawn to the drug because (1) there are higher rates of smoking initiation among the poor; (2) reward effects could be stronger, either positive or negative, perceived or actual; (3) greater difficulties in smoking cessation, through lower motivation, higher dependence, or fewer available coping resources (p. 230). The popular explanation that smoking ameliorates stress is not supported by the empirical evidence. Smoking cessation results in higher, not lower, levels of perceived stress. There is no evidence that smoking

alters mood other than from withdrawal relief. Smoking is therefore a stressor, not a stress reliever (Jarvis & Wardle, 2006). In summary, researchers, particularly those interested in individual differences, need to be aware that health behaviours such as smoking are socially patterned, not just a matter of individual choice. Therefore, the health behaviour model in personality-health research may benefit from the inclusion of SES.

Alcohol

Alcohol can cause high blood pressure, cancer and cirrhosis of the liver. It also increases the risk of mouth and stomach cancer. It contributes to traffic related accidents and to domestic violence. But unlike smoking, there are clearly demonstrated benefits to drinking alcohol in moderation (defined as 3-4 units per day) (Home Office, 2004; Rimm, Klatsky, Grobbee, & Stampfer, 1996; World Health Organization, 2002; Peele, 1999; Renaud & Lorgeril, 1992). Binge drinking is defined as drinking twice the recommended daily allowance (Home Office, 2004). However, there are social benefits to alcohol use, because moderate drinkers enjoy larger social networks (Peele, 1999). Red wine has antioxidant cancer-protective properties and reduces the likelihood of coronary heart disease via platelet activity; a mechanisms thought to explain the "French paradox" in which red wine drinkers live longer (Renaud & Lorgeril, 1992). Beer contains B-vitamins (Bamforth, 2002). Alcohol drinkers report that it helps them to relax (Peele, 1999). Alcohol, unlike smoking, has a different type of relationship to SES. In women, the number consuming greater than the recommended daily allowance of alcohol is highest in managerial and professional occupations, rather than routine and manual occupations (Home Office, 2004; Laaksonen et al., 2003). There is a U-shaped relationship between alcohol and all cause mortality (Duffy, 1995), so that moderate drinkers (in the U.K., 12.9 units of alcohol per week) have decreased risk of mortality (in the 50 and 80 year age range), when compared to non-drinkers or heavy drinkers (White, 1999).

Self reports and alcohol. People may find self-reporting their alcohol consumption difficult. There are several different kinds of alcoholic drinks, and several different sizes of measures and glasses. Wine glasses can be refilled by others, and some people drink beer from pitchers (Home Office, 2004; Del Boca & Darkes, 2003). Drinking may occur at different times, and not form a regular pattern, as might be the case in occasional binge drinking. To address these kinds of problems, researchers have experimented with two general approaches to self reports. In the first, estimates of quantity or frequency are sought. This is popular and straightforward, but Del Boca and Darkes (2003) argue that these types of self reports provide *modal* not *average* consumption. That is, people report the most typical drinking frequency, but not the mean level. For example, they are responding in terms of "the usual amount I drink on Fridays". In the second, retrospective and prospective daily estimation is sought. This provides data on actual drinking events, but is more time consuming for the respondent. Using this method, it is possible to ascertain total alcohol consumption over time, mean drinks per drinking session and percentage heavy drinking within a specified time period. This means that the data is suitable for a variety of modeling strategies, including latent growth curve modeling, traditional general linear modeling and survival analysis.

Activity

Higher levels of physical activity are associated with reduced likelihood of CVD, hypertension, non-insulin dependent diabetes, osteoporosis, falling, colon and other cancers, anxiety and depression (Warburton et al., 2006; Home Office, 2004; World Health Organization, 2002). It is linked to all-cause (Batty, Shipley, Marmot, & Smith, 2003) and cause-specific (Batty et al., 2003; D. G. Smith, Shipley, Batty, Morris, & Marmot, 2000) mortality. It also has a dose response relationship with quality of life and independent living in older adults (Spirduso & Cronin, 2001). A dose-response relationship means that higher activity results in correspondingly higher levels of these variables. This is one of the criteria for assessing causality in epidemiology (Armenian & Shapiro, 1998; Bhopal, 2002). Mod-

erate exercises includes brisk walking, gardening, cycling and swimming (Pate et al., 1995). Vigorous activity includes running, weight training and aerobics. Sports and activities are usually defined as those lasting 20 minutes or longer (Boniface & Tefft, 1997). Too much activity can have harmful effects, such as physical exertion. There is increased risk of colds after extreme physical exercise, such as marathon running or skiing (Douglas, Hemila, Chalker, & Treacy, 2007).

Mechanisms linking activity with mortality. The mechanisms linking activity to health are not well understood. This is partly because cause-specific mortality has cause-specific mechanisms (Batty, Shipley, Marmot, & Smith, 2001). Currently, theories about why activity is protective against mortality are underdeveloped. One hypothesis (for cancer) is that physical activity improves cardiovascular risk factors; another is that is reduces other primary risk factors such as smoking (e.g. Taylor, Unal, Critchley, & Capewell, 2006). Batty et al. (2001) acknowledge that their "simplistic measure of physical activity may, in part, explain the weak associations seen" (p. 863). Indeed, activity is very difficult to measure. Questionnaires are often the only feasible way to assess activity. However, few of the available questionnaires have been validated (Powell & Pratt, 1996). Like nutrition (discussed below), self reports of activity contain large amounts of error variance. Activities may not be salient for respondents, or may not be routine. Unlike smoking and alcohol, there is no biological marker for recent activity. Cardio-respiratory fitness measures, such as forced expiatory volume (FEV) in one second (e.g., Cox, 1988), are useful, but are influenced by factors other than activity.

Nutrition

Nutrition is a risk factor for morbidity and mortality. For example, diet accounts for up to 80% of the cancers of the large bowel, breast and prostate (Cummings &

Bingham, 1998). These may be preventable by dietary change. It is useful to classify food groups into *protective* and *risky*. Protective foods include complex carbohydrates (bread, cereals, potatoes), whole grain cereals, fruit and vegetables. At least five portions of fruit and vegetables are recommended (Home Office, 2004). Vegetables protect against cancers of the large bowel, while fruits protect against cancers of the stomach (Cummings & Bingham, 1998). Risky foods include salt, sugar and saturated fat (Robertson, Brunner, & Sheiham, 2006). There is increasing interest in the dangers associated with high sugar diets (Robertson et al., 2006). Red and processed meats are generally considered to be risky (Cummings & Bingham, 1998). The mechanisms linking nutrition with mortality are complex, but include factors such as antioxidants and apoptosis. Apoptosis is a programmed form of cell death, which protect against cancer by suppressing mutations in DNA. It may be possible, in the future, to use DNA damage as a biomarker for poor nutrition (Fenech, Shyong, & Gillies, 2007). In contrast to broad food groups, the relationship between specific food items and mortality, is complex, and less well understood (Cummings & Bingham, 1998; Fenech et al., 2007; Hales & Barker, 2001).

Measuring nutrition. As with activity, nutrition is very difficult to measure. It is also difficult to model. There are many food-disease associations, and even more mechanisms underlying each of these associations (Ness & Powles, 1997). Self reports are prone to error because portion sizes are variable between people. The questionnaires tend to be lengthy and memory problems can prevent participants from recalling accurately what they ate. Several food questionnaires have been shown to be invalid and there is no "gold standard" (unlike salivary cotinine for smoking, for example, Masson et al., 2003) from which to validate self reports of nutrition. However, Masson et al. (2003) did successfully correlate a food frequency questionnaire with a variety of measurable indicators, including saturated fat, alcohol, iron, sugar, starch, vitamin E, vitamin C and zinc. All of the correlations were above .5, showing that self-reports of nutrition are valid indicators of nutrient intake. Dietary fat can be estimated using a food frequency

questionnaire, and matched using knowledge about the fat content of foods (Tefft & Boniface, 2000). Food diaries with 24-hour recall are popular in the literature.

Relationship of activity and nutrition to obesity. Activity and nutrition combine to create an association with obesity. The risks associated with poor nutrition can be reduced by increasing activity. Conversely, the risks associated with low activity can be reduced by increasing good nutrition. This is because activity and nutrition operate through many of the same causal pathways, in their influence on CVD (Ignarro, Balestrieri, & Napoli, 2007). Obesity is often conceptualized as a health outcome, it is not a cause of death. Obesity is independently linked to heart disease and to diabetes, which increase the risk of premature death (Home Office, 2004). Obesity is defined as a body mass index of 30 or higher (weight in kilograms divided by height in metres squared) but waist-hip ratio (ratio of girth of the hips to waist) is a better predictor of all cause mortality (Bigaard et al., 2004).

3.3 Multiple health behaviours

Health behaviours co-occur (Belloc & Breslow, 1972; R. R. Vickers et al., 1990; Chiolero, Wietlisbach, Ruffieux, Paccaud, & Cornuz, 2006) and not in a random way. For example, smokers are more likely drink alcohol, and not to exercise. Finding out the prevalence of multiple risk factors is an important part of the descriptive epidemiology of health behaviours. If health behaviours co-occur and group together consistently, then epidemiological research and health promotion interventions might be guided usefully by this knowledge. It may alert researchers to previously unknown confounding variables (e.g. if smokers also tend to drink), or allow the more efficient targeting of groups of health behaviours rather than single ones (Donovan & Jessor, 1985; Atkins & Clancy, 2004). The causes of co-occurring health behaviours is an important but separate issue. It may also be important to find out which subgroups of the population have high prevalence for multiple health behaviours. Unfortunately, population based data on multiple health behaviours is rare. There is a vast literature on

health behaviours, unhealthy lifestyles, problem behaviours in adolescence, but not on the causes of multiple behaviours.

3.3.1 Choice of single versus multiple behaviours

It is not always appropriate to study multiple health behaviours. The focus of a study might be on one or more health behaviours, rather than a comprehensive understanding of health risks. Some health behaviours are important in their own right because they are empirically related to specific health outcomes (Tefft & Boniface, 2000; Bigaard et al., 2004; Home Office, 2004; Hampson, Goldberg, et al., 2007; Ignarro et al., 2007; Wasylkiw & Fekken, 1999, 2002). Some health behaviours are maintained by different mechanisms to others, so one will not necessarily lead to others. Interventions designed to change a health behaviour may need to be targeted at that behaviour only (Wasylkiw & Fekken, 1999; Stefansdottir & Vilhjalmsson, 2007).

3.3.2 Mechanisms linking multiple health behaviours with mortality

There is emerging evidence that multiple health behaviours have multiplicative (synergistic) associations with disease risk (Breslow & Enstrom, 1980b; Chiolero et al., 2006). Synergism means that the association is interactive. The risk from one variable depends on the level of the other variable (a statistical interaction; Miles & Shevlin, 2000), so that the net effect can be greater than the sum of the separate effects[3]. Nutrition and activity, for example, have synergistic effects on CVD risk (Ignarro et al., 2007). Arguably, it is therefore important to find strategies for modelling health behaviour covariance, not simply to control for the presence of other health behaviours.

[3]Interactions can also reduce risk. The risk from one behaviour could be lessened, depending on the level of another behaviour (Miles & Shevlin, 2000).

3.3.3 Measuring health behaviours as dimensions

There are several different ways that multiple health behaviours can be measured. Some researchers have opted for additive scoring of behaviours (Belloc & Breslow, 1972; B. M. Brock et al., 1988; Wingard et al., 1982; Laaksonen et al., 2003), structural equation modelling (a method described in chapter five; Boniface & Tefft, 1997; Hampson, Goldberg, et al., 2007), cluster analysis (Schuit, Loon, Tijhuis, & Ocke, 2002; Vollrath & Torgersen, 2002), or factor analysis (Dean & Salem, 1998; R. R. Vickers et al., 1990; Stefansdottir & Vilhjalmsson, 2007). The factor analytic approach is similar to that taken by trait researchers, where one or latent variables (see chapter five) explain the correlation between health behaviours (Vickers & Hervig, 1984; R. R. Vickers et al., 1990; Donovan & Jessor, 1985; Wasylkiw & Fekken, 1999, 2002). However, there is no agreement in the literature on what the best approach should be. There are multiple working hypotheses under consideration (Chamberlin, 1965). When variables are highly correlated, as is often the case for health behaviours, the factor analytic approach is appropriate because it is more parsimonious. We can suggest that the explanation for the correlation is that the measured variables are an expression of a more general, latent variable (Deary et al., 1996; R. B. Kline, 2005; Loehlin, 2004). It is not always possible to fit a single latent variable (e.g., Hampson, Goldberg, et al., 2007). Boniface and Tefft (1997) used structural equation modeling to extract a latent variable for health behaviours from the Health and Lifestyle Survey data. They showed that the latent variable, labelled "lifestyle", was predictive of CVD risk, over and above the contribution from individual health behaviours. I return to this data set in chapter eight.

3.3.4 How many dimensions?

For health behaviours beyond the big four, some authors suggested that a single general factor of health behaviour (g) might account for most of the variance (Boniface & Tefft, 1997; Donovan & Jessor, 1985; Belloc & Breslow, 1972). Later work (Vickers & Hervig, 1984; R. R. Vickers et al., 1990) led to the development

of the Health Behaviour Checklist (hereafter, HBC), suggesting that between two and four factors adequately account for the variability and structure in up to forty different health behaviours. This instrument is described in the next section. This was a more comprehensive approach, in that up to forty behaviours are included. However, a general "health orientation" (Williams & Wechsler, 1972) has not been replicated that can model the variance in all health behaviours (c.f. Boniface & Tefft, 1997). Some behaviours are independent of one another or even negatively correlated (Dean & Salem, 1998; Mechanic & Cleary, 1980). It is important to note that smoking and alcohol do not correlate with each other in every study. Smoking and alcohol are particularly prone to relate inconsistently (Haylett, Stephenson, & Lefever, 2004). This may be due to confounding factors, such as the social patterning of smoking and alcohol use (Laaksonen, Lahelma, & Prttl, 2002; Laaksonen et al., 2003). Therefore, it is important to consider SES in order to understand health behaviours.

3.4 The Health Behaviour Check List (HBC)

The decision was made to use the Health Behaviour Checklist (HBC), in the HAPPLE study (see chapters six and seven). The HBC is a set of 40 questionnaire items designed to assess a wide (although not exhaustive) list of health behaviours and score respondents on these dimensions (R. R. Vickers et al., 1990). This inventory asks participants to rate on a five-point scale how strongly they endorse an item, using tradition Likert response options[4]. Wellness Maintenance and Enhancement (WME) consists of 10 items such as "I exercise to stay healthy". Accident Control (AC) consists of 6 items such as "I have a first aid kit in my home". Traffic Risk (TR) includes seven items such as "I speed while driving". Substance Risk (SR) includes four items, such as "I don't smoke". The TR and SR scales are coded negatively, so that higher scores indicate increased risk taking.

[4]The response options used vary between studies. For example, "Not at all like me" (scored 1) to "Very much like me" (scored 5; R. R. Vickers et al., 1990); or "Disagree strongly" (scored 1) to "Agree strongly" (scored 5; R. R. Vickers et al., 1990; Wasylkiw & Fekken, 2002); or "very uncharacteristic of me" (scored 1) to "very characteristic of me" (scored 5; Roberts et al., 2005).

Scores on the HBC can range from 10 to 50, 6 to 30, 7 to 35, and 4 to 20; on the WB, AC, TR and SR scales, respectively.

3.4.1 Reliability and validity of the HBC

Validity. Criterion validity has been demonstrated for the HBC in several studies, because the scales correlate with personality traits such as Conscientiousness (Booth-Kewley & Vickers, 1994; Goldberg, 1999; Roberts et al., 2005; Wasylkiw & Fekken, 1999, 2002). Goldberg (1999) factor analyzed the HBC and found three factors (Health Concerns, Good Health Practices, and Risk Avoidance), and used these to compare the predictive validity of the IPIP in comparison to the NEO and other existing inventories. The IPIP outperformed these, by achieving higher criterion validity coefficients. The Health Concerns and Good Health Practices scales (not Risk Avoidance) correlated with healthy eating habits, in a different study (Goldberg & Strycker, 2002). Although these were not the original published scales, the coefficients provide additional evidence of validity for the HBC items, albeit, in the form of a different measurement model.

Reliability. Acceptable internal consistency for the HBC has been demonstrated for WB, AC and TR (R. R. Vickers et al., 1990; Wasylkiw & Fekken, 2002). The original study used three samples, comprised of Navy, Army and Marine recruits. Internal consistencies ranged from .74 to .82 (WME), .57 to .73 (AC), .64 to .75 (TR) and .44 to .60 (SR). Clearly, these values could be improved. They may reflect the relatively small number of items on each scale, the fact the health behaviours do not form a coherent scale, or because the content coverage is quite wide (K. Bollen & Lennox, 1991). In subsequent studies, SR has again shown low reliability (Wasylkiw & Fekken, 2002). Roberts et al. (2005) excluded the scale entirely, combing WME and AC into "Preventive Health" and using this alongside TR. Low to medium inter-item correlations (R. R. Vickers et al., 1990) suggest that health behaviours are not tightly correlated constructs, when compared to constructs such as mental ability or personality traits.

3.4.2 Unpopularity of the HBC

There are relatively few citations made in the literature to the HBC, suggesting that the measure is not popular. There are at least six reasons that could explain the poor uptake of the HBC by researchers using psychometric methods. First, the SR scale has poor reliability, perhaps because substance use items do not correlate consistently from study to study (Haylett et al., 2004). Second, there are only a small number of items in the SR scale. This necessarily reduces the upper limit for Cronbach's alpha (R. K. Cohen & Swerdlik, 1998). Third, single items may not be appropriate for measuring substance risk. It may be necessary to ask "do you smoke" and "if you do smoke, how many cigarettes per day?". This is more face valid than asking respondents to agree or disagree with items such as "I don't smoke", on a five-point scale. Fourth, despite their comprehensiveness, forty items is a large number of items, particularly if researchers wished to use techniques that require large sample sizes, such as confirmatory factor analysis, a technique which will be described in chapter five. A long scale is unlikely to be used in medical or clinical settings. In clinical settings, questionnaires must be quick, easily processed, and coded scores easily interpreted (Ware et al., 1993; Ware, Kosinski, & Keller, 1994). There may even be insurance issues, if time taken to use questionnaires is not reimbursed by insurance companies (Solari, 2005). Few health scientists would consider replacing familiar and face valid measures of key health behaviours such as smoking and alcohol use, with single items such as "I don't smoke" measured on a Likert scale ("strongly agree" to "strongly disagree"). Fifth, large amounts of missing data may result from administering items relating to driving for participants who do not own a car. Sixth, the HBC is quite far removed from behaviour. The items are very similar to personality items, which may not be appropriate.

3.5 Causes of health behaviour

3.5.1 Theories of health behaviour

There are several theories about the causes of health behaviours. It is beyond the scope of this thesis to consider theoretical perspectives of health behaviour, beyond that which involves personality traits. However, trait researchers are increasingly drawing on theoretical resources to improve their models. Sociological approaches tend to focus on income, poverty and other measures of SES (Brunner & Marmot, 2006). Health Psychologists have discussed the role of beliefs about health or intentions to perform healthily (Leventhal, Weinman, Leventhal, & Alison, 2008). Geneticists have found genetic contributions to health behaviours. For example, there are candidate genes for smoking (Posthuma, Cherny, & Boomsma, 2006). There has also been some recent interest in the notion that health behaviours can be "programmed" early in life. For example, the thrifty genotype hypothesis proposes that early poor nutrition produces later permanent changes in glucose-insulin metabolism (Hales & Barker, 2001). Early life environments can also influence cortisol production, a stress hormone that can influence health and illness (Wust, Entringer, Federenko, Schlotz, & Hellhammer, 2005). Similarly, life course epidemiology seeks to understand the longitudinal aspects which determine health and illness (Bhopal, 2002). Referring to personality traits and health, Friedman (2000) considered a metaphor of health "trajectories" that people may follow, trajectories that are partly influenced by personality traits. It is possible that early development and personality traits program patterns and trajectories of health behaviours. He argued that health is determined across the lifespan by dynamisms (processes or trajectories), mechanisms (mediators) and tropisms (push and pull factors that have positive or negatively influence).

3.5.2 Socio-economic status (SES) and health behaviours

Health behaviours are influenced strongly by structural factors, both social and economic. The term socio-economic status (SES) is used to describe measures of social and economic standing that take into account these factors. SES is an umbrella term for at least four different forms of social inequality: educational attainment, social class, social status, and material circumstances such as income (Blane, 2006). Features of a neighbourhood or place can also contribute to SES, in a process called area based deprivation (Stafford & Mccarthy, 2006). Each measures some unique aspect of SES which the other does not. SES could be defined by social class, social status, education, or material circumstances. Therefore, SES is a construct, in the sense that it can not be measured directly. SES is a potential confounder of the relationship between health behaviours and health outcomes (morbidity, mortality). This is because health behaviours are not evenly distributed at all levels of SES. Neither are multiple health behaviours (Laaksonen et al., 2002, 2003). It is important to consider SES as a mediator of relationships between traits, health behaviours and health outcomes. Health behaviours are socially patterned (Jarvis & Wardle, 2006). Lower income individuals are more likely to smoke, drink alcohol, be sedentary and have poor nutrition (Brunner & Marmot, 2006). SES is a candidate mechanism that might explain why traits are associated with health (see chapter seven).

Risk factors for multiple risky health behaviours

There is evidence that multiple health behaviours are influenced by structural factors (e.g., SES, Laaksonen et al., 2002). Schuit, Loon, Tijhuis, and Ocke (2002) found that multiple behaviours were more common in the low educated, unemployed and those whose health had deteriorated in the previous year. Higher age, ethnicity minority status, low education, non-married marital status, the presence of chronic diseases, mental distress, and lack of health insurance; have also been linked to multiple risky health behaviours (Fine, Philogene, Gramling, Coups, & Sinha, 2004). Highly educated, older and female individuals have a

lower prevalence of multiple risky health behaviours (Berrigan, Dodd, Troiano, Krebs-Smith, & Ballard, 2003; Fine et al., 2004; Schuit et al., 2002). Structural factors alone do not account for all the variance in multiple health behaviours. Therefore, psychological (e.g., traits Booth-Kewley & Vickers, 1994) and social factors (e.g., education Hampson, Goldberg, et al., 2007) are also involved.

3.5.3 Conclusion

This chapter has introduced the concept of health behaviours, their importance as predictors of health outcomes, strategies for measuring them. Dimensional approaches are particularly attractive to trait researchers (e.g., Booth-Kewley & Vickers, 1994; Wasylkiw & Fekken, 2002), because they are easily integrated into personality-health models. However, I have illustrated several reasons why the HBC has not gained popularity. The SR scale, in particular, requires improvement. The HAPPLE study, introduced in chapter six, will aim to address the limitations of the HBC, by collecting additional data on separate health behaviours (e.g. cigarette smoking and alcohol use). Unlike traits (chapter two), there is no widely agreed strategy for measuring health behaviours. There is, however, an agreed measurement model for self reported health outcomes, which is the focus of the next chapter.

CHAPTER 4

Measuring health outcomes

> *Self ratings of health cannot serve as a substitute for epidemiologic diagnoses. These ratings clearly measure something more — and something less — than objective medical ratings. However, our data demonstrate that self assessment of health is not random but persistently and positively related to objective evaluations of health status* (Maddox & Douglass, 1973, p. 92).

4.1 Introduction

As the final chapter concerned with measurement, this chapter describes how self reported functional health status can be measured. It clarifies the rationale for adopting the Short Form 36 Health Outcome Questionnaire (SF-36, Ware et al., 1993). The SF-36 will be used in the HAPPLE study (chapter six) and the cortisol study (chapter eight), to measure health status. The first section defines functional health status. The second section introduces the SF-36, with particular relevance to its reliability and validity. In contrast to dimensional approaches to health behaviour described in the last chapter, there is broad agreement that the SF-36 dimensions are a valid measure of health status. As in the big five model of personality, the dimensions are organized into a hierarchy, containing broad and specific measures.

4.2 Functional health status

There are many approaches to conceptualizing health: (1) medical approach: This simply defines health as the absence of disease; (2) holistic approach: This defines health in terms of physical, mental and social components; (3) wellness approach (Bowling, 1997; Korotkov & Hannah, 2004). This adds to the holistic approach, and includes self reports such as happiness and quality of life; and (4) eclectic approach: This is a "catchall" approach for unusual definitions of health, such as worker productivity (Korotkov & Hannah, 2004). In this thesis I adopt a combination of the second and third approaches. Health is operationalized as *functional health status*: "what people are able to do and how they feel" (Ware et al., 1993, p. 9:1). This is congruent with how the World Health Organization (World Health Organization, 2002) and the Department of Health (Home Office, 2004) define health.

4.2.1 Functional health status as a latent variable

Functional health status is a latent variable. It cannot be observed directly, but it may be possible to measure its effects. It is a construct. Several different aspects of physical and mental functioning need to be considered, in order to describe individual differences in health status (Bowling, 1997). These indicators have been widely researched, and were amalgamated into the SF-36 (Ware et al., 1993). A large body of literature already exists, supporting the validity of the SF-36 (Ware et al., 1993; Ware, Kosinski, & Keller, 1994; Ware & Gandek, 1998). Like personality traits (chapter two) but unlike health behaviours (chapter three), a hierarchical measurement model is widely accepted in health outcome research (Bowling, 1997; Ware & Gandek, 1998). A suitable measurement model has already been established. The SF-36 is not the only inventory, but it is the most validated (Bowling, 1997; Ware et al., 1993). Rowan (1994) argued that the search continues for "the optimal measure — simple, cheap, short, acceptable to both users and respondents, valid, reliable and sensitive to change" (p.66). The aim of

this chapter is to defend the decision to adopt the SF-36 in my research, by showing that it meets all these criteria. Therefore, it is suitable for use in personality-health models.

4.2.2 The SF-36

In order to know the impact of health care, treatments and operations on functional health status, it is necessary to measure functional health status. Any measure of functional health status must do three things. First, it must measure how people feel. Second, what people can do. Third, it has to be able to measure changes in "what people are able to do and how they feel" (Ware et al., 1993, p. 9:1). It is not always possible to administer health status inventories before and after treatments, so an additional scale that could detect recent changes, without the need for longitudinal measurement, would be valuable. The SF-36 is a reliable and valid measure of all three of these indices (Ware et al., 1993). Furthermore, health status is not a unitary phenomenon. Like personality, it has principal components. That is, there are different dimensions of health upon which people vary. A good self-report measure of health status would need to measure several dimensions of health status. In the section below, I explain the historical development of the SF-36 and how the dimensions were selected.

4.2.3 History of the SF-36

Three major studies formed the groundwork for the development of the SF-36 (see Ware et al., 1993). It was amalgamated from instruments and items that were in use since the 1970s and 1980s, including the General Psychological Well-Being Inventory, various physical and role functioning measures, the Health Perceptions Questionnaire, other measures from the Health Insurance Experiment, and the Medical Outcomes Study (see Ware et al., 1993, p. 2:1-2:2). As with the NEO inventories, the SF-36 model is a benchmarking tool that helps integrate the literature. It retains traces of earlier instruments and questionnaire items from older studies, like a palimpsest. The development of a measurement

model for functional health status followed a similar pattern to the development of the big five. That is, the health sciences were faced with the same problem that trait researchers faced. A multitude of terms to describe health had proliferated in the literature (Bowling, 1997; Ware et al., 1993), and many constructs described the same things — they were "jangles" (Deary, 2000, p. 111) for what researchers now call functional health status. There was need to take stock of the constructs in the literature, and provide a measurement model that could be used as a benchmark (cf. de Raad & Goldberg, 2002). This is an important process. Piaget warned that too many controversies in psychology occur because antecedents are sought for consequential schemes before the nature of the outcome has been agreed (M. Chapman, 1988). Defining central concepts precisely substantiates later theories and explanatory accounts (Gholson & Barker, 1985). In other words, there was a rush to explain constructs before their descriptive measurement model had been agreed (see also Deary, 2000). In the competition between multiple working theories, poorly described concepts will begin to fail (Gholson & Barker, 1985).

4.3 The SF-36

The eight scales in the SF-36 are listed below, with a description corresponding to the highest score on each scale[1]:

PF. Physical Functioning. "Performs all types of physical activities including the most vigorous without limitations due to health".

RP. Role Physical. "No problems with work or other daily activities".

BP. Bodily Pain. "No pain or limitations due to pain".

GH. General Health. "Evaluates personal health as excellent".

EV. Energy Vitality. "Feels full of pep and energy all of the time".

SF. Social Functioning. "Performs normal social activities without interference due to physical or emotional problems".

RE. Role Emotional. "No problems with work or other daily activities".

[1] http://www.sf-36.org/tools/SF-36.shtml

ME. Mental Health. "Feels peaceful, happy, and calm all of the time".

Two component summary scales are defined by the eight scales. Physical Component Summary (PCS) is defined by PF, RP, BP and GH. Mental Component Summary (MCS) is defined by EV, SF, RE and ME[2]. Originally, these summary scales were developed using factor analysis. The criterion of eigenvalues greater than one was used to determine the number of factors (dimensions) needed to describe functional health status[3]. The SF-36 scales were then rotated to simple structure "to facilitate interpretation" (Ware et al., 1993, p. 3:8), which means that items correlate highly with their parent scale, and the least with scales they are not designed to measure.

Scoring using the RAND method. The eight SF-36 scale scores can be calculated using the RAND 36-Item Health Survey 1.0 method[4]. The raw data are first converted to scale scores. For interpretability, these are often transformed into T-scores. To calculate T-scores, Z scores are first calculated for each of the eight scales. Z scores have a mean of 0 and a standard deviation of 1 (Field, 2005). The Z scores are then multiplied by the factor score coefficients. These are factor loadings for the PCS and MCS factors, developed from a large study that contained a representative sample of the U.K. population (Jenkinson, Stewart-Brown, Petersen, & Paice, 1999). The values are then summed. Finally, the PCS and MCS scores are multiplied by ten and added to 50. This produces T scores, which have a mean of 50 and a standard deviation of 10 (R. K. Cohen & Swerdlik, 1998). U.K. norms, and factor score coefficients are available (Jenkinson, Coulter, & Wright, 1993; Jenkinson, Layte, Wright, & Coulter, 1996; Jenkinson, Stewart-Brown, Petersen, & Paice, 1999; Lloyd, Jenkinson, & Stewart-Brown, 1999). The

[2]However, GH, VT and SF have significant correlations with the other summary measures.

[3]This criterion, referring to the proportion of variance a factor explains, is less popular in contemporary research because it can lead to too many factors (P. Kline, 2000) but does ensure that the variance of a factor is greater than the variance of a single variable. A eigenvalue of 1 is best understood as a minimum value, rather than a way to choose the number of factors.

[4]This method, used for the studies described in this thesis, differs slightly from the method described in the SF-36 manual (see `http://tinyurl.com/2sj8t9`).

norms allow researchers to compare SF-36 scale and summary measures with the general population.

4.3.1 Reliability of the PCS and MCS

Internal consistency and test-retest reliability coefficients range from .89 to .94 (PCS) and .84 to .91 (MCS). In the U.K., estimates are .92 (PCS) and .89 (MCS, Ware et al., 1993, p.5:2). PCS and MCS define more levels of health than any of the eight sub scales, so they are more reliable, which increases the capacity for validity. The PCS and MCS scales are more reliable than the eight scales. For example, the range in reliability coefficients for the eight scales is .68 to .93, whereas the range for the PCS and MCS reliability coefficients is .93 and .88 (Ware et al., 1993, p. 4:4).

4.3.2 Validity of the PCS and MCS

Validity for the PCS and MCS was found in the Health Outcomes Studies, where patients were compared on levels of severity, and with healthy volunteers. Health criteria were chosen that were clinically (e.g. a diagnosis) and socially (e.g. job loss) important, plausible, and measured independently of PCS and MCS scores (Ware et al., 1993). The following criteria were selected by the authors:

- Presence of four chronic medical conditions.
- Two levels of severity of hypertension.
- Four levels of diabetes.
- Two levels of severity of congestive heart failure.
- Presence of one of 16 co-morbid conditions and a count of 10 others.
- Frequency of symptoms in ear, nose, throat, central nervous system, musculoskeletal system, gastrointestinal system, genitourinary system.
- Comparisons across age for healthy volunteers and patients with uncomplicated hypertension (cross-sectional and longitudinal).
- Comparisons of patients with and without depression (cross-sectional and longitudinal).

The authors of the SF-36 (Ware et al., 1993) approached the concept of validity theoretically, by testing six hypotheses in relation to these groups:

1. People with more severe conditions would score worse.
2. Advanced conditions would score worse.
3. Greater frequency of symptoms would score worse.
4. Self reported change in physical, mental and general health at 1 year would be most related to changes estimated from repeated measures of the same concept.
5. Physical health but not mental health would decline with age.
6. Mental health would be better for those without depression and improve recovery from major depression.

All six of these hypotheses were supported. Relative Validity (RV) coefficients were estimated using the ratio of F-statistics from MANOVA (multivariate analysis of variance) models when comparing these illness severity levels, and control groups (see Ware et al., 1993, p. 6:4). Both scales correlate with specific symptoms, in patterns consistent with construct validity for the scales. For example, the highest PCS correlations are shortness of breath (-.55), stiffness (-.53) and types of pain (range -.41 to -.42). For MCS, highest correlations were for headaches more than usual (-.38), waking up early or being unable to get to sleep (-.35) or feeling dizzy when standing (-.32).

Construct validity for the PCS

The PCS was also found to correlate with important health criteria: severity of disease (range .83 to 1.00), co-morbid conditions (.89), symptom clusters (.55 to .86), age differences (.20 to .82), self-reported change in physical health (.74 to .79), self-reported change in mental health (.05), clinical depression (.01) (Ware et al., 1993, p. 6:6). The pattern of correlations indicates convergent and discriminant validity, because the higher correlations are for physical health criteria and the lower correlations are for mental health criteria. Predictive validity for all

Table 4.1: PCS scores predict odds ratio of death within five years of measurement

Level of PCS	T-score mean	Unadjusted odds ratio	Age adjusted odds ratio
1	55-72	1	1
2	45-54	2.7	2.0
3	35-44	3.6	2.2
4	25-34	9.7	4.8
5	8-24	14.8	6.8

cause mortality has been demonstrated for the PCS. Table 4.1 shows that death within five years is more likely for people with lower PCS scores (Ware et al., 1993, p. 7:11). All of the odds ratios are significant, because their confidence intervals do not include 1 (Tabachnick & Fidell, 2000). The hypothesis that PCS scores would decline with age, but not MCS scores, was supported. PCS scores are, on average, two points lower for each increasing year of age (Ware et al., 1993, p. 7:18).

Construct validity for the MCS

The RV coefficients between MCS and health outcomes were: severity of disease (.24), co-morbid conditions (.89), symptom clusters (.55 to .86), age differences (.35), self-reported change in physical health (.18 to .29), self-reported change in mental health (1.02), clinical depression (1.03 to 1.47) (Ware et al., 1993, p. 6:6). It is also noteworthy that MCS correlates highly with cognitive functioning (.70) but neither the PCS nor MCS scale correlates with sexual functioning. These two outcomes were included in the original Health Outcomes Studies because they are health outcomes. However, they are not included in the measurement model for the SF-36 (Ware et al., 1993). Therefore, functional health status (as measured by the SF-36) does not include sexual health functioning, but may share variance with cognitive functioning. For MCS, there is no equivalent "ultimate" outcome, but depression is an important mental health outcome (World Health Organization, 2001). Table 4.2 shows the prevalence of these outcomes at different levels of MCS (Ware et al., 1993, p. 7:12). Validity is shown because higher MCS scores are associated with a lower prevalence of depression and stress. In contrast, life satisfaction is higher for higher MCS scores. The MCS also correlates with the Beck Depression Inventory (-.52) and with the Hospital Anxiety and Depression Scale (HADS) Depression sub scale (-.64) (Birks, Roebuck, & Thompson, 2004). MH and MCS cut-off scores of 52 and 42, respectively, can detect depression (Ware et al., 1993; Silveira et al., 2005). Therefore, the MCS can operate as a surrogate for clinical measures of mental health, or to screen for clinical depression. It "does as

well or better than the "best scale in mental health tests" (Ware, Kosinski, Bayliss, et al., 1994, p. AS274) and these correlations support that.

Choice of bandwidth or fidelity

There are advantages and disadvantages in using the eight scales of the SF-36, rather than the PCS and MCS component summaries. Many of these parallel the advantages and disadvantages of personality facets discussed in chapter two. The key advantage is that each sub scale has unique variance that is not accounted for by the PCS and MCS. There may be effects that are specific to one or more of the eight scales. The disadvantages are threefold. First, the PCS and MCS are more reliable because they retain only shared variance from the sub scales which define them. Second, the PCS and MCS account for 80 to 85% of the variance of the eight factors. Third, the PCS and MCS "have the potential to reduce the number of statistical comparisons...without substantial loss of information" (Ware et al., 1993, p. 2:1). Conducting tests on eight scales, rather than two, increases the type I error rate — rejecting the null hypothesis, of no association, when it is true. Researchers can choose to operate at the level of an SF-36 total score, at the level of the PCS and MCS, or at the level of the eight sub scales (Rowan, 1994). There is validity supporting the use of a single item from the SF-36: "Compared to others of your same age and gender, would you say that in general your health is (1) Poor, (2) Fair, (3) Good, (4) Very Good, or (5) Excellent" (e.g., B. P. Chapman, Duberstein, & Lyness, 2007; Hampson, Goldberg, et al., 2007)[5]. Finally, The PCS and MCS are only weakly positively correlated. This means that it is possible to separate physical from mental health variance. The PCS measures physical morbidity. The MCS measures psychological, or mental, morbidity. Empirical evidence, supporting the reliability and validity for these scales, is now described.

[5]Similarly, Wasylkiw and Fekken (2002) assessed self reported health using a single item: "In general, would you say your health is...".

Table 4.2: PCS scores predict odds ratio of death within five years of measurement

Level of MCS	T-score mean	Percent depressed	Percent stressed	Percent satisfied with their life
1	65-74	10.7	0.0	78.6
2	60-64	12.3	6.7	73.4
3	55-59	18.4	11.5	66.3
4	50-54	29.1	18.4	47.5
5	45-49	45.6	30.5	31.9
6	40-44	58.9	37.6	32.4
7	35-39	73.9	42.5	18.4
8	30-34	89.2	54.9	7.0
9	9-29	93.5	78.9	5.7

4.3.3 Strengths of the SF-36: Summary

The authors of the SF-36 approached validity by treating it as set of hypotheses, designed to demonstrate construct validity (discriminant and convergent validity, and for the PCS, ultimate validity). All of their hypotheses were supported, showing that the SF-36 is a valid measure of functional health status. Furthermore, the PCS and MCS are independent predictors of many illnesses. This confirms that, although the PCS and MCS are correlated, they describe independent dimensions of functional health status. A summary of the benefits of the SF-36 is now provided:

Model. The SF-36 has an established measurement model. The model is hierarchical, allowing choice of specificity in measurement.

Reliability. The SF-36 is reliable. The PCS and MCS are also reliable.

Length. The SF-36 is short. This reduces respondent burden and time taken to complete it.

Norms. The SF-36 has norms, including norms for the U.K (Jenkinson et al., 1993). This is important because I planned to recruit healthy volunteers from the U.K. and needed to ensure that they matched healthy norms.

Validity. The SF-36 is a valid measure of health status, and of recent change in health status. The PCS additionally predicts death and other key outcomes. The MCS additionally acts as a surrogate for clinical measures of anxiety and depression. There is a nomological network providing construct validity.

Popularity. The SF-36 is widely used. It has a large nomological network. Over 300 articles were published between 1988-1995 that cited the SF-36 (Jenkinson et al., 1999).

Benchmark. The SF-36 is a common yardstick for measuring burden of disease and comparing treatment success. The SF-36 scores are comparable across studies. When a nomological network becomes moderately large, systematic reviews of the literature are often conducted (Torgerson, 2003). These are more straightforward to conduct when the measures of the variables are comparable across studies (Shenkin, Starr, & Deary, 2004; Torgerson, 2003).

Coverage. The SF-36 measures all important dimensions of health status[6].

Change. The SF-36 can detect, explain and track changes in health over time.

Outcomes. The SF-36 acknowledges patient's total functioning when choosing among treatments.

Efficiency. The SF-36 is an efficient use of health care resources. It is quick, cheap, reliable and valid, meeting the criteria for use in applied settings (Solari, 2005).

Models of personality and health must contain reliable and valid measures of health (Matthews, Deary, & Whiteman, 2003). The SF-36 meets these criteria, and the criteria for being an "the optimal measure — simple, cheap, short, acceptable to both users and respondents, valid, reliable and sensitive to change" (Rowan, 1994, p. 66). The SF-36 manual allows researchers to (a) interpret scores; (b) compare scores to other studies; (c) plan future studies (Hemingway, Stafford, Stansfeld, Shipley, & Marmot, 1997). This also allow comparisons within individuals over time, and individual assessments in medical settings. The SF-36 is unusual and attractive in that it places weight on both personal and social implications of different disease states (Ware et al., 1993). As such, it moves beyond the simple medical model of health (Bowling, 1997), without compromising on validity.

4.3.4 Acknowledging disadvantages and criticisms of the SF-36

The SF-36, like any inventory, is not a perfect measure of functional health status. This section summarizes its three main disadvantages.

Healthy samples. The appropriateness of the SF-36 for healthy volunteers has been questioned. Fitzpatrick (1994) argued that health outcome measures are not suitable for survey or general population studies because the modal value is usually zero. This means that there may be ceiling or floor effects (Ware et al., 1993, p. 4:4), where many respondents score the maximum or minimum

[6]The SF-36 does not measure sexual and cognitive functioning, which are aspects of functional health status. However, MCS scores correlate highly with cognitive functioning.

on a scale. If there is restricted range, this will limit the size of associations that can be obtained with predictors (see Deary, 2000, p. 12). Scores are unlikely to provide the discrimination needed between different levels of physical health status (Fitzpatrick, 1994, p. 32). Indeed, "Not all questionnaires are appropriate in all settings, for all purposes or for all respondents" (Jenkinson, Bardsley, & Lawrence, 1994, p. 178). In contrast, Rowan (1994) argues that the SF-36 and even single item measures are suitable for routine settings.The SF-36 scales do have problems with ceiling and floor effects, but the higher order PCS and MCS scores do not (Jenkinson, Layte, & Lawrence, 1997). This means that, providing there is sufficient variability, the PCS and MCS can be used in statistical models. For this reason, the PCS and the MCS were used to model health outcomes in the HAPPLE study (chapter seven).

Mode effects. Scores on the MCS are 2.43 points higher when the SF-36 is completed on paper than when completing during a telephone interview (Ware et al., 1993). There remains the possibility that scores are affected by Internet administration of the SF-36. McCue, Buchanan, and Martin (2006) found that the measures of psychopathology (e.g. prospective memory problems, anxiety and depression) are inflated when administered online. The mechanism underlying this finding is not yet known. In the future, it would be wise to conduct a randomized controlled trial of SF-36 responses where participants complete the SF-36 online and on paper. There is, however, evidence that electronic administration does not significant effect SF-36 scores (J. M. Ryan, Corry, Attewell, & Smithson, 2002). This cross-over trial showed that electronic administration reduced missing data, although participants preferred the paper version.

Layout. There have been criticisms of the layout and wording of the SF-36 (Lloyd et al., 1999), particularly version 1. In a published letter, Lloyd, Jenkinson, and Stewart-Brown (1999) argued that the measure is not suitable for older age groups because many of his patients failed to understand the questions and complete it correctly, creating error in the scores obtained. In contrast,

Hayes, Morris, Wolfe, and Morgan (1995) found that the SF-36 was suitable for older groups, but not those older than 75. In response, Jenkinson argued that such criticisms of the SF-36 are rarely supported by evidence, "based upon little more than anecdotes, rather than rigorously conducted qualitative studies" (see , Lloyd et al., 1999). Jenkinson also stressed that there will always be error present in any questionnaire, which is why the SF-36 employs multi-item scales.

4.3.5 The future of the SF-36

For researchers who promote the use of the SF-36, their aim is to have the instrument administered in medical settings as a matter of routine (Ware et al., 1993). This parallels the claim made about traits measures in Matthews, Deary, and Whiteman (2003) that "If personality traits do influence health, then this is one of the prime reasons to measure personality traits in medical settings" (p. 273). However, it is worth noting that self reports of functional health status are not routinely measured. Two-thirds to three-quarters of adults in the US say that physicians rarely or never ask about limitations in performance in everyday activities (physical, social, role functioning), even if they have chronic conditions (Ware et al., 1993). It is well known that depression is often not detected in primary care (Steer, 1999) and up to 50% of depressed patients leave consultations without their condition being recognized (Kessler, Bennewith, Lewis, & Sharp, 2002). This is a sobering thought for personality-health researchers who would like traits to be measured in medical settings. If measures of functional health status are unpopular, how likely is it that measures of traits will be embraced? Fortunately, there is evidence of a sea change in how medics perceive self reports of health. Use of the SF-36 has increased, and health authorities and general practitioners know they have a responsibility to evaluate the effectiveness of health care (Hemingway et al., 1997). Ware et al. (1993) said they hope that standardized health surveys could become the new "laboratory tests" of medical practice (p.10:1). It remains to be seen whether medical practitioners adopt measures of personality traits in their practice.

4.3.6 Conclusion

In conclusion, I have shown that the SF-36 is a reliable and valid measure of functional health status. Although the scales of the SF-36 have restricted range, when used with healthy volunteers, calculating scores for the PCS and MCS should address this. The component summary scores should address problems with ceiling and floor effects that can occur in the shorter scales (Jenkinson et al., 1997). Therefore, they should be appropriate to use with samples of healthy volunteers.

CHAPTER 5

Latent variable modelling: a framework

> *The purpose of modeling data is to describe the structure of a data set in a simple way so that it is more understandable and interpretable. Essentially, modeling data amounts to specifying a set of relationships between variables* (Muthén & Muthén, 1998–2007, p. 510).

5.1 Introduction

Operationalizing constructs involve selecting "measures of the variables represented in the model" (R. B. Kline, 2005, p. 64). Once constructs have been operationalized and validated, they can be used in statistical models (Cliff, 1983). The previous three chapters concerned measurement of traits, health behaviours and health outcomes. This chapter is a move from measurement, to modelling. It introduces a general framework for modelling personality-health associations: the generalized latent variable modelling (LVM) framework (Muthén, 2002). This provides a starting framework and reference point for models described in later chapters of this thesis.

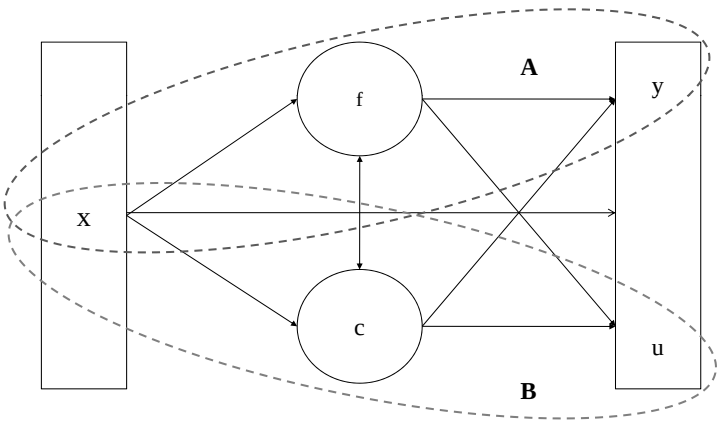


Figure 5.1: Path diagram illustrating the generalized latent variable modelling framework

5.2 The generalized LVM framework

Most statistical techniques can be described as special cases of LVM (Miles, 2003; Muthén, 2002). Figure 5.1, adapted from Muthén and Muthén (1998–2007), helps to illustrate this. LVM is often presented using a path diagram. Boxes represented observed variables, circles represent latent variables, and arrows represent parameter estimates which are labelled with the size of the association between the connected variables. Although not shown in this diagram, double-headed arrows represent correlations, and triangles could be used to represent means. Residuals are shown with a circular arrow that enters and leaves a box, but are sometimes not shown at all. In Figure 5.1, the rectangles are used to represent several observed variables. Muthén (2002) describes how different statistical techniques, or "galaxies of work" were historically developed in isolation, and each new development took a long time to reach journals and even longer to reach software. Even when included in software, users may not take advantage of the new techniques immediately. It is now possible, and more parsimonious, to assume that all these methods can be understood within a single

framework. Muthén (2002) argues that LVM provides "connections between different modelling techniques and suggests interesting extensions" (p. 112). In both ellipses, arrows represent associations between variables. The choice of appropriate model depends on whether the downstream variables are continuous (y) or categorical (u). Intervening latent variables can be either continuous (f) or categorical (c), or a mixture of the two.

Ellipsis A models

Ellipsis A refers to continuous latent variables (f) and outcomes (c). Most "traditional" techniques used in psychology (Miles & Shevlin, 2000) can be incorporated into this framework. Regression, correlation, partial correlation, factor analysis, t-tests, ANOVA, ANCOVA and MANOVA are all special cases of LVM in ellipsis A, and can be represented in path diagram form (Miles, 2003). Most personality-health models can be described as ellipsis A models. A popular type of LVM is a confirmatory factor analysis (CFA) where latent variables are proposed that explain the correlation between observed variables, such as test items. If the number or nature or latent variables is not known from prior research or theory, then exploratory factor analysis (EFA) is often used instead. The term structural equation modelling (SEM) is often used to describe ellipsis A models where factors are related to other variables and to each other. However, LVM is a more general term which includes these types of models.

Ellipsis B models

Ellipsis B refers to categorical latent variables (c) or outcomes (u). Covariates can be categorical or continuous (x). Another important feature of the framework is that categorical latent variables can be included, such as clusters or latent classes (groups) of people. The terms "within" and "between" refer to multilevel modelling, not discussed here. Cluster analysis, mixture modelling (e.g. latent class analysis), logistic regression and survival analysis, are examples of

ellipsis B models. Table 5.1 compares the features of traditional statistical models with the options available in the LVM framework. The LVM is less restrictive, so that the apparent restrictions on the presence of categorical outcomes, latent variables and indirect effects, is relaxable. Most techniques can be modified, so that multiple working models can be considered side by side (Chamberlin, 1965, cf.).

5.2.1 Decision to use Mplus software

The decision to use Mplus was based partly on the flexibility of the program. Mplus is designed to handle cross-sectional and longitudinal data, missing data, and a mixture of ellipsis A and ellipsis B models. This should make it very attractive to personality-health researchers. Personality-health researchers have not utilized ellipsis B models fully, especially survival analysis and latent class analysis.

5.2.2 The measurement model

The measurement model is "a multivariate regression model that describes the relationships between a set of observed dependent variables and a set of continuous latent variables" (Muthén & Muthén, 1998–2007, p. 43). Therefore, the measurement model describes *which* indicators load onto which factors (Scott Long, 1983).

5.2.3 The structural model

Given that the measurement model has specified a set of factors, defined by observed variables, the structural model links these factors to each other, and to other observed variables which are not indicators. The structural model specifies the strength of the relationships in predicting the variance in dependent variables from the independent variables. It also relates observed variables to each other (Scott Long, 1983; Muthén & Muthén, 1998–2007). The structural model therefore describes *how much* the indicators load onto factors, factors that have been

Table 5.1: Traditional modelling techniques within the LVM framework

Technique	Outcome	Categorical latent variable?	Continuous latent variable?	Indirect effects?	Change over time?
Multiple regression	No	No	No	No	No
Confirmatory factor analysis	No	Yes	No	No	No
Item response theory	Yes	Yes	No	No	No
MANOVA	No	No	No	No	No
Logistic regression	Yes	No	No	No	No
Mediation and path analysis	No	No	No	Yes	No
Survival analysis	Yes	No	Yes	No	Yes
Latent class analysis	No	No	Yes	No	No
Residualized change score analysis	No	No	No	No	Yes

created by the measurement model (Scott Long, 1983). The distinction between dependent and independent variables is discussed in more detail below.

5.2.4 Assumptions of LVM

All models make assumptions about data. As described by R. B. Kline (2005), LVM models have to be multivariate (that is, contain more than three variables). Second, they must be linear (although it is possible to model log, quadratic and cubic relationships). Third, they must be *identified* (explained below). Fourth, there must be unobserved heterogeneity, which can be modelled by one or more latent variables. Latent variables are now described in more detail.

5.2.5 Latent variables: Circles

Continuous latent variables (which have been variously termed proxies, factors and indexes) are unobserved variables that have heterogeneity (Muthén, 2002). That is, they are constructs (see chapter one) that cannot be measured directly, but which have a structure. Researchers attempt to recreate this structure by measuring several indicators of the construct. This general definition is crucial to convey, and very useful. Measurement models are designed to measure latent variables. Personality traits are latent variables (chapter two). Dimensions of health behaviour are latent variables (chapter three). The MCS and PCS scales of the SF-36 are latent variables (chapter four). Latent variables are defined by at least two, and preferably three or more indicators. For example, a latent variable Conscientiousness could be defined by six facets: Competence, Order, Dutifulness, Achievement Striving, Self-Discipline and Deliberation (C1 to C6). Indicators contain at least some measurement error, so it is always best to have multiple indicators of a latent variable. The covariance between three or more observed variables provides latent variables with their source of homogeneity. By definition, the latent variable has less measurement error than any of its indicators alone. This is why latent variables are so popular with trait researchers — they reduce measurement error. When multiple indicators are not available, single

observed variables can be used. Two indicators are usually not sufficient, because this means that the LVM is not identified (R. B. Kline, 2005, p. 174), a term which is later explained.

5.2.6 Observed variables: Boxes

Observed variables represent data that has been measured or recorded. In path diagrams, these are represented by boxes. Observed variables come in two forms: exogenous (upstream) and endogenous (downstream). A model that contains only observed variables is simply a latent variable model without a latent variable. Traditional approaches to modelling (e.g. multiple regression, path analysis, ANOVA) are special cases of the LVM framework without latent variables (Miles, 2003).

Exogenous, upstream variables

Upstream variables could be described as predictor variables "because their causal sources lie external to the path diagram; they are causally independent with respect to other variables in the diagram — straight arrows may lead away from them but never toward them" (Loehlin, 2004, p. 4). Demographic variables, such as age, are upstream variables.

Endogenous, downstream variables

Downstream variables have some causal (in the mathematical sense) influences within the path diagram: arrows may point toward them as well as lead away from them (Loehlin, 2004).

Mediators, upstream and downstream variables

Mediators are variables that have endogenous and exogenous status. They are predicted by upstream variables but also predict other variables further downstream. Variables positioned as mediators are usually done so for substantive

theoretical reasons. For example, a researcher may suspect that health behaviours explain the association between traits and health outcomes. As such, the health behaviour variables lie downstream from traits and upstream of health outcomes. They are candidate mechanisms for the association. Missing data is permitted for mediators, because they are not covariates with upstream status.

Residuals

We could say that residuals are explained by something "outside" the box we want to operationalize, or perhaps that the trait researcher's toolkit can only capture some of the variance (the box), but not all of it (the residuals). Only a portion of the variance in health behaviour can be explained.

Parcelling items into boxes

Observed variables (boxes) sometimes represent the total score of a set of questionnaire items. However, they can be aggregated and promoted to the status of boxes (observed variables) if psychometric work has been conducted to show that they form a single reliable scale, and there is substantive prior research supporting this. This technique is called parcelling (Clara, Cox, & Enns, 2003; Hagtvet & Nasser, 2004; R. B. Kline, 2005; Marsh & Hau, 1999). It reduces the complexity of models and the sample size required. When the nomological network provides enough validity for personality facets, as in the NEO-PI-R, it is appropriate to parcel items into facets and use these in models (Saucier & Goldberg, 2002). Modelling at the item level, without parcelling, would require a separate box for every item, increasing considerably the sample size required to model the facets and higher order traits (Hagtvet & Nasser, 2004).

5.2.7 Arrows: Parameter estimates

The arrows in Figure 5.1 represent an association between two variables: a change at the tail transmitted to a variable at its head (Loehlin, 2004). That is, as one variable changes, it changes the value of the variable it points to. Effect sizes show

how strong this association is. These are the parameter estimates, used to label each arrow. There are some rules about where arrows can appear in LVMs. Loops are not permitted, moving forward then backward is not permitted, and more than one arrow per path is not permitted. When the downstream variables are continuous variables, effect sizes are usually standardized beta weights. These represent a standard deviation change at the tail transmitted to a variable at its head.

Standardized beta weights

Standardized beta weights are estimates of the effect size for a parameter. They have four important features: (1) they are partial, in the sense that they account or control for all of the other arrows in the diagram; (2) they are standardized, in the sense they represent changes in standard deviations of variables, not a change in the units that were measured originally; (3) if there is no arrow connecting two variables, then this means that they are not associated in the proposed model. This is the same as fixing the size of the beta weight to zero; (4) The covariance between two variables in a path diagram is the same as the "summed compound paths between them" (Loehlin, 2004, p. 27). That is, the product of weights from one variable to another. If there are multiple pathways, products representing both paths are summed.

Estimates for categorical outcomes. Different estimates are required when the downstream variables are categorical. For example, logistic regression estimates might be needed to describe the odds of belonging to a "smoker"' category, compared to the "non-smoker category". Log odds (logits) are used to express the probability of a categorical outcome per unit change in a predictor. Taking the exponential gives the odds (Muthén & Muthén, 1998–2007).

5.3 The modelling process

The next section describes the stages of model building advocated by R. B. Kline (2005), which are used as subheadings. Under these subheadings, I have expanded on key points which are particularly relevant to this thesis.

5.3.1 Model specification

When researchers choose a model, they know that the data will not perfectly fit that model. All statistical models imply a source of misfit between the data and the model (Miles & Shevlin, 2000). When a LVM is specified, a set of parameters are implied. Parameters can be left "free" for the computer program to estimate, or they can be fixed. In a path diagram, when no arrow connects two variables, this is the same as saying that the parameter between them is fixed to zero. Sometimes parameters are fixed to values which the researcher estimates. For example, a parameter between health behaviours and age might be fixed to .3, because the researcher knows, from previous research, that the association is a medium effect size. Frequently, the researcher does not know what the parameter value is, so the parameter is left free.

5.3.2 Model identification

A latent variable model is identified if it is "theoretically possible to identify a unique estimate of each parameter" (R. B. Kline, 2005, p. 105). If there are more observations than parameters, the latent variables are given a metric, and there is only one possible solution, then the model will be identified. Starting values can also help. The metric of latent variables is set by constraining the first factor loading to 1.

5.3.3 Model believability

Models should be plausible representations of reality. For example, age and sex are always upstream. Variables which occur later in time cannot predict

events that occurred earlier in time. For some models, the distinction is more controversial. Socio-economic status (SES) is sometimes defined by educational level, occupation, income and social deprivation. As a latent variable, the arrows run from SES to the observed variables, suggesting that SES causes them. This may not be appropriate. Some researchers adopt an emergent variable model, in which the pathways run from the observed variables to the latent variable. This is more difficult to model, but has greater believability.

A note about causality. Models are empirical tests of associations in data, but they do not explain these associations on their own. Arrows in models are hypotheses, and therefore they can be falsified. Structural models need to be replicated, other possible models rejected, evidence from experimental studies provided, and interventions based on models predicted successfully (R. B. Kline, 2005, p. 118). Only when findings are added to the wider literature, the nomological network, can associations be interpreted in terms of the criteria for judgement (see chapter three), and as *causal* (see also Borsboom et al., 2003; Cliff, 1983).

5.3.4 Operationalizing the constructs for the model

Operationalizing a construct for a model involves more than just selecting measures of the variables. It is important that constructs selected from the literature are reliable and valid. It is important to check the reliability and validity of measures *before* they are added to models (Cliff, 1983). Reliability is normally assumed when the model is presented[1].

5.3.5 Using a computer program to estimate the model

Mplus uses two main types of estimators to determine the discrepancy function (F). This is a statistical criterion that the observed covariances are as close as possible to those implied by the model (R. B. Kline, 2005). The choice of estimator, to obtain F, is usually determined by the scale of the dependent variable (Muthén,

[1]When observed variables are modelled at the item level, reliability can be modelled explicitly in confirmatory factor analysis (e.g. Miles & Shevlin, 2007)

2002). If the downstream variables are continuous, than the maximum likelihood (ML) estimator is appropriate. If the downstream variables are categorical, then the weighted least squares estimator is appropriate. If the downstream variables are a mixture of continuous and categorical, then the researcher will probably have to choose a more specialized estimator. ML is the most common estimator used in LVM, in fact R. B. Kline (2005) argues that not using ML, requires explicit justification.

Maximum likelihood estimators

Maximum likelihood (hereafter, ML) is an estimator suitable when the downstream variables are continuous. It maximizes the likelihood that the data (the observed covariance matrix S) were sampled from the population implied by the model Σ (R. B. Kline, 2005). That is, it looks for parameter estimates that maximize the likelihood that the parameter estimates were actually observed — given the model chosen by the researcher: "Given the data, and our choice of model, what values of the parameters of the model make the observed data most likely?" (Crawley, 2007, p. 324). ML is a statistical criterion which tries to minimize the discrepancy function F. Clearly, this is why computer software is needed — testing the fit of different parameters by hand would not be possible. The fit is sought iteratively. This is similar to the least squares criterion in traditional multiple regression. The aim there is to find a regression line that minimizes the size of the squared residuals from the line. ML is conducted on the unstandardized parameters (it is scale free and scale invariant). This is one reason why covariance matrices are generally preferred as data for LVM rather than correlation matrices (R. B. Kline, 2005). ML assumes that downstream variables have a multivariate normal distribution. ML is considered a full information estimator because all of the parameters are calculated at the same time, rather than one at a time.

Statistical significance of parameters. For unstandardized parameter estimates, statistical significance is achieved when the estimate divided by its standard error exceeds 1.96 (Muthén & Muthén, 1998–2007).

Weighted least squares estimators

Weighted least squares (WLS) is a family of estimators, some of which are suitable when the downstream variables are categorical (when we are working in ellipsis B). WLS estimators are more flexible, because they do not assume underlying multivariate normality (Muthén & Muthén, 1998–2007). They do not provide a chi square statistic, but they do provide a log likelihood statistic (see below).

5.3.6 Statistical power

There are not agreed guidelines for sample size in LVM. Less than 100 cases is "untenable", and between 100 and 200 is considered "medium", with over 200 considered "large" (R. B. Kline, 2005, p. 15). However, this does not provide precise information about the statistical power of LVM. There are several, more complex, ways to test the statistical power of a parameter. One is to deliberately mis-specify a parameter to zero, and treat the implied covariance matrix as the null hypothesis. This is then compared to the fit of a correct model, using the chi-square statistic generated by the discrepancy function (Miles, 2003). A second method is perform a bootstrap simulation. Data are generated from a population that is implied by the model. This is repeated a large number of times, and the parameter estimates are averaged (Muthén & Muthén, 2002). Inspection of the parameter estimates can show whether a specified model and sample size have adequate statistical power.

5.4 Missing data estimation

Missing data is common in LVM (Loehlin, 2004). Deleting cases with missing data may not be appropriate, and may result in small sample sizes. This is because when cases are missing at random (related to other variables[2]), deleting them causes bias (R. B. Kline, 2005; K. A. Bollen & Curran, 2005; Loehlin, 2004). Although imputation (e.g. series mean, regression methods) and multiple imputation (the generation of multiple data sets using different imputation techniques, then running analyses of all the imputed data sets as a sort of average data set) are popular, these all involve working with data that was not observed. There are other problems. For example, imputation can result in non-positive definite covariance matrices (Wothke, 1993). This occurs when one of the eigenvalues in the covariance matrix is negative, perhaps because the values replaced are not consistent with the pattern of correlations in the data that was observed (Wothke, 1993). It can also raises ambiguity when interpreting the data (K. A. Bollen & Curran, 2005). For example, if I report the cortisol mean and standard deviation, is this the mean before imputation or after? What is the sample size, before and after? In contrast, the ML approach "opens up the possibility for a variety of analyses with missing data" (Muthén, Kaplan, & Hollis, 1987, p. 459). The estimator uses information for each case only for variables that are available. This method has several advantages: (1) is maintains the asymptotic properties of ML estimators (that is, parameter estimates and significance tests are available). The criticism that imputation is creating data that does not exist, does not apply to ML because no data is imputed.

Missing data on covariates. Many researchers (Shipley et al., 2007, e.g.) exclude cases with missing data on covariates, referred to as casewise deletion. Casewise deletion is very controversial, but modelling missing data on covariates is more difficult. Muthén and Muthén (1998–2007) are worth quoting at length in relation to this issue: "In all models, missingness is not allowed for the observed

[2]Not to be confused with missing completely at random, where missingness is unrelated to other variables.

covariates because they are not part of the model. The model is estimated conditional on the covariates and no distributional assumptions are made about the covariates. Covariate missingness can be modeled if the covariates are brought into the model and distributional assumptions such as normality are made about them" (p. 401-402). This requires numerical integration. Numerical integration is a technique in which expressions of data are brought together by approximating their values, using statistical algorithms and calculus. A number of integration points are used to approximate a solution. The more integration points, the more precise the estimation. However, a large number of integration points "can be computationally demanding" (p. 19). Models can fail to converge if they are too computationally demanding.

5.5 Model fit

When a model has been specified, the estimated parameters of the model are used to generate an implied covariance matrix. That is, a covariance matrix implied by the researcher's model. The sigma symbol (Σ) is used to represent the implied covariance matrix. The fit of the specified model is the difference between the actual covariance matrix (the data, S) and the implied covariance matrix Σ. The difference between them is called the discrepancy function (F). The discrepancy function is estimated by maximum likelihood, describe above. That is, it starts from some values and keeps searching until it finds the parameter values which minimized the size of F. F values follow a chi-square distribution, a familiar distribution that can easily be used to test the statistical significance of the model as a whole: "A properly specified model should lead to non-significant differences between the model-implied and data covariance matrices ($p > .05$) 19 times out of 20" (Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007, p. 843). Fit indices which use the F distribution and chi-square test belong to a family of tests called approximate fit indices (AFI), and are discussed in more detail below. In summary, LVM can be assessed for goodness-of-fit as a whole, and for each parameter individually. A parameter is statistically significant if its

estimate divided by its standard error is greater than 1.96 (Muthén & Muthén, 1998–2007, p. 575). Statistical models are always accompanied by a statement about how well the model fits the data. There is no agreement in the literature concerning the best measures of fit to use.

5.5.1 Goodness of fit criteria

There is not agreement among researchers about appropriate "cut offs" for assessing the goodness of fit of a LVM. Therefore, the values presented below should not imply that there are "golden rules" (Markland, 2007; Marsh, Hau, & Wen, 2004), but they should be used as approximate rules of thumb (R. B. Kline, 2005). Model fit should be assessed using a variety of strategies (L. T. Hu & Bentler, 1998), including the sign and size of the parameter estimates (R. B. Kline, 2005). Indices based on the discrepancy function (the measure F, of how well the model fits the data) can be grouped under the umbrella term approximate fit indices (hereafter, AFI). Fit indices can be considered as multiple working hypotheses (Chamberlin, 1965) regarding the fit of a model.

Approximate Fit Indices

Chi-square. The higher this value, the worse the fit. A non-significant chi-square implies a perfect fit to the data (R. B. Kline, 2005, p. 136). However, chi-square increases with larger sample sizes, and with larger correlations. Therefore, it cannot be used alone to judge model fit. It should always be reported because several other fit indices are derived from this value (R. B. Kline, 2005, p. 136). In addition, two models can be compared using the difference in chi square values and their degrees of freedom, to see if one is a significantly better fit than the other.

Log likelihood. This can be used to compare two nested models. The log likelihood test is the LL from the baseline nested model H0, minus the LL from the comparison model H1, multiplied by -2 (Field, 2005, p. 222). This statistic has a chi-square distribution, so it is possible to state that a second model

is significantly better than a starting model, if the two models are nested. When using the MLR estimator, a correction needs to be applied (Muthén & Muthén, 1998–2007).

CFI. Comparative Fit Index. Values of .90 are reasonable (R. B. Kline, 2005, p. 140), .95 is considered good (L. T. Hu & Bentler, 1998).

TLI. Tucker-Lewis Index (also called non-normed Fit Index, NNFI). Values above .95 are described as good (L. T. Hu & Bentler, 1998)

RMSEA. Root Mean Square Error of Approximation. This measure includes a penalty for models which are not parsimonious (i.e. models that have more pathways). Values lower than .05 are considered good, and values from .05 to .08 are considered reasonable (R. B. Kline, 2005, p. 139). L. T. Hu and Bentler (1998) suggest that .06 is a good fit.

SRMR. Standardized Root Mean Residual. SRMR Values lower than .06 are considered good (L. T. Hu & Bentler, 1998), values lower than .10 are considered favorable (R. B. Kline, 2005, p. 141). This fit index is particularly suitable when working with small sample sizes (Paul Dudgeon, personal communication, June 2007).

AIC. Akaike Information Criterion. The smaller AIC value is a better fit (R. B. Kline, 2005). AIC can be used to compare models even when they are not nested.

BIC. Bayesian Information Criterion. This is the same as AIC but penalizes for small sample size and extra parameters in the model. Therefore, it is more reliable when sample sizes are small or there are large numbers of parameters (R. B. Kline, 2005).

Like many aspects of LVM, these criteria are controversial, "and will continue to be controversial for the foreseeable future" (Miles & Shevlin, 2000, p. 199). There is a large literature discussing fit indices, and opinion changes regularly.

5.5.2 Limitations to fit indices

The goodness-of-fit (e.g. CFI, TLI) and badness-of-fit (e.g. RMSEA, SRMR) indices described above are not an exhaustive list. Again, it is important to stress

that they should not be regarded as golden rules (Markland, 2007; Marsh et al., 2004), because there remains a fierce debate about using them appropriately. Tests of fit using AFIs can fail "for trivial or ignorable reasons" (p. 860), particularly with large sample sizes (McIntosh, 2007). For example, non-normality in the data, and specification errors (e.g. when two variables should have an arrow connecting them but do not). These would lead to a significant chi-square, which indicates poor fit by the AFI criteria. The failure of AFIs does not state why the model has failed (McIntosh, 2007). Claiming that the model is a good or bad fit based on AFIs alone is dubious (op cit). But it is equally dubious to ignore the AFI by arguing that the statistic is sensitive to sample size (McIntosh (2007) called this a "mantra" of using sample size to justify a poorly fitting model). Others argue that AFIs are part of a toolkit, "a multi-faceted strategy for determining model accuracy" (p. 851). They should not, Markland (2007) argues, be "elevated to the status of golden rules"(p.858).

A note about R-square. R square (in LVM or in multiple regression) provides the proportion of variance accounted for in each dependent variable (and latent variable) by the independent variables, but it is not directly related to model fit. The quality of the model depends on the context, and R-square should not be overinterpreted (e.g., Pitt & Myung, 2002). Researchers use R-square as part of a multi-faceted strategy, taking into account other factors such as theory, individual parameter estimates, the structure of the residual variance, and the implications of applying the model to interventions or across populations (R. B. Kline, 2005).

5.5.3 Parsimony and model fit

A final debate in the LVM literature centres on the publication of well fitting models which are not parsimonious. It is easy to generate a model which fits the data very well, if there are arrows from every variable pointing to every other variable (a saturated model). Similarly, in traditional multiple regression it is easy to increase R-square if there are many variables in the model. But this is not

desirable, because of the principle of parsimony. Parsimonious models are more efficient; they have fewer parameters (Crawley, 2007; R. B. Kline, 2005). There are those who call for a greater focus on the individual parameter estimates for this reason (e.g., R. B. Kline, 2005). However, the specific parameter estimates do depend on the overall fit of the model, so the two are connected (McIntosh, 2007). This means that a poorly fitting model can change the parameter estimates. In this sense, parameter estimates are "subordinate" to the AFIs (McIntosh, 2007).

5.5.4 Consider alternative models and re-specify

There is disagreement among researchers about how to approach model modification, also known as model trimming (e.g. R. B. Kline, 2005). A strict confirmatory approach implies that a model should be tested once, against a set of data. However, researchers frequently remove non-significant pathways from models, "starting with the lowest t-value" (Diamantopoulos et al., 2006) until the fit indices begin to worsen substantially. Models usually have to be re-specified several times before a final model is presented. Important parameters, not specified by the researcher, may need to be added to a model in order to improve its fit. Without them, a model may suffer from specification errors. There are several procedures available to uncover specification errors. Lagrange Multiplier (LM) statistics tell the researcher how much a model would improve if they added a particular parameter to the model (R. B. Kline, 2005, p.148). The greater the value of LM, the better the overall improved fit in the model if that path were added (R. B. Kline, 2005). LM statistics then, can reveal otherwise unnoticed pathways between predictors and criteria[3]. This might be a useful tool for personality-health researchers.

5.5.5 Multiple models

Ultimately, the model is a researcher's choice. And that means it must be guided by theory (R. B. Kline, 2005; Miles & Shevlin, 2000). Simpler theories are consid-

[3]LM statistics are limited to ellipsis A models.

ered preferable. Research programs that begin to fail often do so because over-complex theories have been developed (Gholson & Barker, 1985). As warned by Pitt and Myung (2002), "choosing between competing models is no easier or less subjective than choosing between competing theories" (p. 425). Models are supposed to simplify data — they should be more simple than the raw data itself while minimizing error (Miles & Shevlin, 2000). However, just as it is often profitable to explore multiple working hypotheses (Chamberlin, 1965), it is advisable to test more than one model. The LVM framework allows several plausible models to be tested, since the position of observed and latent variables can be changed. The fit of multiple working models could be equally as good, preventing researchers from relying too heavily on their preferred initial choice (Chamberlin, 1965).

5.5.6 Final step: Apply the model

Very few LVMs developed by psychologists reach the stage of application. This is particularly the case in personality-health models. This is partly because research is focused on finding good measurement and structural models for personality and health variables. This is the "groundwork" of personality-health research. The mechanisms underlying these associations may not be understood, making application and intervention premature. In the concluding chapter, I argue that it may benefit personality-health research to attend to potential applications of models earlier, by thinking about model audience.

Part II

Structural models for health criteria

CHAPTER 6

The HAPPLE study

> *Internet sampling techniques only permit the generation of diverse, not representative, samples* (Best, Krueger, Hubbard, & Smith, 2001, p. 131).

6.1 Introduction

Chapters two to four explained how personality traits, health behaviours, and health outcomes, can be operationalized and measured. Chapter five introduced a general framework for latent variable modelling. In this chapter, the methods and descriptive statistics from the HAPPLE study are presented.

6.2 The HAPPLE study

The Health and Personality Processes: Linked Explored (HAPPLE) study was launched in 2003. This study was first described in Hagger-Johnson (2004), a thesis which contained a limited set of analyses from an initial set of respondents (N = 189) who enrolled during the first 40 days of the study (8th July to 16th August 2004). The aim of the HAPPLE study was to create a large data set containing measures of health behaviours, big five personality traits, health outcomes, and symptom reports; via internet mediated data collection over two time points. This data set could then be used to model both traits and facets,

and test their relationship with a variety of health criteria, including health behaviours. When the HAPPLE study was designed, the internet mediated research (hereafter, IMR) was a relatively new method. It was anticipated that the IMR method would produce sufficient variability on demographic variables (Bailey, Foote, & Throckmorton, 2000; Buchanan, 2000), in addition to high participation rates, low attrition, comparable to higher reliability and little missing data. Preliminary IMR studies had suggested that diverse, if not representative, samples could be obtained (Best et al., 2001). Psychological mechanisms, and decision making processes, are thought not to differ between IMR and paper-and-pencil testing modes (Best et al., 2001). Some of these hypotheses have been borne out by later research. For example, demographic variance has been observed across many IMR studies, often as good as traditional paper-and-pencil studies (Gosling, Vazire, Srivastava, & John, 2004; Rhodes, Bowie, & Hergenrather, 2003). The measures used in the HAPPLE study, and the order in which they were completed, are now described. Not all of the available variables were analyzed in this thesis (e.g. sexual health behaviours, past smoking history, self-examination, and symptom reports). They are available to be modelled in future analyses. A full version of the questionnaire is appended to this thesis as an appendix.

6.3 Methods: The HAPPLE study

6.3.1 Protocol and questionnaire

The protocol for the study was approved by the NHS Multi-Region Ethics Committee for Scotland. It was stated in the protocol that participants should be able to skip any question which they would prefer not answer. Advertisements linking to the study were posted on four IMR research portals: the Psychological Research on the Net[1], Online Psychology Research U.K.[2], Online Social Psychol-

[1] http://psych.hanover.edu/research/exponnet.html
[2] http://onlinepsychresearch.co.uk/

Figure 6.1: Web button advertisement for the HAPPLE study

Figure 6.2: Animated banner advertisement for the HAPPLE study: Frame 1

Figure 6.3: Animated banner advertisement for the HAPPLE study: Frame 2

ogy Studies[3]and the Web Experiment Psychology Lab[4]. In addition, advertisements were posted on Jiscmail discussion lists, voluntary and community sector organization web sites, and on NHS intranet web pages (Hagger-Johnson, 2004). Web site designers from the voluntary and community sector were invited to link to the study using a button (Figure 5.1) or animated banner advertisement (Figures 5.2 to 5.4). The questionnaire was converted to Hypertext Markup Language (HTML) format, using MySQL and Personal Home Page (PHP) language to transfer the data into Structured Query Language(SQL) database format (see Welling & Thomson, 2003). 128-bit encryption via Secure Sockets Layer (SSL), provided by Demon Webhosting Ltd., was used for data encryption between the web questionnaire and the database.

[3]http://www.socialpsychology.org/expts.htm#sother
[4]http://www.psychologie.unizh.ch/sowi/Ulf/Lab/WebExpPsyLab.html

Figure 6.4: Animated banner advertisement for the HAPPLE study: Frame 3

Self reported access to the HAPPLE study. Self reported data on how participants found the study was available (N = 460). Responses were grouped and coded as: an email from the researcher, including announcements on non-academic e-mail lists (108, 23.5%), voluntary and community sector organizations (64, 13.9%), a recommendation (N = 53, 11.5%), Google (N = 51, 11.5%), a friend (N = 44, 9.6%), a link on another web site, not VCS (N = 35, 7.6%), one of the four portals (N = 21, 4.6%), browsing (N = 14, 3.0%), a University list or email (N = 12, 2.6%), a search engine (non-Google) (N = 10, 2.2%), an academic e-mail list, including Jiscmail (N = 6, 1.3%), the yahoo search engine (N = 5, 1.3%), MSN (N = 5, 1.1%), NHS Intranet (N = 4, .9%).

Page 1: Consent form

The consent form ended with five items, designed to ensure that participants met the inclusion criteria for the study:

Healthy volunteer. "I have a long-term illness, health problem or disability, which limits my daily activities or the work I can do" (reverse keyed: "no" required).

U.K. resident. "I live in the United Kingdom (U.K.)" ("yes" required).

Consent. "I would like to participate in this study" ("yes" required).

Informed consent. "I have read the Participant Information Sheet" ("yes" required).

Withdrawal. "I understand that I can leave the study at any time by e-mailing the researcher" ("yes" required).

As shown in the appendix, the initially selected response options were randomized, as a further measure to ensure that participants could not enter the study unless they checked the appropriate response.

Page 2: E-mail address and health behaviours

The second page of the study requested the participants' e-mail address, and contained health behaviour items, described below. To address the problems of poor reliability in the SR scale, described in chapter three, the strategy adopted in the HAPPLE study was to collect additional information on smoking and alcohol use. A set of health behaviour items were taken from the European Health Behaviour Survey (Wardle & Steptoe, 1991).

E-mail. E-mail address was used to invite participants to complete the second wave, at least three months after the first wave. It was also used to match participants' responses across each wave.

HBC. The HBC (R. R. Vickers et al., 1990) was described in chapter three. Minor changes were made to wording, to ensure that the items were appropriate for British participants.

Smoking. Participants were asked "Do you smoke?" ("yes", "no"). For those answering "yes", they were classified as CURRENT smokers. Current smokers were asked "How many cigarettes do you smoke per day?", "How long have you been a smoker? (years)" and "How long have you been a smoker? (months)". For those not classified as CURRENT smokers, they were asked, "have you ever smoked cigarettes for more than one month?" ("yes", "no"). Those answering "no" were classified as NEVER smokers. Those answering "yes" were classified as EVER smokers. EVER smokers were asked "how many cigarettes per day did you used to smoke?", "How long were you a smoker? (years)", "How long were you a smoker? (months)".

Alcohol. Alcohol items were rewording using items from the International Health Behaviour Survey (Wardle & Steptoe, 1991; Ussher et al., 2004) which provide information on number of alcoholic drinks in the previous two weeks, self-ratings of drinking habits and intentions to reduce alcohol intake. Participants were asked "Would you describe yourself as" (a non drinker, a very occasional drinker, an occasional drinker, a regular drinking) and asked

"on how many days over the past two weeks (14 days) did you have a drink?", "On the days that you did drink, how many units did you have, on average?" and "Would you like to reduce the amount that you drink?"

Other. A set of additional health behaviours, not included in the HBC, were included. Four of them refer to sexual health behaviours: "How many casual partners have you had in the last six months?", "Do you use condoms with your main partner?","Do you use condoms with casual partners?", "Have you gone to a GP, hospital or clinic for a test for an HIV test?". Additional items relating to breast and testicular cancer were added from the European Health Behaviour Study (Wardle & Steptoe, 1991; Ussher et al., 2004). These were, "Women only. Do you know how to examine your own breasts for lumps", '"If 'YES', about how many times a year do you examine your breasts for lumps?', "Women only. How long has it been since you had a cervical (Pap) smear test?","Men only. Do you know how to examine your own testicles for lumps?", "If 'YES', about how many times a year do you examine your testicles for lumps?". Participants indicated how well each statement described them ("very characteristic of me" to "very uncharacteristic of me") on a five-point scale.

Pages 3 to 5: IPIP NEO

Participants completed the IPIP representation of the NEO-PI-R (Costa & McCrae, 1992a), hereafter termed IPIP NEO[6](Goldberg, 1999). The IPIP NEO is a self-report inventory consisting of 300 items, which were divided across pages three, four and five of the web questionnaire (100 items per page). Participants clicked a radio button to select one of five options ("strongly agree" to "strongly disagree"). As described in chapter two, the IPIP NEO allows measurement of the Big Five traits and their facets. For example, the Neuroticism dimension measures the facets of Anxiety, Anger, Depression, Self-consciousness, Immoderation and Vulnerability (labelled N1 to N6 respectively). Correlations between

[6]"IPIP" might also refer to 50-item and 100-item questionnaires available, this is usually clear from the description.

Table 6.1: Descriptive statistics for HAPPLE data set: Personality traits

Variable	Alpha reliability	Test-retest stability[5]	Mean	SD	Missing
N	.95	.72	178.88	13.07	138
N1	.84	.83	29.39	7.32	138
N2	.89	.83	29.25	7.85	138
N3	.91	.87	28.56	9.07	138
N4	.79	.83	29.05	6.65	138
N5	.81	.76	30.11	6.96	138
N6	.86	.84	28.20	7.42	138
E	.94	.78	182.13	13.85	138
E1	.90	.82	31.98	8.23	138
E2	.84	.81	30.57	7.12	138
E3	.86	.88	31.53	7.02	138
E4	.61	.70	31.20	4.97	138
E5	.75	.86	29.79	6.66	138
E6	.89	.90	31.44	8.19	138
O	.97	.70	181.67	10.89	138
O1	.90	.87	32.42	8.81	138
O2	.94	.90	33.01	10.82	138
O3	.92	.89	32.18	9.33	138
O4	.87	.87	32.38	7.96	138
O5	.94	.90	32.42	10.19	138
O6	.80	.91	31.53	7.34	138
A	.97	.68	192.99	11.02	138
A1	.88	.86	31.90	7.43	138
A2	.86	.90	31.98	7.84	138
A3	.95	.86	32.81	9.67	138
A4	.81	.86	31.93	7.29	138
A5	.77	.87	30.35	6.46	138
A6	.88	.88	32.18	8.15	138
C	.96	.61	186.16	12.71	138
C1	.93	.90	32.46	8.25	138
C2	.85	.87	30.55	7.46	138
C3	.93	.90	32.83	9.25	138
C4	.91	.85	32.55	8.29	138
C5	.80	.79	30.71	6.20	138
C6	.82	.79	31.09	6.19	138

the traits and facets of the IPIP NEO and the commercial NEO-PI-R are high (range .86 to .99). The IPIP NEO has predictive validity (Goldberg, 1999, p. 18). Test-retest stability is shown in Table 5.1.

Validity scales. Validity scales, designed to correct for response sets such as socially desirable responding, were not included. The rationale for this was based on recommendations in Piedmont, McCrae, Rieman, and Angleitner (2000) and advice in the NEO-PI-R manual (Costa & McCrae, 1992a). These authors argue that self-reports do not enhance validity, for several reasons. First, "Even the most sophisticated validity scales offer very limited guidance about the meaning of test responses." (Piedmont et al., 2000, p. 591). Given that validity scales are self-reports, and so are trait items, they may share method variance, and both may be subject to socially desirable responding. Trait items are not a perfect way to measure traits, "and validity scales will not make them so" (Piedmont et al., 2000, p. 591). The authors of the NEO-PI-R instead emphasized the need to develop well validated inventories, and ensure that participants understand the importance of providing honest answers, by providing clear instructions. In applied clinical and health settings, they argued, clinical judgement is more important than reliance on "validity scales" designed to correct for faking (Costa & McCrae, 1992b).

Page 6: SF-36

The SF-36 version 1 (Ware et al., 1993) was described in chapter four (Ware et al., 1993). It consists of 36 items, producing eight scales which measure different aspects of physical and mental health status. PCS and MCS scores were derived from these scales using the UK factor score coefficients (Jenkinson et al., 1999).

Page 7: Physical symptom checklist

The physical symptom checklist is an 18-item scale designed to measure physical symptoms (Saboonchi & Lundh, 2003): tiredness, daytime sleepiness, bodily tension, insomnia, headache, weakness/dizziness, loss of appetite, back pain,

Table 6.2: Descriptive statistics for HAPPLE data set: Health behaviours

Variable	Alpha reliability	Test-retest stability[a]	Mean or N	SD or %	Missing[b]
Wellness Maintenance and Enhancement	.64	.73	29.77	5.97	19
Accident Control	.61	.81	36.41	11.76	19
Traffic Risk	.77	.79	30.61	7.43	130
Cigarettes per day	NA	NA	3.49	6.98	15
Self reported drinker type (range 1 to 4)	NA	NA	3.16	.85	41
Weekly units	NA	NA	12.66	17.87	41
Would like to reduce drinking (1 = yes)	NA	NA	153	29%	41
Current regular smoker (1 = yes)	NA	NA	157	28%	1

[a]Wave 2 measures were taken at least three months after wave 1. Test-retest correlations for the HBC scales were calculated using all available data from wave 1 (N = 587) and maximum likelihood estimation for missing data at wave 2, which increases reliability of standard error estimation. Covariance coverage ranged from 29% to 31%.

[b]If cigarettes per day was reported as zero, then non-smoker was inferred.

Table 6.3: Descriptive statistics for HAPPLE data set: SF-36 scales

Variable	Alpha reliability	Test-retest stability[a]	Mean	SD	Missing
Physical Functioning	.87	.25	51.97	7.68	157
Role Physical	.76	.26	49.60	9.62	157
Role Emotional	.84	.45	45.08	12.76	157
Energy Vitality	.85	.55	32.82	10.22	157
Mental Health	.67	.67	43.39	9.03	157
Social Functioning	.74	.57	61.02	11.01	157
Pain	.81	.42	50.39	8.96	157
General Health Perception	.80	.86	45.67	10.02	157
Physical Component Summary	NA	.51	53.28	8.81	157
Mental Component Summary	NA	.61	41.84	11.39	157

[a]Wave 2 measures were taken at least three months after wave 1. Test-retest correlations were calculated using all available data from wave 1 (N = 434) and maximum likelihood estimation for missing data at wave 2, which increases reliability of standard error estimation. Covariance coverage ranged from 28% to 29%.

Table 6.4: Comparison of HAPPLE SF-36 scale scores with the general population, by gender.

Variable	U.K. Males'	U.K. Females	HAPPLE Males			Happle Females		
Variable	Mean (SD)	Mean (SD)	Mean (SD)			Mean (SD)		
PF	89.76 (18.78)	86.66 (20.15)	94.78 (9.54)	-5.61	$<.0001$	91.18 (14.68)	-4.85	$<.000$
RP	89.01 (21.09)	85.83 (22.52)	87.60 (26.35)	0.60	.55	85.06 (28.25)	0.44	.66
BP	81.25 (22.21)	76.97 (23.44)	85.12 (20.24)	-2.13	.04	81.72 (18.26)	-4.15	$<.0001$
GH	70.86 (20.29)	71.28 (20.54)	65.96 (19.61)	2.79	.01	64.21 (19.98)	5.73	$<.0001$
EV	60.81 (18.93)	55.91 (19.85)	56.58 (19.54)	2.42	.02	53.62 (19.85)	0.50	.62
SF	84.71 (22.56)	81.33 (23.62)	79.63 (19.25)	2.93	$<.0001$	78.20 (20.18)	2.46	.01
RE	88.08 (19.91)	84.07 (21.79)	72.09 (40.35)	4.48	$<.0001$	65.94 (40.67)	7.33	$<.000$
MH	74.32 (17.24)	70.05 (18.65)	64.24 (15.16)	7.39	$<.0001$	61.96 (15.09)	8.57	$<.000$
PCS	50.63 (9.41)	49.54 (10.40)	54.22 (7.33)	-5.34	$<.0001$	53.23 (8.29)	-6.96	$<.000$
MCS	51.16 (9.34)	49.17 (10.39)	42.35 (11.14)	8.83	$<.0001$	41.12 (11.11)	11.56	$<.000$

muscle pain, stomach-ache, constipation/diarrhea, fever, nausea, cold, allergic complaints, itching, heart palpitations, sweating. Six responses options are presented on a Likert scale (none, very mild, mild, moderate, severe, very severe).

Page 8: Demographic variables

The demographic variables are presented below. Several of these variables were converted, based on visual inspection of the frequency distributions, into categorical variables:

Age. Date of birth, converted into years of age (continuous).

Relationship. Relationship status, recoded into "in a relationship" (1) or "not in a relationship" (0).

Gender. Gender was recoded into "male" (1) and "female" (0).

Ethnic minority. Ethnicity, recoded into "minority ethnic status" (1) or "non-minority ethnic status" (0).

Sexual orientation. Sexual orientation identity, on scale (1 to 7), recoded into "non-heterosexual orientation" (1; exclusively identifying as "gay" or "lesbian") and "heterosexual orientation" (0 Herek, Garnets, & Nolen-Hoeksema, 2007).

Student. Student status recoded into "full or part time student" (1) and "non student" (0).

Employment status. Employment status, recoded into "employed" (1) and "not employed" (0).

Occupation. Participants were asked to report, in open text fields (open-ended items), their job title, job description and industry of work. These were coded into the Standard Occupational Classification (SOC, 2000) codes. The SOC codes were updated in 2000 to reflect recent changes in society. The main categories are:

1. Managers and senior officials.
2. Professional occupations.
3. Associate professional and technical occupations.
4. Administrative and secretarial occupations.
5. Skilled trades occupations.

6. Personal service occupations.
7. Sales and customer service occupations.
8. Process, plant and machine operatives.
9. Elementary occupations.

The Computer Assisted Structured COding Tool (CASCOT)[7] was used to make the coding simpler, quicker and more reliable. This program allows researchers to code responses according to the U.K. standards developed by the U.K. Office for National Statistics. It is suitable for situations like the HAPPLE study, where the item is open-ended, and the researcher must categorize "the huge range of all possible answers to a pre defined set of categories (each category having a unique code)". As described on the Internet page for CASCOT, "quality of coding performed by CASCOT depends on the quality of the input text". However, the software can be considered reliable and valid because "performance of CASCOT has been compared to a selection of high quality manually coded data. The software authors report that 80% of records receive a score greater than 40 and of these 80% are matched to manually coded data".

Deprivation. Area based social deprivation concerns households and neighbourhoods, not just individual differences, which might influence health above and beyond individual level social and economic factors (G. D. Smith, Whitley, Dorling, & Gunnell, 2001). Postal sectors were converted into area based deprivation indices (continuous). Carstairs scores were chosen because they are available for England, Wales, Scotland and Northern Ireland; and because they have been validated (see Morris & Carstairs, 1991). Carstairs scores summarize, as a Z score, four indicators of social deprivation:

Unemployment. Unemployed male residents over 16 as a proportion of all economically active male residents aged over 16.

Overcrowding. Persons in households with 1 and more persons per room as a proportion of all residents in households.

[7] http://tinyurl.com/2pnehe

No car ownership. Residents in households with no car as a proportion of all residents in households.

Low social class. Residents in households with an economically active head of household in social class IV or V as a proportion of all residents in households.

At least the middle two numerical values are required, in order to calculate a Carstairs score (i.e. EH8 9). These data are referred to as postcode sector. Postcode sectors were converted to Carstairs scores in four steps:

1. Carstairs data files were imported into a Microsoft Access database[8].
2. Postcodes were linked to Census wards using the WARD ID (Field 35) in the National Statistics Postcode Directory[9].
3. Postcodes in the HAPPLE data set were validated using the CONVERT tool[10]. Postcodes that could not be validated were treated as missing data.
4. Microsoft Access was used to link the postal sectors, from the validated postcodes, to the Carstairs scores.

Working hours. Hours worked per week.

Education. Educational qualifications (and Scottish equivalents) were converted into a variable capturing educational level (continuous), ranging from 1 to 7, using the Office of National Statistics proposed harmonization categories for educational level[11]. Here, 1 = no qualifications, 2 = other qualifications level unknown, 3 = qualifications at level 1 and below, 4 = trade apprenticeships, GCSE grade A*-C, vocational level 2 and equivalents, 5 = A Levels, vocational level 3 and equivalents, 6 = other higher education below degree level, 7 = degree or degree equivalent and above. Educational attainment is a useful indicator of socio-economic status because, unlike occupation and deprivation, it is largely stable once attained. In contrast, people can move to a more deprived area, or take a lower status job. Education, and any benefits it brings, would still be available to that person: "Educational

[8] http://tinyurl.com/37u33y
[9] http://datalib.ed.ac.uk/EUDL/NSPD.html
[10] http://convert.mimas.ac.uk
[11] http://tinyurl.com/5cvt6j

attainment characterizes a persons life pathway in a single, summary variable. It is typically achieved in the early adult years and remains stable, so there is no question of reverse causality in its association with later outcomes" (Hampson, Goldberg, et al., 2007, p.122). It is a powerful variable associated with many health outcomes, through many possible pathways. It reflects a range of other variables: health related knowledge, problem solving skills, control over one's life, and material resources (Braveman et al., 2005).

Economic inactivity. Economic inactivity comprised those reporting that they were retired (N = 8), students (N = 206), looking after home or family (N = 50), or long term sick or disabled (N = 4). These were recoded into "economically inactive" (1) and "not economically inactive" (0). Responses to this item did not always correspond to the earlier item concerning student status. For example, respondents might indicate they were a student but then not endorse this item in relation to economic activity. A strategy for resolving this issue is described in chapter seven.

A high drop out rate was observed between the pages. As shown in Tables 6.1 to 6.3, this resulted in an increasing proportion of missing data across each page.

Wave two

After three months, participants were sent a maximum of two e-mail reminder messages, inviting them to complete the second wave of the study. Given the high drop-out rate observed between pages of the first wave of the study, the demographic items were moved to the first page of the questionnaire. A matching process was then conducted, to replace missing data on demographic variables for the first wave, using available demographic variables from the second wave. The response rate for the second wave was approximately 28%. The total available proportion of demographic data, after the matching process, is shown in Table 6.5. Demographic data is therefore not available for all participants. The

Table 6.5: Descriptive statistics for HAPPLE data set: Demographic variables

Continuous variables	Mean	SD	Missing
Age (years)	31.15	7.02	284
Education (years)	20.64	2.71	152
Deprivation score (range -1.6 to 3.3)	1.15	3.72	382
Occupation code (range 1 to 9)	4.7	2.60	393
Categorical variables (1 = yes)	**Count**	**Percentage**	**Missing**
Male	134	32%	147
In a relationship	250	60%	148
Ethnic minority status	82	19%	141
Student	183	45%	152
Economically inactive	195	35%	139
Non-heterosexual orientation	161	31%	45

second wave of the study was then used to provide test-retest stability coefficients, displayed in Tables 6.1 to 6.3.

6.3.2 Descriptive statistics

Tables 6.1 to 6.4 show the descriptive statistics for HAPPLE measures, and the proportion of missing data (maximum N = 561). They also show, where relevant, the test-retest reliability of each measure.

6.4 Discussion

There are three key points to make about the HAPPLE data set. First, the proportion of missing data was higher than anticipated, because a large number of participants did not complete all seven pages of the questionnaire. Second, a large proportion of the sample are economically inactive: a group largely comprising students. Third, demographic variables are not available for all respondents because these items appeared at the end of the questionnaire in the first wave. An important limitation of IMR research is that the population sampled may have different characteristics to the general population, and this can affect results. Therefore, the HAPPLE sample was compared to the general population.

6.4.1 How does the HAPPLE sample differ from the U.K. population?

A key strength of the HAPPLE data set is that many demographic variables were measured. It is important to gather data about a sample (Schmidt, 1997) so that population level judgements can be made. Where available, this allows the sample to be compared to the U.K. population, based on the 2001 Census Office of National Statistics (2001). Since sexual orientation is not included in the Census, the U.K. study of sexual attitudes and lifestyles was used (Wellings, Field, Johnson, Wadsworth, & Bradshaw, 1994) for comparison. T tests performed on differences in means confirmed that the sample was significantly older, more educated, less socially deprived, and higher in occupational status, than the U.K.

population (all ps $<.01$). Results from Z tests of proportions showed that, compared to the U.K. population, the HAPPLE study contained significantly larger proportions of: females, those in a relationship, ethnic minorities, economically inactive people (including students), and those with non-heterosexual orientation. Table 6.3 compares the HAPPLE SF-36 scale scores with the normative SF-36 data. T tests were conducted, assuming unequal variances and sample sizes. Males scored significantly higher on PF and BP; significantly lower on GH, EV, SF, RE and MH. Females scored significantly higher on PF and BP; significantly lower on GH, SF, RE and MH. For the PCS, male and females scores were significantly higher. For the MCS, male and female scores were significantly lower. The PCS and MCS score differences demonstrate that the HAPPLE sample is higher in physical health status (as anticipated), but lower in mental health status, than the general population. Figures 6.7 and 6.8 show the frequency distribution for the PCS and MCS scores.

On the generalizability of IMR samples

Given the lack of representativeness observed in the HAPPLE data set, it is worth noting some of the points made by other IMR researchers in this regard. Best, Krueger, Hubbard, and Smith (2001) discussed some problems about the generalizability of IMR research. Chiefly, access is an important factor. They argue it is not appropriate "to make inferences about the psychological mechanisms underlying much broader populations — populations often comprising individuals without internet access" (p. 132). However, there are cases where psychological mechanisms might not differ between those with and without Internet access. They "conclude that the use of Internet samples should be limited to those circumstances in which some demonstrable evidence exists that the decision-making hypotheses being tested are uniformly applicable to the entire population" (p. 132). They call this "coverage error". An internet sample is considered to have sufficient coverage if two conditions are met:

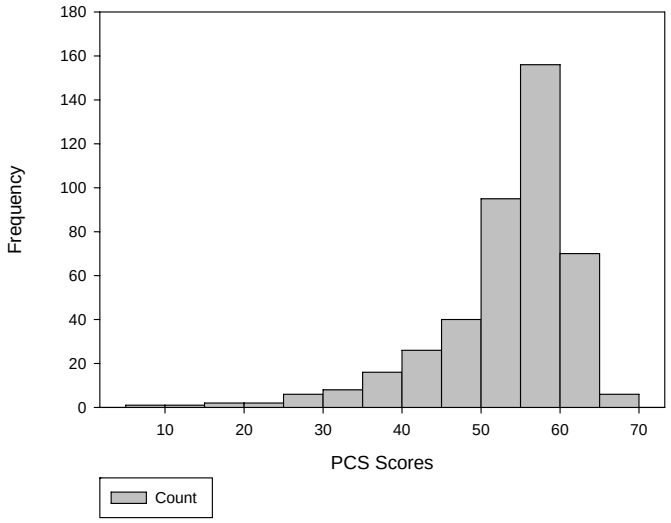


Figure 6.5: Histogram of SF-36 Physical Component Summary T scores, HAPPLE study

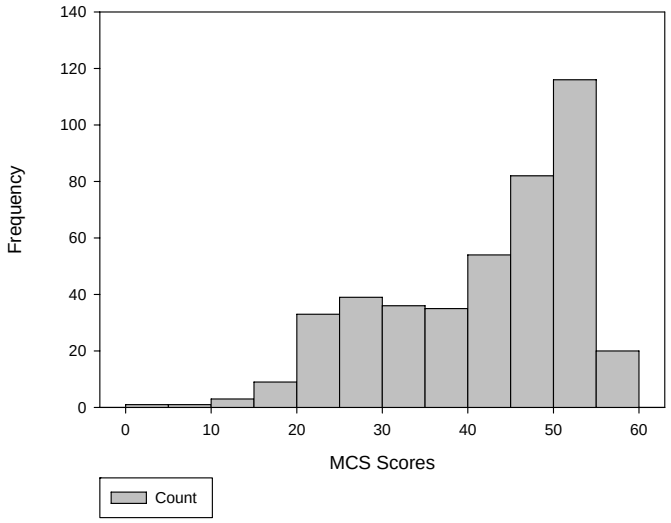


Figure 6.6: Histogram of SF-36 Mental Component Summary T scores, HAPPLE study

Similar mechanisms. Decision-making processes used to generated a given attitude or belief are the same for Internet users as nonusers. If psychological mechanisms and associations are the same (e.g. Conscientiousness is associated with health behaviours), then a representative sample of the internet population is not required. Variability on the two variables is required. Conscientiousness would be expected to be correlated with healthy behaviours in IMR users and in traditional paper-and-pencil samples, as long there was a sufficient range on these two variables. Best et al. (2001) argue that "As long as some variance exists for all the variables in the hypothesized mechanism and the missing variance does not correlate differently than the measured variance, then a diverse rather than a representative sample can be used to infer relationships within the population" (p. 135).

Sampling frames. Since "no support exists for the assumption that a representative sample of Internet users can be drawn" (Best et al., 2001, p. 134), it may be more appropriate to define specific, targeted, sampling frames, in future research. There is no true internet "population". The internet comprises a variety of difference networks, including discussion groups, discussion boards, mailing lists, web sites and social networking sites (Hewson, Laurent, & Vogel, 1996). If a study was targeted at a specific discussion group, for example, then it would be possible to say that all members of that group had an equal chance of being selected for the study.

In summary, Schmidt (1997) emphasised that internet access is "growing exponentially" (p. 274) but noted that "biases are known to exist in the population that frequently accesses the Web. Demographic information about Internet users is available from a number of sources" (p. 274). Similar to my point above about specific facets of the Internet, he argued "the validity of WWW research is likely to be strongest for research domains that target specific populations" (p. 274, see also Hewson et al. (1996) who described the "targeting" of internet respondents). Although the HAPPLE study provides demographic diversity, more so than many traditional studies, it is different from the general population: as are

other IMR studies (Best et al., 2001). However, there is variability on the important measures of interest (demographic variables, personality traits, health behaviours, and SF-36 scores). It is therefore appropriate to use the HAPPLE data set, provided that the limitations of generalizability are acknowledged. To my knowledge, there is no evidence that the *mechanisms* linking personality traits, health behaviours, and health outcomes, would differ between Internet users and non-users. There may, however, be confounding by variables such as socio-economic status (SES).

6.4.2 Conclusion

Having introduced the HAPPLE data set, and the variety of variables included in the study, the next chapter aims to illustrate how this data can be modelled within the generalized latent variable modelling framework. Unlike multiple regression, both the traits and facets of the IPIP NEO can be modelled simultaneously, and mediators can be modelled that comprise latent variables, in relation to multiple dependent variables. Two potential mechanisms (educational level and multiple health behaviours) that may underlie the association between Conscientiousness and health outcomes, are considered.

CHAPTER 7

Conscientiousness and health criteria

> *Educational attainment and health behavior mechanisms...explain some, but not all, of the influence of childhood personality on adult health status*
> (Hampson, Andrews, Barckley, Lichtenstein, & Lee, 2006, p. 124).

7.1 Introduction

One key problem with the field of personality-health research is that the size, or indeed, the existence, of associations is highly variable (e.g., Ashton, 1998; Bogg & Roberts, 2004). The bandwidth-fidelity dilemma is one reason for this (J. Hogan & Roberts, 1996a; Ones & Viswesvaran, 1996; Paunonen, 1998; Paunonen & Ashton, 2001; Pruessner et al., 1997; Mershon & Gorsuch, 1988). It refers to the trade-off between fidelity (quality of information) and bandwidth (complexity of information obtained). Ones and Viswesvaran (1996) raised this concern over a decade ago, arguing that psychometrics has traditionally focused on high fidelity, low bandwidth measures. J. Hogan and Roberts (1996b) dismissed this concern, arguing that "the characteristics of the predictor ought to be driven by the characteristics of the criteria...trade-off between fidelity and bandwidth is unavoidable, and we see no 'dilemma' to be resolved. We regard the 'controversy' as a straw man"(p. 628). However, evidence that facets have predictive validity for health criteria, above and beyond traits (Paunonen, 1998; Paunonen

& Ashton, 2001), suggests that the dilemma is relevant to personality-health research. The choice of bandwidth may be at least partly responsible to the large variability in the published size of Conscientiousness-health behaviour associations (Friedman & Booth-Kewley, 1987) and for personality and health outcomes, such as physical symptoms (Watson & Pennebaker, 1989). Long or short versions may increase or decrease the size of the association (Paunonen & Ashton, 2001; Paunonen, 2003). It has been argued that facets and traits should be measured, because this "increases the prediction of relevant criteria, even when those criteria are broad and complex (Paunonen, 1998; Paunonen & Ashton, 2001)" (de Raad & Goldberg, 2002, p. 11). Health behaviours are precisely this kind of criteria (broad and complex), and provide an interesting challenge for those interested in maximizing the predictive validity of traits and facets.

Traditional approaches to traits and facets

The bandwidth-fidelity dilemma has previously been approached in two main ways (Hagger-Johnson, 2004). First, multiple correlations (Paunonen, 1998). Second, multiple regressions, with alpha reduced to control for type I error rates (Paunonen, 1998; Paunonen & Ashton, 2001; Wasylkiw & Fekken, 2002). The first method provides an index of the effect size for facets (e.g. multiple correlation), the second an index of the percentage of variance accounted for in criteria and the percentage change when facets are added to models that already contain traits. For example, Paunonen (1998) compared facets from the Personality Research Form (PRF) inventory and traits from the NEO-FFI. He used partial correlation and set correlation (similar to canonical correlation) and compared the effect size for each, on multiple outcomes. Facets had a multiple correlation of .84, traits only reached .68. He showed that the facets explained 24.2% more variance than traits alone. He next used multiple regression, and assessed the percent change in R-square when facets were added to a model that already contained traits. R-square is a percentage of variance in the criteria, accounted for the independent variables in a multiple regression (Miles & Shevlin, 2000; Tabachnick & Fidell, 2000). I explored the utility of the R-square approach in

multiple regression during my MSc thesis (Hagger-Johnson, 2004), concluding that facets do offer additional predictive validity for the HBC scales, as assessed by increases in R-squared.

The over-emphasis on R-squared.

The R-squared approach has several weaknesses, some of which are acknowledged by (Paunonen, 1998). It is difficult to know if the advantage of facets is due to: (1) the participants; (2) the model specified by the researcher; (3) the choice of variables; (4) the dimensionality of the outcomes. Indeed, the dimensionality problem has sparked a hypothesis that predictors and criteria need to be matched at equal dimensionality (Wasylkiw & Fekken, 2002). There is also the fifth problem that increases in variance explained by regression-type models capitalize on a tautology — it will always be true that variables added to a model will increase R-square (Goldberg, 1993). There is also a risk of type I errors, because comparing traits and facets usually involves more facets than traits. There are three strategies to resolve this fifth problem. First, prediction equations can be cross validated on another sample (Goldberg, 1993). Second, equal number of traits and facets can be chosen to study (Ashton et al., 1995). Third, the type I error rate can be controlled, so that when variable numbers are unequal, the p value is adjusted (Paunonen, 1998). Overall, personality-health researchers seem to prefer multiple regression. My claim in this chapter, is that factorial complexity can be met using the generalized latent variable modelling framework, for different kinds of health criteria.

A focus on Conscientiousness

Conscientiousness was chosen for three main reasons: (1) general traits, especially Conscientiousness, have been widely researched elsewhere including recent factor analytic (Roberts et al., 2005) and meta analytic (Bogg & Roberts, 2004) studies; (2) there has been little investigation of Conscientiousness facets in relation to health behaviours; (3) further research is needed into the mechanisms

that might underlie associations between Conscientiousness and both physical and mental health. The modelling strategies described in this strategy could be applied to other traits and other criteria.

7.1.1 Methods

Data

The data selected for this example comprised the first 345 participants from the HAPPLE study who completed the HBC and IPIP NEO pages of the questionnaire.

Variables

Conscientiousness. Conscientiousness items from the IPIP NEO were parcelled into observed variables representing the six facets (C1 = Self-efficacy, C2 = Orderliness, C3 = Dutifulness, C4 = Achievement-striving, C5 = Self-discipline, C6 = Cautiousness). Reliability coefficients are shown in Table 6.1.

Health behaviours. Three scales from the HBC were selected (WME = Wellness Maintenance and Enhancement, AC = Accident Control, TR = Traffic Risks). The Substance Risk scale was excluded due to its poor reliability (see Table 6.2).

Age. Age was measured in years.

Sex. Sex was coded as male (1) and female (0).

7.1.2 Results

Modelling

A confirmatory factor analysis was tested in which a latent variable (Conscientiousness) was defined by its six facets (C1 to C6). The full information maximum likelihood estimator was used to model missing data. A direct pathway from C5

Self Discipline to a health behaviours latent variable (HB) was specified, to illustrate the unique variance from the C5 facet, after controlling for the general C trait. Latent variables cannot be predicted by observed variables. However, a dummy latent variable can be used with a regression coefficient fixed at 1 pointing to the observed variable, which has an error variance fixed at zero. This provides an equivalent solution (Tabachnick & Fidell, 2000). Sex and age were included as covariates. In the path diagram, solid arrows represent predicted pathways. In this example, the facets have been parcelled from the appropriate IPIP NEO items (Goldberg, 1999) The model was estimated using the full information maximum likelihood estimator. Missing data on outcome variables is permitted, because there is a missing data theory and estimation strategies used by Mplus, for downstream variables (Muthén & Muthén, 1998–2007). In latent variable modelling, maximum likelihood estimation can be used to model missing data, on mediators and outcomes, without the need for imputation (Muthén & Muthén, 1998–2007). A good fit for the hypothesized model was not supported by the data (χ^2 = 81.34, $p < .0001$, CFI = .84, RMSEA = .06, SRMR = .05). In order to improve the fit of the model, it was necessary to allow Cautiousness to cross-load onto the health behaviours latent variable, and allow the errors of C3 and C6 to correlate. Although this pathway is not consistent with facets as predictors of health behaviours, it is possible that Cautiousness and the HBC are describing item content similar enough to warrant this cross-loading. I regressed C6 on C (β = -.03), C6 on HB (β = -.34), then C6 on C (β = -.57) and HB (β = -.50). This indicates the possibility of a suppressor effect. This pathway should therefore be interpreted with caution. These pathways were suggested post hoc by Lagrange Multiplier tests (modification indices), and are shown as dashed arrows in Figure 7.1. These modifications improved the fit of the model substantially (χ^2 = 55.02, p = .03, CFI = .93, RMSEA = .04, SRMR = .04). Although the model failed the chi-square test of exact fit, this is known to be a very strict test. It suggests that the proposed model fits perfectly, which is an unreasonable assumption in most behavioral research (Goffin, 2007). The model had excellent fit by the RMSEA

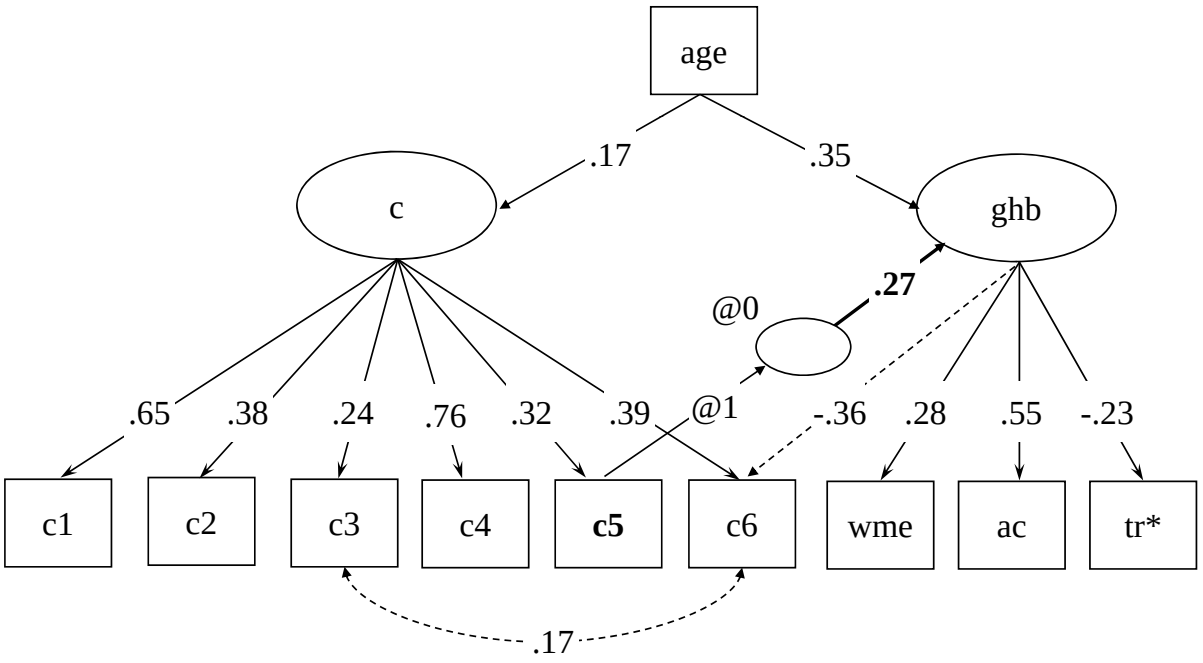


Figure 7.1: Path diagram showing latent variable model for Conscientiousness facets and the Health Behaviour Checklist (HBC), with age as a covariate. Age = age in years, C1 = Self-efficacy, C2 = Orderliness, C3 = Dutifulness, C4 = Achievement-striving, C5 = Self-discipline, C6 = Cautiousness, C = general Conscientiousness, WME = Wellness Maintenance and Enhancement, AC = Accident Control, TR = Traffic Risks, GHB = general health behaviors, * = reverse scored. Solid lines represent hypothesized pathways, dashed lines pathways determined by Lagrange Multiplier tests.

and SRMR criteria (L. T. Hu & Bentler, 1998; R. B. Kline, 2005). The Mplus syntax is shown in the appendix.

7.1.3 Discussion

Principal findings

The results show that the bandwidth-fidelity dilemma can be resolved without resorting to a choice between broad or specific measurement of traits and criteria. It is possible to model trait and facet variance, and structural pathways to health criteria. Traits and facets can be included in the same model, and that predictive validity can be demonstrated for facets for health criteria, "even when those criteria are broad and complex" (de Raad & Goldberg, 2002, p. 11). The model also shows that health behaviours can be modelled as a single latent variable. The main implication of this study is practical. LVM is an appropriate technique to model personality facets. By modeling general Conscientiousness as a latent variable, it is possible to extract pathways from specific facets to various health criteria. By extracting the general Conscientiousness variance from its six facets, it was possible to estimate a pathway from C5 Self Discipline to health behaviors, which represented the "non-conscientious" contribution of that facet.

Strengths and weaknesses

In this model, the C5 pathway to the health behaviour latent variable was specified in advance. Although this extracts variance unique to C5, other facets could be associated with health behaviours. Ideally, LM statistics would suggest that adding pathways from facets to criteria, would improve model fit. This would provide more convincing evidence that facets have predictive utility. However, the technique described is for illustrative purposes. Other measures and other data sets could model traits and facets in a similar way, and use LM statistics to improve model fit. Modification indices that suggested adding pathways from facets to criteria, would suggest that they have incremental validity over and above any traits already in the model. A second limitation is that the SR scale

in the HBC was excluded from this analysis, due its poor reliability. This meant that smoking and alcohol use, important health behaviours, were not included in the model.

Meaning of the study

The choice of traits or facets need not "be driven by the characteristics of the criteria" (J. Hogan & Roberts, 1996b, p.628). The results presented here suggest that traits and facets can be included in the same model. It is misleading to assume that the criteria should drive the choice of bandwidth in the predictor. Broad and specific measures are related in complex ways, and the breadth of content in the criteria does matter. There can be associations between traits and general health behaviour factors; or between general traits and narrow health behaviour factors. Facets have predictive validity, even for "broad and complex" criteria (de Raad & Goldberg, 2002, p. 11), and selecting them does not need to be driven by the characteristics of the criteria. Pathways from general C, or from specific facets, can be differentially associated with health criteria.

Next steps

Parcelling items into facets means that bandwidth and fidelity can exist in the same model, even when a moderate sample size is available. de Raad and Goldberg (2002) noted that "a representation combining broad and narrow constructs, offers a good compromise between efficiency (or parsimony) and fidelity. Moreover, the theoretical structure is much clearer and likely to contribute to an increased understanding of the functioning of personality dimensions" (p. 11). Using LVM brings additional features, such as the ability to add pathways that run between traits and facets. This may contribute to making modelling clearer, as well as theory. In the next section, I extend this approach by considering Conscientiousness, its facets, and health outcomes. Health behaviours, and educational level, are considered as potential mediators of the association between this trait and physical and mental health outcomes.

7.2 Conscientiousness and health outcomes

7.2.1 Introduction

Associations have been found between Conscientiousness and health outcomes (Hampson, Goldberg, et al., 2007), including mental health outcomes (O'Cleirigh, Ironson, Weiss, & Costa, 2007). The mechanisms underlying these associations are less well understood. Educational level and health behaviours explain part, but not all of this association (Hampson, Goldberg, et al., 2007). No study, to my knowledge, has considered Conscientiousness, a wide range of health behaviours, educational level, and both physical and mental health outcomes, in the same model. The aim of this section is to demonstrate that the generalized latent variable modelling framework could be used to explore mediating pathways between traits and health outcomes. A secondary aim is to demonstrate that the HBC can be improved by removing the SR scale and replacing it with items that assess quantities of cigarette smoking and alcohol use. By extending the health behaviour construct to encompass a wide range of behaviours, and extending the criteria to include both physical and mental health outcomes, this may further our understanding of the influence of Conscientiousness on health.

7.2.2 Methods

The hypothesized model is shown in Figure 7.2. The aim was to fit a single factor to the HBC scales, extending them by including cigarette smoking and alcohol units, and the remaining items from the SR scale. Socio-economic status is defined by education, Carstairs deprivation scores, and occupational level. Health behaviours and SES mediate the association between Conscientiousness (defined by its six facets) and health outcomes (physical and mental). Age and sex are additional covariates, not shown in the Figure.

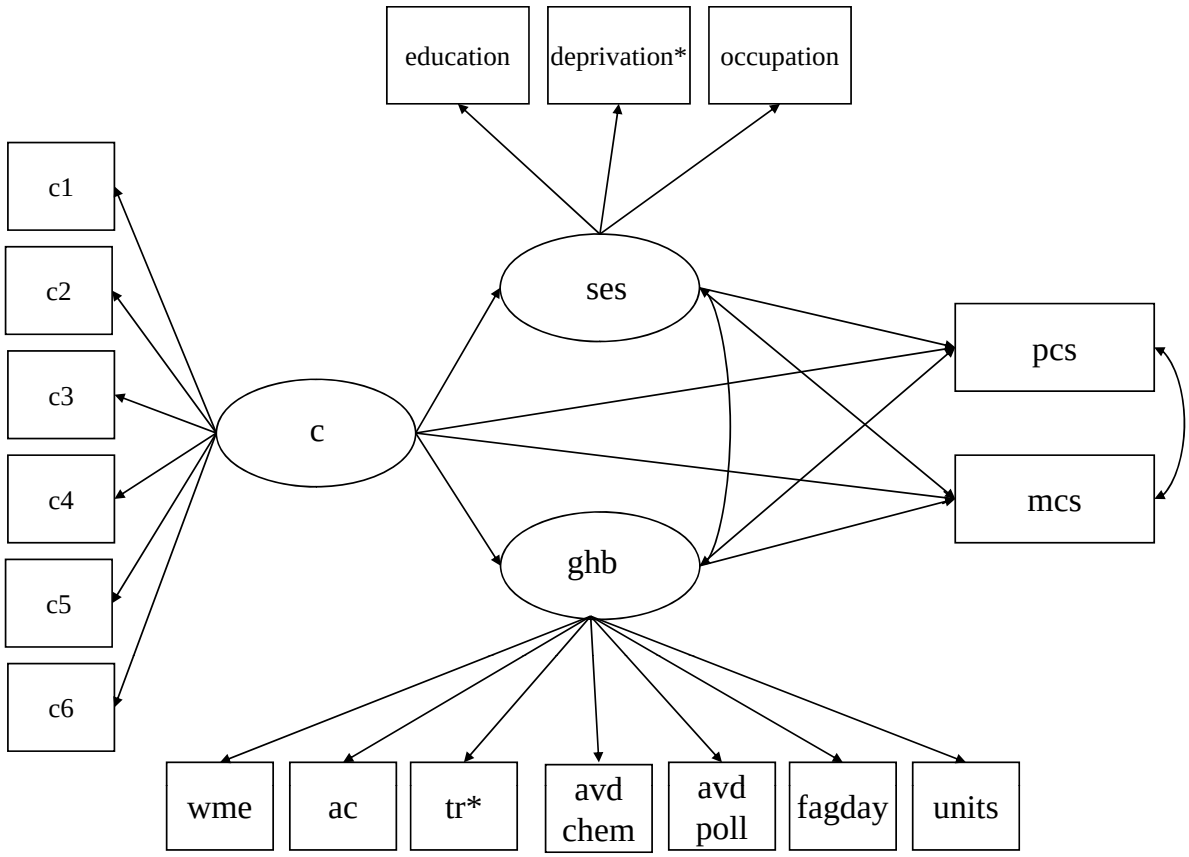


Figure 7.2: Path diagram showing a hypothesized model: SES as a meditating variable between Conscientiousness, health behaviours and health outcomes. C1 to C6 = Conscientiousness facets. C = general Conscientiousness, WME = Wellness Maintenance Enhancement, AC = Accident Control, TR = Traffic Risk, avdchem = "I don't take chemical substances which might injure my health", avdpoll "I avoid areas with high pollution", fagday = number of cigarettes smoked per day, units = units of alcohol consumed per day, ghb = general health behaviours, education = educational level, deprivation = Carstairs deprivation score, occupation = Standard Occupational Classification code, SES = socio-economic status, PCS = Physical Component Summary, MCS = Mental Component Summary. * = reversed scored. Deprivation is multiplied by -1 so that higher scores indicate higher SES.

Data

In the models that follow, Conscientiousness facets are covariates, and health behaviours are mediators (upstream and downstream variables). This means that missing data is not permitted on traits, but it can be modelled for health behaviours. For this reason, a database was created in which all those participants with data available on the IPIP NEO were retained for analysis. This excluded participants with missing data on all the personality items. For example, those who completed only the first page of the questionnaire (health behaviours), were excluded. Missing data on covariates is excluded casewise. Since age, sex and educational level are also covariates, the sample size varies as a function of which covariates are included in the model.

Variables

Conscientiousness facets. Conscientiousness facets were included as above (C1 to C6).

HBC. HBC scale scores were calculated to allow one missing item per scale. For TR, two items specific to drivers were excluded due to high levels of missing data, addressing the problems highlighted above. Three items were reversed keyed so that higher scores on all TR item reflected healthy traffic related behaviours. This was designed to help interpretation of the health behaviour latent variable.

Cigarette smoking. Cigarette smoking was coded as 0 for non-smokers, and the number of cigarettes smoked per day or smokers.

Alcohol units. Alcohol was converted into units consumed in the previous two weeks, and coded 0 for non-drinkers.

Physical health. The PCS scores from the SF-36 were used to measure physical health status.

Mental health. The MCS scores from the SF-36 were used to measure mental health status.

Age. Age was recorded in years.

Sex. Sex was coded 1 (male) and 0 (female).

Socio-economic status (SES). Education, occupation and deprivation (see chapter six) were converted into Z scores, and deprivation was multiplied by -1 so that all indicators reflected higher SES. These two strategies are designed to help convergence, which can fail if indicators are measured on very different scales (Boniface & Tefft, 1997; Muthén & Muthén, 1998–2007).

Economic inactivity. As described in chapter six, a greater proportion of participants were economically inactive than originally expected. These comprised those reported that they were retired (N = 8), students (N = 206), looking after home or family (N = 50), or long term sick or disabled (N = 4). These were recoded into "economically inactive" (1) and "not economically inactive" (0). Responses to this item did not always correspond to the earlier item concerning student status, which was coded as "full or part time student" (1) and "non-student" (0). To resolve this issue, the maximum of these two variables was taken to indicate economic inactivity. This is the most conservative strategy.

7.2.3 Results

Modelling

The analysis was restricted to those who did not identify as economically inactive or as students. The rationale for excluding these participants is that they may not have reached their final level of educational attainment, creating a ceiling effect below the degree level on the education variable. It would not be appropriate to control for differences in educational level for people who have not yet left the education system. The maximum likelihood estimator was used to estimate

missing data. It is robust against non-normality, which was present for the PCS and MCS scores (see Figures 5.6 and 5.7).

SES model. Preliminary analyses showed that it was not possible to model SES as a latent variable, with education, occupation and deprivation as indicators. This is most likely due to the poor covariance coverage (available data for covariances between variables), resulting from the large amounts of missing data on occupation and deprivation. The covariance coverage ranged from just 12.9% to 21.1% for occupation, and 12.9% to 60.8% for deprivation. In contrast, the covariance coverage for education ranged from 21.1% to 60.8%. Given that the least amount of missing data was available for education, the decision was made to use education in further analyses (models 5 to 10). In addition, the decision was made not to conduct multi-group analysis for each gender, because of the resulting small sample sizes in each group (N = 140 females, 87 males). Educational level was tested as a meditator first, followed by health behaviours. The model was then modified to remove non-significant pathways. In all models, missingness is not permitted for covariates, which are excluded casewise. Models were tested sequentially, described below and fit indices listed in Table 7.1.

Model 1. The first step in mediation is to show that the independent variable (Conscientiousness) is associated with the dependent variable (health outcomes)(Miles & Shevlin, 2000). PCS (not significant) and MCS (β = .43, $p < .001$) were therefore regressed onto Conscientiousness. Age (β = .19, p = .04) and male (not significant) were also regressed onto Conscientiousness. Multiple group analysis was deemed inappropriate, given the small sample size and availability of demographic data (N = 140 for females, N = 87 for males). Two sets of correlated errors were permitted, within the Conscientiousness factor.

Model 2. The second step in testing for mediation is to show that a mediator (educational level), is associated to the independent variable). Educational level was therefore regressed on Conscientiousness (β = .15, p = .05). This parameter was just below the threshold for statistical significance.

Table 7.1: Goodness-of-fit statistics for the latent variable models containing health behaviours and years of education

Model number	Chi square	d.f.	*p*	CFI	TLI	AIC	BIC[a]	RMSEA	SRMR
Model	Chi square	d.f.	p	CFI	TLI	AIC	BIC	RMSEA	SRMR
1	79.41	27	$< .001$	92	87	-6073.85	-6034.15	.10	.06
2	74.48	22	$< .001$	92	87	-4957.71	-4920.47	.11	.07
3	144.36	34	$< .001$	83	74	-6439.25	-6367.07	.13	.07
4	57.84	31	$< .001$	82	75	-5071.22	-5042.30	.07	.06
5	173.65	93	$< .001$	90	87	-8886.20	-8799.37	.07	.07
6	155.75	82	$< .001$	91	88	-9227.56	-9149.69	.07	.07
7	206.42	115	$< .001$	89	86	-10335.50	-10232.29	.06	.07
8	213.17	116	$< .001$	88	85	-10338.88	-10232.29	.06	.07
9	178.07	98	$< .001$	90	88	-9492.06	-9403.03	.06	.07

[a]Sample size adjusted.

Model 3. The third step in mediation is to show that the mediator is associated with the dependent variable, after controlling for the independent variable. Neither PCS nor MCS were associated with education in this model. Therefore, educational level is not a mediator of the association between C and health outcomes. Education was included in subsequent models as a covariate.

Model 4. A model was tested in which the WME, AC and TR scale from the HBC formed a latent construct, together with cigarette smoking, alcohol units and the remaining items from the SR scale of the HBC. This was a good fit by the RMSEA and SRMR criteria, but not by the CFI and TLI criteria. Age was associated with the general health behaviour factor (β = .50, $p < .001$).

Model 5. Having already demonstrated that Conscientiousness is associated with MCS (mode1, education models above), the second step is to test whether the mediator (the general health behaviour factor) is associated with the independent variable (Conscientiousness). The parameter estimate for this pathway was large but the fit of the model was not adequate. Allowing correlated errors between TR and C6 improved model fit. The parameter estimate for the pathway was large (β = .37, $p < .001$). Although these do not belong to the same construct, it is possible that shared method variance results in unexplained correlations between the HBC and IPIP-NEO.

Model 6. The third step is to test whether PCS and MCS are associated with the health behaviour factor, controlling for C. The model was a poor fit to the data when covariates were included, and many modification indices were suggested. To focus on the relevant pathway, the model was re-run without covariates. The pathway was not significant, suggesting that health behaviours could not plausibly mediate the association between Conscientiousness and health outcomes.

Model 7. The decision was made to include the direct and indirect pathways in the model, with covariates, so that mediation could be formally tested.

Model 8. As a formal test of mediation, the chi-square from model 7 (206.42, 115 d.f.) was compared with a nested model in which the direct pathway from Conscientiousness to MCS was constrained to zero (213.17, 116 d.f.). The

chi-square difference of 6.75 with 1 degree of freedom was statistically significant at the $p < .01$ level. This indicates that the more restrictive model is a significantly worse fit than the restrictive model. However, the true test of mediation is the indirect effect, defined as the reduction in the direct pathway once the mediator is included in the model. Given that health behaviours were not associated with MCS, this is a more conservative strategy. The indirect test can be tested in Mplus using the MODEL INDIRECT feature. The reduction in the direct pathway was not statistically significant ($p = .74$). Therefore, health behaviours did not mediate the association between C and MCS.

Model 9. Non-significant pathways were removed one at a time, starting with the least significant. The final model is shown in Figure 7.3. It shows that there is a direct pathway from conscientiousness to mental health ($\beta = .28$, $p < .001$) that is no accounted for age, education, or health behaviours. Significant pathways were also observed from conscientiousness to general health behaviours ($\beta = .36$, $p < .001$), age to health behaviours ($\beta = .45$, $p < .001$), age to Conscientiousness ($\beta = .18$, $p = .04$), Conscientiousness to educational level ($\beta = .14$, $p = .05$). Age predicts C, which in turn, predicts educational level, multiple health behaviours and mental health outcomes. The factor loading for cigarette smoking was no longer significant. The model accounted for 8% of the variance in MCS scores and 40% of the variance in general health behaviours.

In summary, physical and mental health status was influenced by variables other than age, sex, educational level, Conscientiousness and multiple health behaviours. The final model was a good fit to the data, and explained a modest proportion of the variance in PCS and MCS scores. Neither educational level nor multiple health behaviours mediating the association between Conscientiousness and MCS.

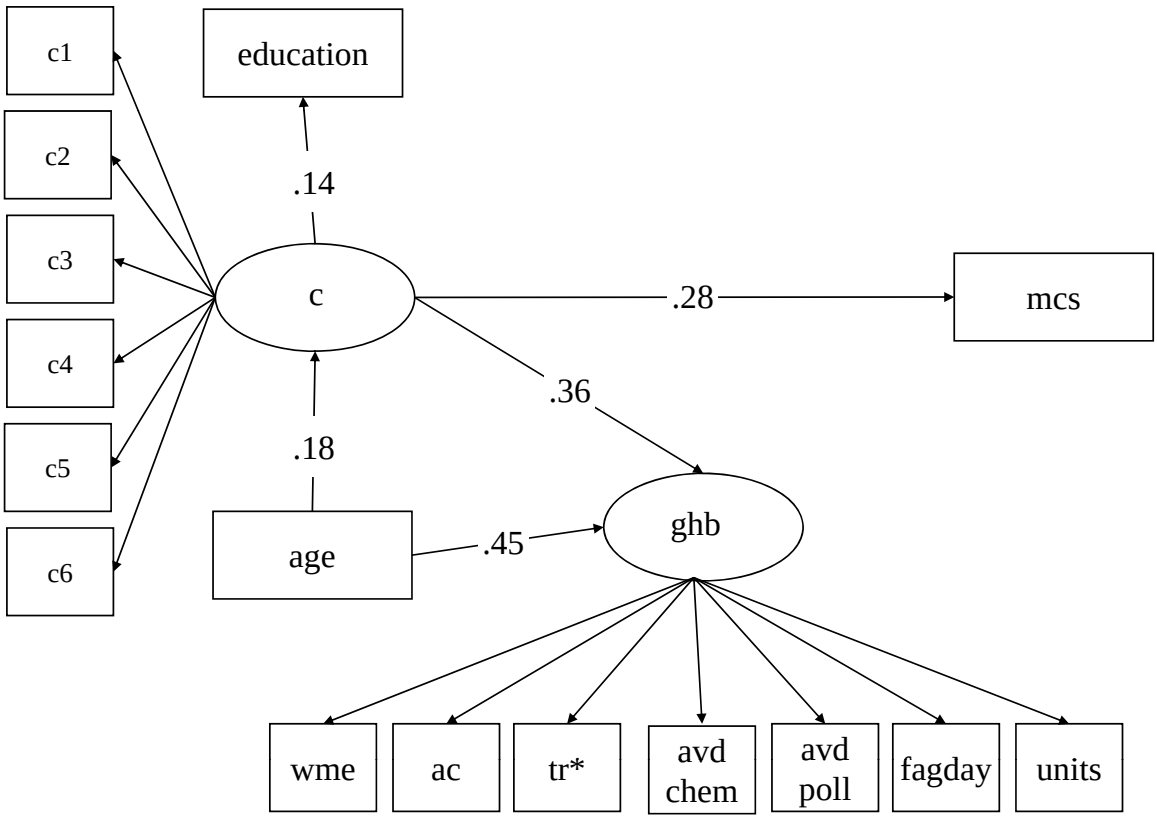


Figure 7.3: Path diagram showing pathways between Conscientiousness facets, multiple health behaviours, age and education to SF-36 Mental Component Summary scores. MCS = SF-36 Mental Component Summary, C1 to C6 = Conscientiousness facets. C = general Conscientiousness, WME = Wellness Maintenance Enhancement, AC = Accident Control, TR = Traffic Risk, avdchem = "I don't take chemical substances which might injure my health", avdpoll "I avoid areas with high pollution", fagday = number of cigarettes smoked per day, units = units of alcohol consumed per day, ghb = general health behaviours.

7.2.4 Discussion

Principal findings

The results show that in a sample of economically active adults in the HAPPLE data set, Conscientiousness was associated with higher MCS scores. The association found was not mediated by education or health behaviours. Second, the results show that a general health behaviour latent variable can address limitations of the HBC inventory, by removing items relating to drivers, and including more detailed information and cigarette smoking and alcohol use.

Advantages over traditional multiple regression. Given that this technique is not possible in standard multiple regression, it is a further example of how the generalized latent variable modelling framework can provide additional insights for personality-health research. The framework can be used to test more than one dependent variable (e.g. PCS and MCS), mediators that are latent variables (e.g. the general health behaviour factors) and covariates (e.g. age, educational level). Multiple regression is rather limited in its capability to handle more than one outcome measure, latent variables, and mediating variables (Miles & Shevlin, 2000). The associations are not causal relationships, but can be considered suggestive of potential pathways that could be explored in future research, and considered alongside similar findings in the nomological network (Cliff, 1983).

Strengths and weaknesses

The model extends the findings of the previous section, which showed that facets and traits can be modelled simultaneously. It illustrates that mediating variables, between traits and health, can include latent variables. A key strength was that a single factor could be fitted to the HBC scales. Concerns about the poor reliability of the SR scale in the HBC were addressed by recording the number of cigarettes smoked, and units of alcohol consumed, rather than the Likert response items from the HBC. These variables capture quantities of substances consumed, which may prove to be a more fruitful approach in the future. The remaining items from

the SR scale were added to the health behaviour latent variable. Missing data for items relating to driving were addressed by including only non-driving related items from the TR scale. Importantly, the results suggest that the HBC scales may not require two to four dimensions. In model 9, the health behaviour latent variable accounted for 16.1% of the variance in WME, 35.1% for AC, 9.0% for TR. This shows that a moderate proportion of the variance in each HBC scale is explained by a general factor. However, it also shows that health behaviours are also explained by variables and pathways not included in the model. The factor may have arisen from shared method variance. Future research should consider using multiple methods for health behaviour measurement, perhaps with more than one health behaviour factor. The main limitations of this analysis were: (1) SES could not be modelled as a latent variable; (2) education was not a significant mediator of the relationship between C and health outcomes; (3) only a moderate proportion of the variance in MCS scores was explained; (4) characteristics of the HAPPLE sample may have influenced the parameter estimates. These points are now discussed in turn.

SES. Unfortunately, there were large proportions of missing data for the SES indicators. Postcodes was the item with the highest rate of missing data, resulting in relatively few Carstairs deprivation scores. It was a requirement of ethical approval for the HAPPLE study that participants could skip any items they did not want to answer. They may have felt that a postcode could identify them, and therefore chosen to leave this field blank. A model in which SES was defined by years of education, occupational social class, and deprivation, did not converge. However, educational attainment was available for most participants, and is a valid proxy for adult SES. Hampson, Goldberg, et al. (2007) have argued that is "characterizes a person's life pathway in a single, summary variable" (p. 122). The decision was made to adopt education as a proxy for SES, for this analysis. A second problem, related to creating an SES variable, was that a large proportion of the sample were economically inactive, including students. This reduced the available sample size considerably.

Education. Education is a reasonable indicator of SES (Fernander et al., 2005; Mink et al., 2004), and is used by other personality-health researchers as a proxy for SES (Hampson, Goldberg, et al., 2007). However, it should be noted that education is an imperfect indicator of SES. Even among non-students, educational attainment may be influenced by age, because older respondents tend to have less years of education. Age and education were often correlated in research conducted between thirty and forty years ago, but recent evidence suggests "these variables are now much less correlated with one another" (Holbrook, Krosnick, Moore, & Tourangeau, 2007, p. 339). This is because education has been made available to the whole population, whereas in the past further education was less available. In the US, the correlation dropped from -.20 in 1968 to .05 in 2002 (Holbrook et al., 2007). However, these findings suggest that age and education should not be confounded in newer surveys. This could be modelled in the HAPPLE data set, by including an interaction term between education and age.

Gender. Multiple group analysis for males and females were not performed on the data, because the sample size was small, particularly in the male group. However, Hampson, Goldberg, et al. (2007) found that in their model of traits, health behaviours, education and health outcomes, only one parameter differed significantly between males and females. Females, but not males, who were less Agreeable as children, were more likely to smoke.

Variance explained. The model did not explain a large proportion of variance MCS scores. No association was found for PCS scores. This may reflect the fact that PCS scores were generally higher, and MCS scores lower, than the U.K. population norms (see chapter six). These measures were skewed (see Figures 6.7 and 6.8). However, maximum likelihood estimation was used, which is robust against non-normality of dependent variables (Muthén & Muthén, 1998–2007, p. 8). Despite this, a potential limitation of the HAPPLE data set is that it comprises healthy volunteers, with SF-36 scores that differ from the general population. The SF-36 is designed, first and foremost, to measure the impact of treatments or deterioration in health related quality of life. Although the PCS scale has

variance throughout the healthy population, the MCS may be more amenable to connecting with upstream variables.

Self selection. A limitation of the HAPPLE data set is that self selection resulted in a far greater proportion of students than expected. Another limitation related to self selection is that SF-36 scale scores differ from the general population. Participants' SF-36 scores could be influenced by the internet mediated assessment (c.f. McCue et al., 2006), or because their physical or mental health influenced their decision to take part. A randomized cross-over trial of SF-36 questionnaires, conducted online and offline, could clarify why these differences have been observed. It may also be useful to ask participants what their motivation for taking part in a study was. Despite these limitations, a large number of demographic variables were measured in the study. Minority groups selected to take part, in greater proportions than would be expected from a U.K. wide study. Although the sample is not representative, it is diverse. Further insights could be gained by bringing additional demographic variables into other models, using this data set.

Comparison to other studies

The association between Conscientiousness and MCS stands in contrast to Straten et al. (2007), who recently found no association between this trait and SF-36 scale scores. In that study, Neuroticism, Extraversion, Openness to Experience and Agreeableness *did* show an association, which was particularly strong for mental health outcomes: MH and MCS. Associations between these other big five traits, and MCS scores, were not tested in the current analysis. However, Straten et al. (2007) sampled patients with mood and anxiety disorders, and although these diagnoses were controlled for, differences in sampling between that study and the HAPPLE study may account for the differences in results. The results also differ from Hampson, Goldberg, et al. (2007), who found a direct pathway from Conscientiousness to (physical) health status (β = .18). Health status was measured differently, as a latent variable, with body mass index as an indicator, alongside

the general health item and items measuring "bodily pain, physical role functioning, work functioning, emotional role functioning" (p. 123) from the SF-36. Therefore, differences in how physical health status was operationalized in that study, may account for differences in results. Hampson, Goldberg, et al. (2007) also found an association between Conscientiousness and education, which was not replicated here. Finally, the results are consistent with Ware et al. (1993) in that PCS and MCS are were uncorrelated. Taken together, the results of all these studies highlight the need to consider a wide range of self reported health outcomes, meditators, and to describe demographic characteristics of samples in detail.

Meaning of the study

Conscientiousness is associated with better mental health outcomes among healthy volunteers. This association is not mediated or explained by educational level or by a wide range of health behaviours. Other variables, not measured in the HAPPLE study, may account for the association. Higher Conscientiousness was associated with higher educational level, higher age and healthier behaviours. There is scope for exploring the reasons behind this association in this, and other, data sets. It goes some way to providing clues about how this trait might be related to mental health outcomes. Other studies have found that Conscientiousness is associated with physical health outcomes (e.g., Friedman et al., 1995; Hampson, Goldberg, et al., 2007) and mental health outcomes (O'Cleirigh et al., 2007). Education and health behaviours do not fully mediate these associations, although they do operate as partial mediators in some data sets (Hampson, Goldberg, et al., 2007). Taken together, the findings suggest that personality-health researchers need to look beyond health behaviours and educational attainment in order to explain why Conscientiousness is related to health criteria. The findings also suggest that measuring many different kinds of health behaviours, improving content validity, does not necessarily remove the association between traits and health outcomes. It may be useful to model multiple health behaviours, but

we cannot assume that any trait-health association is simply an artefact of unmeasured health behaviours (O'Cleirigh et al., 2007). Furthermore, the fact that a latent variable *can* be fitted to multiple health behaviours, may not convince other researchers that there is any utility is doing so. In epidemiology, there is a long-standing tradition of controlling for the effects of confounding factors, rather than modelled them explicitly (Shenkin et al., 2004). This issue is considered in the next chapter.

Next steps

The fact that the association between Conscientiousness and physical and mental health remained after considering age, educational level and multiple health behaviours, suggests that something else underlies this association. Other studies have come to similar conclusions. O'Cleirigh et al. (2007) found that Conscientiousness predicted immune system deterioration in HIV disease, even after controlling for health behaviours, adherence to medicine regimes, and mental health. Hampson, Goldberg, et al. (2007) found that health behaviours and educational level were associated with physical health, but there was much unexplained variation. The wider implication is that biological variables may better explain why Conscientiousness is associated with health. The question of how Conscientiousness influences health outcomes, remains open. There is very little research considering the psychobiological mechanisms that involve this trait. Stress reactivity processes, governed by the HPA axis and cortisol production, have been proposed as candidate mechanisms (O'Cleirigh et al., 2007). Cortisol is considered in section three of this thesis, for this reason.

CHAPTER 8

The big four and survival

> *Whereas much is known about the prevalence of single health risk factors and their associations with demographic characteristics including pairwise associations between behaviors and other lifestyle-related health factors, only a modest literature addresses the relationships among multiple lifestyle-related health factors or the clusters of such factors and their demographic correlates* (Glasgow, Goldstein, Ockene, & Pronk, 2004, p. 26).

8.1 Introduction

Chapter seven demonstrated that multiple health behaviours could be modelled as a single, continuous, latent variable, in the HAPPLE data set. However, the fact that a latent variable can fit the data is the start, rather than the end, of construct validity. A more immediate aim, is to demonstrate predictive validity, for outcome beyond self reports (Matthews, Deary, & Whiteman, 2003). Latent variables are not widely used in epidemiology to study health behaviours (Muthen, 1992). One exception is a study by Dean and Salem (1998) who presented results from a study in which seven health behaviours were combined into a behavioural index. They chose to model lifestyle as a continuous latent variable with categorical indicators, a restricted form of the Rasch model. The Rasch model is essentially a confirmatory factor analysis where the indicators

are categorical variables. It requires moderate to high correlations are required for all indicators (Embretson & Reise, 2000). They concluded that the model did not fit, that "behaviours have quite distinct meanings that should be studied separately" (Dean & Salem, 1998, p. 197) and that "risk behaviours do not represent sufficiently similar types of phenomena to form an additive scale of risk taking" (Dean & Salem, 1998, p. 199). These conclusions have gone unchecked in the literature. The lack of popularity of latent variables of health behaviour is also apparent in the relatively few citations made to the original HBC items (R. R. Vickers et al., 1990). It is therefore worth asking if any practical utility, in using latent variables of health behaviours, can be demonstrated. Is there any mileage outside personality-health research, in this approach? This chapter aims to answer that question, by showing that survival analysis, a popular technique in epidemiology, can be combined with the latent variable approach to multiple health behaviours.

8.1.1 Multiple health behaviours

When the focus of research is on health behaviours, and the aim is to predict or explain them, they are usually considered individually. Laaksonen, Lahelma, and Prttl (2002) note that "the majority of research has concentrated on one behaviour at a time" (p. 225). This is partly convenience, partly because interventions too address one behaviour at a time, and is partly a relic of the limitations of traditional statistical modelling, which encourages a focus on one dependent variable. Treating health behaviours as outcomes and looking for explanatory variables is not necessarily a simple way to understand them. A vast body of research relates individual behaviours to socio-demographic variables, and several biopsychosocial variables. There are multiple hypotheses about how the mechanisms related to the behaviours. "While some of the hypotheses are behaviour-specific others may apply to a wider range of behaviours" (p. 225). This has led to a large literature in which separate measures, models, hypotheses, and theories are created for each behaviour. It may turn out that grouping multiple

behaviours into latent variables makes the overall picture *less* complex. Following Laaksonen et al. (2002), I use the term "lifestyle approach" (p. 230) to refer to the practice of studying multiple behaviours (lifestyle) and the relationship of lifestyle to other variables.

8.1.2 An interdisciplinary problem

Multiple behaviours are not a new research topic. Indeed, there is widespread understanding and discussion of their existence, across several disciplines. However, there has been little cross fertilization of ideas, and a proliferation of terms has occurred. Addictive behaviours researchers speak of "polyproblem individuals" (Christo et al., 2003). Epidemiologists have defined "health-related lifestyle, which refers to health behaviour as a whole" (Laaksonen et al., 2002, p. 229). Social researchers are generally aware that multiple health behaviours are socially patterned (Jarvis & Wardle, 2006). Psychologists have identified and studied multiple health behaviours in several ways, using a varied terminology, such as dimensions of risk behaviour (Booth-Kewley & Vickers, 1994), polyproblem individuals (Christo et al., 2003) and a risky personality for health behaviour (Donohew et al., 2000). Across research as a whole, there is understanding that multiple health behaviours increases mortality and morbidity, and therefore so do predictors of these behaviors, and there is understanding that the effects may be synergistic not necessarily additive. However, there has been no resolution about how best to model them. Can drawing on models developed in personality-health research, and models explored earlier in this thesis, offer any insights?

8.1.3 The Health and Lifestyle Survey (HALS)

The HALS is a two-wave prospective study, and is available from the U.K. Data Archive. A large number of variables, including the big four health behaviours, were measured in 1984, and the sample was reasonably representative of the

U.K. population at this time point. It contained more females, fewer single people and fewer older women than the general population. The study comprised an interview in the home and a second visit, where physical measurements and a personality questionnaire (H. Eysenck & Eysenck, 1964) were administered. The second wave took place in 1991 (Cox, 1995). Despite being broadly comparable to the 1991 Census data, HALS2 was not representative of the general population and has no data for those aged 18 to 24. Therefore, it is not used in the analyses presented below. As described above, no change was observed in patterning by social class although all behaviours were more healthy seven years later. Non-manual occupations, home-owners and those in full-time work were over-represented (Boniface, Cottee, Neal, & Skinner, 2001). Deaths have since been recorded at the NHS Central Deaths Registry, up to 2005 (Cox, 2005).

8.1.4 HALS in previous research

Health behaviours in the HALS study have been examined by various authors. Bartley, Fitzpatrick, Firth, and Marmot (2000) used the data set in conjunction with a later data set, the Health Survey for England (conducted in 1993). They noted that attempts to intervene on CVD risk have focused on the big four behaviours. However, the social patterning of CVD risk has increased, rather than decreased. Could this reflect changes in the social patterning of the big four? The authors examined the patterning of health behaviours in relation to occupational social glass. All of the behaviours were patterned so that healthier behaviours were associated with higher levels of occupational social class. No change was observed from 1984 to 1991 in the patterning by social class. Behaviours were somewhat more healthy, but this reduction was observed in all social classes. However, they did not consider the correlation between multiple behaviours and whether this was related to social class. Nor did they consider the change in the correlation between 1984 and 1991 for the HALS. Describing the prevalence of single behaviours can be misleading if those behaviours are correlated (Laaksonen et al., 2002).

The big four as a factor

The correlation between the big four was considered in a sample of males from the HALS study (Boniface & Tefft, 1997). Interestingly, exercise was positively correlated with smoking, alcohol and fat intake. As a healthy behaviour, this is surprising because the other three behaviours are unhealthy. The authors did not discuss possible reasons for this negative association. A single factor (lifestyle) was a good fit to the data in both 1984 and 1991. Deaths data were not available, but a cardiovascular disease (CVD) risk index score was derived from age, blood pressure and waist-hip-ratio. This was related to the lifestyle factor. The results suggest that lifestyle is related to CVD risk. Given that the deaths data is now available up to 2005, it is pertinent to ask if this can be verified by considering CVD related deaths. The departure point for the modelling described in this chapter takes Boniface and Tefft (1997) as a platform for studying lifestyle as related to actual mortality risk.

8.1.5 Survival analysis

Survival analysis is an appropriate technique to study predictive mortality of health behaviours. In many studies, the focus of interest is the time to a particular event (e.g. death). However, death will not be observable for all participants, since many will be alive at the end of a followup period in a study. In the HALS study, many of the participants were still alive in 2005. Because their survival time is not yet known, these data are unobserved. All we know if they survived at least as long as the study period up to the follow-up point. A special kind of latent variable is required to capture this information, called censoring. Censoring occurs when the variable is missing artificially by the end of the follow-up period, before death occurred (Kirkwood & Sterne, 2001). Survival analysis, the name given to models of survival data, is a well known latent variable model (it has been available since the 1970s), particularly in the health sciences and epidemiology. It is another example of how standard techniques would not be

appropriate (rank methods for death would be possible if all subjects were followed to death). It is less well used in personality research, with several recent exceptions (Shipley et al., 2007; Weiss & Costa, 2005). A second type of latent variable is often needed if people drop out of the study. This missing data needs to be modelled, but is not necessary for HALS because all deaths are flagged at the NHS central deaths registry.

8.1.6 The current analysis

It is important to extend Boniface and Tefft (1997) by testing the predictive validity of their lifestyle factor for all-cause and cause-specific mortality. The big four are associated with CVD risk (Bonow, 2002; Mokdad et al., 2004; Pronk et al., 2004). Such a model would lend support to the notion of modelling multiple health behaviours as a latent variable, and address concerns about shared method variance artifactually creating the factor.

8.2 Method

8.2.1 Participants

Following Boniface and Tefft (1997), several sets of exclusions were applied. The relationship between health behaviours, and their relationship to mortality, may differ among people with existing physical illnesses and those taking medications or on special diets. The analysis was restricted to males aged 40 years or higher, without medical diets or hypertensive treatment. Those with heart disease, blood pressure and diabetes were also excluded. It is noteworthy that researchers have analyzed the HALS data without applying exclusions. Shipley et al. (2007) excluded 1868 participants who were missing from the second home visit and 1711 with missing data on variables used in their model. This left 5424 (2991 males, 2433 females) participants who were analyzed. In contrast, Boniface and Tefft (1997) used a strict exclusion criteria so that there analysis was "limited

to males aged 40 years or over who had not reported CHD, diabetes, antihypertensive treatment or dietary therapy in 1984" (N = 1295).

8.2.2 Variables

The Big Four

Fat intake. Fat intake was provided by the authors of an existing study where self-reported dietary behaviours were converted into estimated grams of saturated fat per week (Tefft & Boniface, 2000).

Alcohol. Units of alcohol consumed per were recorded as non-drinker (0), drinker but not in the previous week (1), 1 to 4 units (2), 5 to 8 units (3), 9 to 16 units (4), 17 to 33 units (5), 34 to highest units (6).

Exercise. Analysis of frequency distributions for sporting activities suggest that activity could be dichotomized into active (1) and inactive (0) participants.

Smoking. The number of cigarettes smoked per day was recoded as non- and ex-smokers (0), 1 to 15 (1), 16 to highest (2).

CVD risk factors

Waist-hip ratio. Waist-hip ratio is a well-established risk factor for CVD (Bartley et al., 2000) and is superior to BMI at predicting cardiovascular mortality (Dagenais et al., 2005). It is calculated by dividing waist (cm) by girth (cm). In the HALS data, this measure has already been shown to predict all-cause and cause-specific mortality. Waist-hip ratio was not available for participants with missing data for the second home visit.

Blood pressure. Lowest diastolic and lowest systolic blood pressure were recorded.

Time related variables

Age. Age is recorded in years, at 1984. Age of death is also recorded, for those participants who died up to 2005.

Death at 2005. Deaths are recorded at the NHS central deaths registry and were last updated in 2005 (Cox, 2005).

Cause of death. Cause of death was recoded into death from cardiovascular disease (1: ICD codes 390-459 inclusive) and no death from cardiovascular disease (0). Deaths from all causes were coded "1" and other values were coded "0" (death-all). Deaths from cardiovascular diseases were coded "1" and other values were coded "0" (death-cvd). The underlying cause was used, not secondary causes often provided on the HALS deaths certificates (Shipley et al., 2007). Age of death represented the time-to-event variable. Censoring was applied to those participants who were still alive in 2005. In the data file, age in 2005 replaced the age of death for those participants still alive. Used in conjunction with censoring information, this ensures that the time-to-event variable contains age information for all of the cases.

8.3 Results

8.3.1 Confirmatory Factor Analysis

A confirmatory factor analysis is presented as a measurement model for the lifestyle and CVD risk latent variables. Lifestyle is defined by saturated fat intake, alcohol intake, activity and cigarette smoking in 1984. CVD risk is defined by waist-hip ratio, systolic blood pressure and diastolic blood pressure. Observed variables were standardized to have zero mean and a unit standard deviation, to assist convergence (Boniface & Tefft, 1997). Both factors were regressed on years of age. Correlated errors, proposed by modification indices, were permitted within constructs but not between constructs. Non-complete

Table 8.1: Descriptive Statistics for the HALS study variables

Variable	Mean (SD)	Mean (SD)
	Males	Females
Age (years)	43.46 (17.05)	43.59 (16.90)
Fat intake (grams per week)	760.88 (271.23)	531.49 (180.10)
Alcohol (1 to 6)	3.11 (2.06)	1.61 (1.55)
Active (%)	42.4%	35.3%
Smoking (0, 1 or 2)	.60 (.81)	.50 (.74)
Waist-hip ratio	.91 (.06)	.80 (.06)
Systolic blood pressure	128.12 (15.89)	121.47 (18.35)
Diastolic blood pressure	77.43 (10.96)	73.16 (11.85)

covariance coverage ranged from 58.6% to 96.8% (males) and 58.8% to 95.1% (females), due to missing data on fat intake and physical measurements at the second home visit. Missing data was estimated using the maximum likelihood estimator. Modification indices did not suggest that adding pathways from age to observed covariates would significantly improve the fit of the model, in both genders. The model for males was a good fit to the data by several criteria (χ^2 = 178.92, 16 df, $p < .001$, CFI = .96, TLI = .93, RMSEA = .06, SRMR = .04). A non-significant chi-square would indicate good fit but is rarely observed with large sample sizes. The model was females did not converge. Inspection of the results suggested that the factor loading for fat intake was negative, suggesting that fat intake was negatively correlated with the other three indicators (χ^2 = 42.42, 16 df, $p < .001$, CFI = .99, TLI = .99, RMSEA = .03, SRMR = .02). Seemingly, the big four do not form a latent construct in the same way for both genders.

8.3.2 Survival analysis

Having established a well-fitted measurement model, the next stage was to bring these factors into a Cox proportional hazards regression (Larsen, 2005, e.g.). Covariates are used to predict survival time to follow-up, which can be categorical or continuous, observed or latent. Age of death up to 2005 was used to convey information about time to event, and age at 2005 was recorded for participants who were still alive. Lifestyle and CVD risk were allowed to correlate. Maximum likelihood with robust standard errors and a numerical integration algorithm (two dimensions of integration) was used to estimate the model. The model was estimated separately for males and females, and separately for all-cause and cause-specific mortality. The all-cause model for males (N = 1307, LL = -14451.73 , AIC = 28957.46, BIC = 2907.20) showed a significant pathway from lifestyle to risk of death (β = 1.45). This pathway was not significant in the cause-specific model (N = 1307, LL = -12055.66, AIC = 24169.31, BIC = 24227.28), suggesting that lifestyle is not associated with cardiovascular deaths, or that this association could not be detected in the data. The pathway from CVD risk to survival was significant in this model. In females, the lifestyle factor was changed so that the

fat intake factor loading was negative. CVD risk, but not lifestyle, was associated with all-cause mortality (N = 1369, LL = -13387.50, AIC = 26832.99, BIC = 26892.30) and cause-specific mortality (N = 1369, LL = -11460.22, AIC = 22978.45, BIC = 23037.76). Therefore, lifestyle was only predictive in the model predicting all-cause mortality in males. This model is shown in Figure 8.1. The parameter estimates are standardized for factor loadings and regression coefficients involving continuous variables. For pathways leading to all-cause mortality, the estimates are log odds. The correlation between lifestyle and CVD risk, and the pathway from CVD risk to mortality, are both constrained to zero because they were not significant.

Interpretation

The hazard of death refers to the rate of death, the parameter estimates can be interpreted as the hazard per unit change in the predictor variable. This is sometimes called the death intensity (hazard of death) per unit change, or per standard deviation change, in the predictor variable. The regression coefficients therefore describe the loglinear regression of survival time on the covariates age, lifestyle and all-cause mortality risk. The pathways from the predictors to survival are calculated by exponentiation of the coefficients. The odds of death from all causes were 1.05 times higher for each additional year of life. Expressed differently, ten additional years of age increase death intensity by 22.31%[1], conditional on lifestyle and CVD risk. The influence of lifestyle is greater. The odds of death from all causes are 4.26 times greater per standard deviation increase in the lifestyle factor. The odds are calculated by taking the exponential of the logit (log odds) coefficients. This final model is shown in Figure 8.1. The non-significant pathways were constrained to zero in this model.

[1] exp(10 x .05 - 1) x 100% =

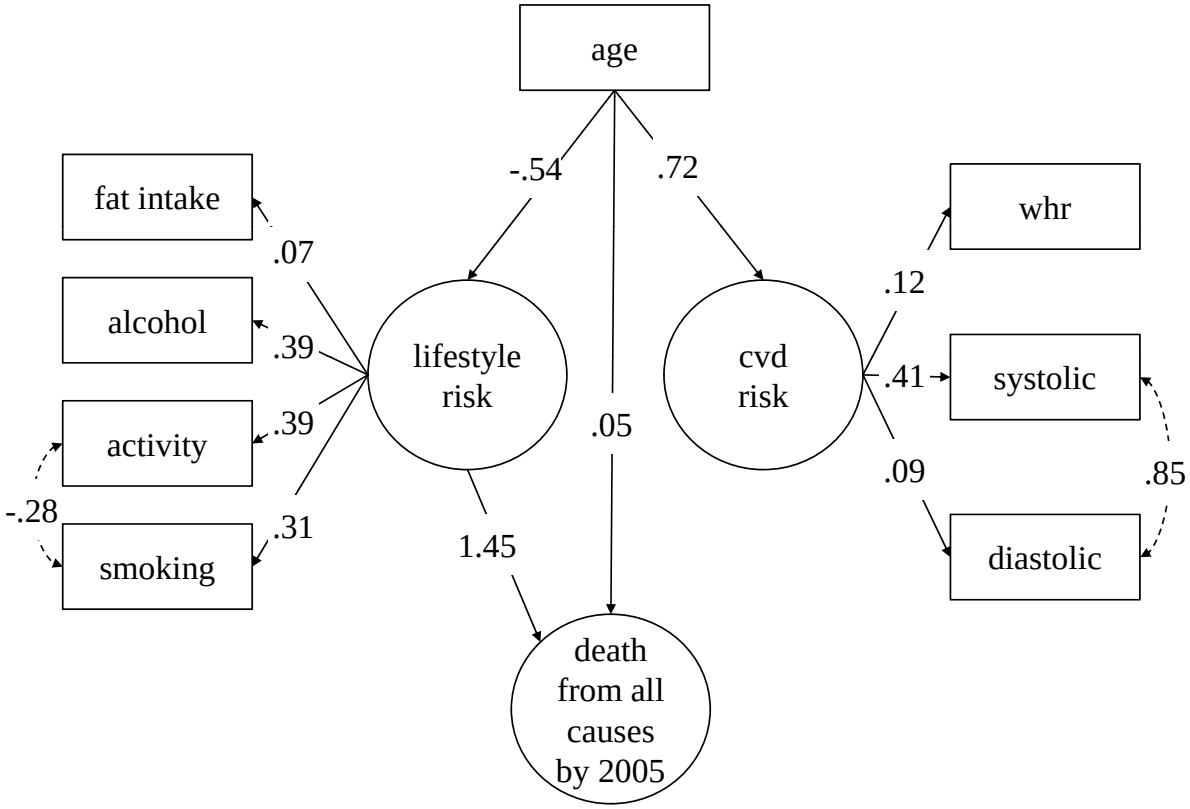


Figure 8.1: Path diagram showing age, lifestyle and CVD risk factors predicting all-cause mortality in males.

8.4 Discussion

The results illustrate that multiple health behaviours have predictive validity for all-cause mortality in males. Lifestyle was not significantly associated with cardiovascular deaths in either gender.

8.4.1 Strengths

The results address the limitations of the analysis presented in the last chapter. The HAPPLE study was essentially mono-method, relying on self reports of health behaviours and health status. Age of death is an objective variable,

which does not share method variance with the self reports in the HALS study. In addition, there is no question of reverse causality, since death is an unambiguous outcome measure. The results replicate and extend Boniface and Tefft (1997) who did not have available the CVD deaths up to 2005. The results show that an assumption of a continuous underlying variable for lifestyle defined by continuous observed indicators, in tenable. Since survival analysis is part of the generalized latent variable modelling framework (Asparouhov, 2006), a further strength is that it can be adopted alongside the existing models preferred by trait researchers. Trait researchers have tended to neglect ellipsis B models. They may reach wider audiences in other disciplines if they are able to demonstrate that structural pathways from traits to health behaviours predict survival, over and above traditional risk factors for death. Although not predictive in all contexts, lifestyle does predict all-cause mortality in males in this data.

8.4.2 Limitations

Limitations of the analysis include the small number of covariates that were included in the model. The HALS data contains many other variables which may be related to mortality, not modelled here. Given that the model requires numerical integration, it was not feasible to include a large number of variables. As the number of integration points increases, convergence becomes less likely because the estimator becomes unstable. It would be interesting to explore other covariates, such as personality traits, in future work. It will be important to consider interactions between latent variables and time, because the proportional hazards assumption could have been violated without these terms. This assumption is more difficult to evaluate in the latent variable context (Larsen, 2005), although interactions between the factors and time can be included in the parametric part of the hazard function (Muthén & Muthén, 1998–2007). Overall, the model illustrates that health behaviour covariance can be modelled in relation to survival. It does not illustrate an exhaustive set of possibilities, or the best model for predicting survival. These multiple avenues are left open for future study. There are many ways in which the variables of this model could related to each other and

to survival. The relatively smaller number of cardiovascular deaths may explain why the factor was not predictive of cause-specific mortality.

8.4.3 Strengths and limitations in relation to other studies

It is not always possible or sensible to model multiple health behaviours as a latent variable. Dean and Salem (1998) found that a Rasch model was not appropriate for categorical variables for eight health behaviours. They concluded that creating a scale for multiple health behaviours was inappropriate. The predictive validity of a health behaviour factor suggests that their conclusion, "behaviours have quite distinct meanings and should be considered separately" (Dean & Salem, 1998, p.195), is not warranted. A Rasch model is equivalent to a confirmatory factor analysis where the indicators are categorical and the factor loadings are probabilities based on the latent variable (Muthén & Muthén, 1998–2007). In situations where a factor cannot be formed, more than one factor may be required (Boniface & Tefft, 1997). Previous research has suggested that to be fully comprehensive, between two and four dimensions are probably necessary (R. R. Vickers et al., 1990). Where the aim is to predict morbidity or mortality, not all behaviours will have mechanisms that tie them to disease. The big four, since they influence the cardiovascular system, have substantively important mechanisms that warrant their inclusion.

Alternative approaches

It may be preferable to model the health behaviours as separate entities but allow them to correlate (Hampson, Goldberg, et al., 2007). Epidemiologists tend to adjust or control for health behaviours which are correlated. Correlated predictor variables are often considered problematic in survival analysis, because collinearity can cause problems with interpretation of the coefficients. It can result in suppressor effects, where the sign of a coefficient is reversed (Mackinnon, Krull, & Lockwood, 2000). One solution is to remove correlated predictors from the model entirely. The approach taken in this chapter acknowledges, rather

than controls for, the health behaviour covariance. Covariance is interesting in its own right and should not be treated as a nuisance artefact. Clearly, there is a need to clarify which behaviours are correlated, which are negatively correlated, and it what contexts these patterns occur. Research into multiple health behaviours now faces a similar task that faced trait researchers. There is also a need to develop a measurement model that researchers can use to locate health behaviours. This is likely to differ for males and females.

8.4.4 Next steps and future research

Clear avenues for future research are proposed. First, it would be useful to bring Neuroticism and Extraversion into the model. No association was found between Neuroticism and mortality when Shipley et al. (2007) adjusted the association for confounding factors, including smoking, alcohol and activity but not saturated fat intake. It may be useful to test a model in which Neuroticism predicts the big four as a factor, which predicts survival in turn. A model in which lifestyle mediates an association between Neuroticism and cardiovascular mortality may fit the data. Neuroticism could increase risk of all four behaviours, although the literature here is equivocal. Unfortunately, Conscientiousness in not available in the HALS data, which was one of the reasons for creating the HAPPLE data set. In future studies, there is a clear need to include the big five and monitor cause specific mortality. Second, another important area for future research is the influence of exclusions on the results. The original lifestyle model involved careful inclusions of participants without CHD, diabetes and medication, for example. In contrast, Shipley et al. (2007) did not exclude these participants from their analyses. Given that associations between traits and health criteria can sometimes differ in patient groups compared to healthy controls, the impact of decisions to exclude on results must be studied. Third, it is important to find out why the factor structure for lifestyle is different for females. Other variables that change the relationship between health behaviours, such as SES,

gender and age, could moderate the factor or influence its relationship to survival. Greater number of CVD deaths may become available in the future, increasing the potential for predictive validity.

8.4.5 Conclusion

The role of gender in multiple health behaviour covariance should be clarified. Socio-economic status is an important variable that should also be considered. Occupation, deprivation and education have been found to predict health behaviour and health outcomes in other models but the associations are not consistent, and SES is rarely modelled as a latent variable. One problem here is that the association between behaviour is partly determined by their prevalence. As noted by Laaksonen et al. (2002), "having two unhealthy behaviours is influenced by the prevalence of each of the individual unhealthy behaviours, examining only one of these proportions at a time may be misleading" (p. 227). If smoking are alcohol are more prevalent in low SES groups, they the association between them can be increased because they are more prevalent. There are higher rates of smoking and drinking together, if there are higher rates for each behaviour. If SES is related to unhealthy behaviour, it does not necessarily follow that SES explains their covariance. Covariance, then, depends on both the prevalence of behaviour and the strength of their association (Laaksonen et al., 2002). SES is a substantive variable, that should be incorporated into future models. Finally, it would be useful to explore the possibility of health behaviour types on survival. Health behaviour types can be modelled using latent class analysis, which is a technique growing in popularity. It proposes that there is unobserved heterogeneity in the data, and that two or more latent classes exist. Mean levels of different health behaviours may differ in each latent class. The debate about whether traits are continuous or categorical may extend to health behaviours, which can be approached from either modelling technique. There is potential here for communicating the results from personality-health models to

audiences that are trained to think categorically, such as clinicians (B. P. Chapman et al., 2007). When models reach the stage of application, it may be more feasible to target classes of individuals, rather than dimensions of behaviour.

Part III

Structural models for the diurnal cortisol profile

CHAPTER 9

Cortisol

> *Circadian rhythms are a potentially important indicator of the regulatory competence of stress response mechanisms because they reflect the capacity of a system to turn on and off appropriately* (S. Sephton & Spiegel, 2003, p.323).

9.1 Introduction

Cortisol is a stress hormone that could be central to understanding health inequalities (Kristenson, Eriksen, Sluiter, Starke, & Ursin, 2004; T. Smith, 2006). It may mediate the association between traits and health outcomes, an association that is not fully accounted for by health behaviours or socio-economic status (O'Cleirigh et al., 2007; Hampson, Goldberg, et al., 2007). This chapter introduces cortisol, and describes why it is relevant for many different health outcomes, and why it may reflect general health status. A methodological set of criteria are proposed, that constitute a protocol for measuring the diurnal (circadian) profile of cortisol. Results and discussion and considered in the next chapter.

9.2 The function of cortisol

Cortisol is a steroid hormone regulated by the hypothalamic pituitary axis (HPA axis; E. O. Johnson, Kamilaris, Chrousos, & Gold, 1992; Praag, Kloet, & Os, 2004). The HPA axis is a set of glands which interact: the hypothalamus, the pituitary and the adrenals (E. O. Johnson et al., 1992). The function of this axis is to regulate homeostasis, the ability of the body to return cortisol levels to normal, in response to stressors (E. O. Johnson et al., 1992). However, cortisol also affects glucose production, fat metabolism, inflammatory responses, vascular responsiveness and central nervous system and immune functioning (A. A. Stone et al., 2001). Responses to stress can be adaptive, or become maladaptive. When adaptive, this "involves a redirection of both behavior and energy" (E. O. Johnson et al., 1992, p. 116), which is helpful in the short term (e.g. the fight or flight response). Cortisol release raises blood pressure, reduces inflammation and mobilizes energy stores mobilize energy stores (E. O. Johnson et al., 1992). Chronic over-activation of the HPA axis is harmful and is involved in many pathologies. These include hypertension, insulin resistance, abdominal obesity, diabetes, impairment of immune function, infectious illnesses and depression (e.g. Goodyer, Park, Netherton, & Herbert, 2001; Kunz-Ebrecht, Kirschbaum, & Steptoe, 2004; McEwen et al., 1997; Walker et al., 1998). It is for these reasons that chronic stress is well established as a predictor of physical illness. Over-activation of the HPA axis has been linked to many illnesses (see below).

Psychological activation of the HPA axis. There are many psychological factors that activate the HPA axis. These include novelty, unpredictability, uncontrollability, anticipation (Gordon, 1997), ego involvement (Knight, Atkins, & Eagle, 1979) and threats to self esteem (Gruenewald, Kemeny, Aziz, & Fahey, 2004). It is worth noting that HPA is engaged by psychosocial as well as physical threats. The HPA axis may not be able to distinguish modern psychosocial stressors from the threats that existed in humans' evolutionary history (Segerstrom & Miller, 2004). Crucially, there are large individual differences in the magnitude of HPA

axis reactivity. Some of this variation is genetically determined, some environmental, and some by early life experiences (E. O. Johnson et al., 1992, p. 117). A large proportion is not explained, suggesting that psychosocial variables are substantively important.

9.2.1 Salivary measurement

Salivary cortisol is a very well validated index of adrenal cortisol activity, correlating between .71 and .96 with serum (blood) measures of cortisol (Kirschbaum & Hellhammer, 1994). It accurately reflects the unbound, *free* component of cortisol circulating in plasma (Kirschbaum & Hellhammer, 1994). That is, it is a valid index of the physiologically active component of cortisol (S. Edwards, Hucklebridge, Clow, & Evans, 2003, p. 614). Only the free cortisol exerts glucocorticoid effects when it reaches tissues, making it a relevant marker to study (Kirschbaum & Hellhammer, 1994). Saliva samples are stable at room temperature for seven days (De Weerth, Zijl, & Buitelaar, 2003). Taken together, these factors make salivary cortisol measurement attractive for studies of participants in everyday, naturalistic settings (Kunz-Ebrecht et al., 2004).

9.3 Cortisol: The three profiles

At any given time, cortisol level is potentially affected by three processes (waking, time of day, and stress reactivity). There are three cortisol profiles, whose function is to regulate these processes: a waking profile, a circadian profile, and a stress-reactive profile. These vary in their influences (Schreiber, Van Hulle, Clark, Lemery, & Goldsmith, 2003). Therefore, there are at least three separate sources of information provided by salivary cortisol measures. The profiles are superposed on top of one another. Therefore, it is important for researchers to clarify which profile is being measured in a research study.

9.3.1 The awakening profile

The waking profile is called the cortisol awakening response . It is known that salivary cortisol increases two-fold during the first 30 minutes after waking (Pruessner et al., 1997). It then drops, forming a peak. The CAR is a robust pattern, which is reasonably stable across days and weeks. It can therefore be used to characterize individuals — a trait characteristic (Pruessner et al., 1997). The CAR is measured by saliva samples taken at wakening, then 15, 30 and 45 minutes. It is possible to measure the CAR using only two measures. For example, Vedhara, Stra, Miles, Sanderman, and Ranchor (2006) asked participants to provide samples (1) directly after waking (before breakfast); (2) 30 min later. The CAR is calculated using one of several indices:

MnInc. The mean increase from waking to a specified time point (MnInc).

AUC_g. The area under the curve (AUC) with respect to ground.

AUC_i. The area under the curve with respect to increase.

AURC. The area under the response curve, in relation to the first sample after waking (S. Edwards, Clow, Evans, & Hucklebridge, 2001).

The CAR is attenuated or flat in a range of health problems. Kudielka and Kirschbaum (2003) found that the CAR was flattened in cardiovascular, psychiatric, and autoimmune and other illnesses. The CAR is not related to age, weight, smoking status, alcohol consumption the previous night, duration of sleep, physical activity, morning routines and time of awakening (Pruessner et al., 1997).

9.3.2 The diurnal profile

Description

The diurnal profile (hereafter, DP) of cortisol (that is, its daily cycle) involves 10 to 15 secretions (secretory episodes) every 24 hours. However, "both the frequency, duration and magnitude of secretory bursts are increased" in certain illnesses, including depression, resulting in higher cortisol output overall (Praag et al.,

2004, p. 155). The strongest secretory activity of the adrenal cortex is in the early morning (for day-active people). The most amount of cortisol is secreted during the CAR, with a steady decline across the day, as long as there is significant stimulation (A. A. Stone et al., 2001). The lowest levels of cortisol are found at midnight, which gradually increase before waking (Wilhelm, Born, Kudielka, Schlotz, & Wust, 2007). The DP is a response to waking, but this gradual increase during the night, toward morning, is part of the DP (Wilhelm et al., 2007). The CAR is superimposed onto the DP. Cortisol declines across the day in virtually all healthy (non-ill) individuals (A. A. Stone et al., 2001). Exceptions include those subjected to environmental extremes, psychological distress, changes in diet, or shift work (A. A. Stone et al., 2001). Given that the DP is so robust, and that many illnesses (somatic and psychological) disrupt the daily decline of cortisol, it could be argued that the diurnal cortisol profile is a putative biomarker for health status (see below).

Measurement

A short DP can be measured by taking saliva samples three, six, nine and 12 hours after waking (Wust, Federenko, Hellhammer, & Kirschbaum, 2000). The profile can be calculated as (1) the difference between a sample three hours after waking and a sample 12 hours after waking (2) the mean of four samples taken across the day. However, more complex modelling strategies might be required (Hruschka, Kohrt, & Worthman, 2005; Willoughby, Vandergrift, Blair, & Granger, 2007). It is also worth noting that S. Edwards et al. (2003) took four waking samples (1, 2, 3, 4) and four diurnal samples (5, 6, 7, 8). They synchronized the DP *from* and to the waking sample, so that DP referred to samples 1, 5, 6, 7 and 8 (c.f. Wust et al., 2000) not to samples 5 to 8. It is therefore important to clarify exactly what is meant, when using the term "diurnal" in cortisol research. Here, the term is used to refer to cortisol values 3, 6, 9 and 12 hours after waking.

Function

The function of the circadian profiles is to regulate the *predictable* light/dark cycle. The retina of the eye sends information about light/dark to the paired suprachiasmatic nuclei (SCN), a part of the hypothalamus. The SCN is also responsible for circadian regulation of pituitary and adrenal activity (A. A. Stone et al., 2001). It regulates the sleep/wake cycle (Pace-Schott & Hobson, 2002). These functions are consistently present in healthy individuals.

9.3.3 Association between the CAR and the DP

Crucially, the CAR and the DP are unitary phenomena, correlating only around .1 and non-significantly (S. Edwards et al., 2001). S. Edwards et al. (2001) reported a strong correlation between the area under the curve (amount of cortisol produced) for CAR and DP (.60 approximately). However, there was no association between the dynamic of the two profiles (amount of cortisol with reference to the first sample on waking). They suggest that the sample taken 6 hours after waking is most representative of the DP. This sample was moderately correlated with the waking response AUC (.46) but not AURC. This suggests that although the shape of the circadian profiles are not correlated, the cortisol output may be. The notion that each is a unitary phenomena is supported by evidence that genetic factors are important for the CAR but not the DP. Genetic control of cortisol is stronger in the morning than in the afternoon and evening Wust et al. (2000).

9.3.4 The reactivity profile

The function of cortisol in the reactivity profile is to regulate reactivity to *unpredictable* life events or stressors. Activation of the HPA axis by evaluated with salivary cortisol measures. Salivary measures are particularly popular with stress researchers interested in inducing stress experimentally (Oswald et al., 2006). For example, the Trier Social Tress Test (TSST) is a validated way to induce stress. It involves a speaking and mental arithmetic task, in front of an audience, and the test is filmed. The TSST induces "cortisol responses in 70 to 80% of participants"

(Wust, Federenko, Van Rossum, Koper, & Hellhammer, 2005, p. 201-202). Cortisol could be measured before and several times after an experimental stressor (Willoughby et al., 2007). Repeated exposure to the same stressor causes habituation (Kirschbaum et al., 1995; Wust, Federenko, et al., 2005), and high responders do not habituate to repeated stressors (Kirschbaum et al., 1995). Chronic over-activation of the HPA axis can change basal levels of cortisol (Praag et al., 2004). This means that it may be possible to examine individual differences in the CAR and DP within the stress moderation model (T. Smith, 2006). Here, circadian measures reflect the cumulative burden of psychosocial stressors on an individual, termed allostatic load (Korte, Koolhaas, Wingfield, & McEwen, 2005; McEwan, 1998).

9.3.5 Test-retest reliability

Measures of cortisol can be taken over more than day, to improve test-retest reliability. It is worth noting that individuals have trait characteristics in the CAR and DP (A. A. Stone et al., 2001). S. Edwards et al. (2001) tested the DP of 45 normal subjects on two consecutive days. They reported individual consistency over two days in the CAR (.52) and the DP 12 hour mean (.65). Therefore, the CAR and DP have moderate test-retest reliability, more so for the DP. This is primarily evidence for the moderate reliability of salivary CAR and DP measures, although the *stability* of these profiles over weeks, months or years is far less researched. Until recently, very few cortisol studies collected samples over more than one day.

9.3.6 A proposed methodological protocol

Thorn, Hucklebridge, Evans, and Clow (2006) warned that there are several unresolved methodological challenges to measuring cortisol. Many of these have been addressed in recent years, several are ongoing, and some have not been researched at all. Early cortisol studies used overall mean of basal levels of cortisol (Kirschbaum et al., 1990), which are less informative than CAR, DP or reactivity

measures. Early studies also tended to synchronize to clock time, rather than awakening time. Taking all of the above factors into account, and some additional information from the British Psychological Society Psychobiology Section Stress Measurement Workshop (2003), a protocol is proposed that is suitabled for studying personality traits in relation to salivary cortisol profiles:

Focus. Select either the CAR or the DP, or both. They are unitary phenomena (S. Edwards et al., 2001), and should not be conflated.

Weekdays. Distinguish weekdays from weekends. The CAR is attenuated at weekends (Thorn et al., 2006).

Retest. Repeat sampling across two days, to provide an estimate of test-retest reliability, and reduce missing data.

3 hours. If measuring the DP, the first sample should be 3 or more hours after waking.

Syncronization. Synchronize to waking time not clock time.

This protocol was adopted in the study described in the next chapter.

9.4 The diurnal cortisol profile and health

Cortisol is well validated and widely used as a biomarker for HPA axis activity (e.g., Oswald et al., 2006; Schmidt-Reinwald et al., 1999; Wust, Federenko, et al., 2005). It is also validated as an indicator of allostatic load, the cumulative burden of stressors on bodily systems (McEwan, 1998). There is also strong evidence that cortisol is involved in several illnesses. In the examples below, the term "flat" refers to the failure or disruption to cortisol decline throughout the day. "Low" and "high" refer to levels of cortisol. "Slope" refers to the decline of cortisol across the day, where flatter slopes can result from high evening or low morning cortisol[1]. In the DP, it is therefore important to distinguish cortisol output (total cortisol across the day) from cortisol slopes (the rate, or angle,

[1]It is often not possible to discern whether changes in morning or evening levels are responsible.

of decline), particularly when relating cortisol to health criteria[2]. The following is a brief summary of how circadian cortisol profiles have been associated with various health outcomes. It illustrates the wide range of physical and mental illnesses that cortisol output is involved in:

Aging. Age correlates (.33) with cortisol levels (S. Edwards et al., 2001). The elderly show flatter cycles (Wolf, Convit, Thorn, & De Leon, 2002).

Asthma. Cortisol has been found to be low in the morning, high in the evening (Fujitaka et al., 2000).

CFS. Chronic fatigue syndrome (CFS, Catley, Kaell, Kirschbaum, & Stone, 2000; Machale et al., 1998) is associated with low morning cortisol and low evening cortisol (but see Gaab et al., 2002).

Depression. A meta analysis (Burke, Davis, Otte, & Mohr, 2005) confirmed that cortisol is high during the evening in depressed patients. This finding is often replicated (see Praag et al., 2004).

Diabetes. Glucose tolerance, insulin sensitivity, and insulin secretion were all associated with high evening cortisol (Plat et al., 1996).

Fatigue. Higher fatigue in breast cancer patients was associated with flatter cortisol slopes (Bower et al., 2005).

Fibromyalgia. This is a type of chronic fatigue syndrome accompanied by muscular pain and rheumatoid arthritis (McCain & Tilbe, 1989). Cortisol values were very high in the morning and high the evening in this study.

Discord. Marital discord was associated with low morning and high evening cortisol (S. E. Sephton, Sapolsky, Kraemer, & Spiegel, 2000).

Memory. Worse memory functioning is associated with flatter cortisol slopes (Abercrombie et al., 2004).

Mortality. In breast cancer patients, the slope of the diurnal decline was found to be a powerful predictor of survival, stronger than immune system markers such as natural killer (NK) cells (S. E. Sephton et al., 2000). In stroke patients, Marklund, Peltonen, Nilsson, and Olsson (2004) found that low cortisol output was an independent predictor of 28-day mortality. Low and

[2]See `http://tinyurl.com/39atln`

high output were associated with increased 1-year mortality. High and low circulating cortisol levels were associated with increased mortality after stroke.

Social support. Lower perceived social support was associated with flatter cycles (Abercrombie et al., 2004).

Panic disorder. Panic disorder (Bandelow et al., 2000) was associated with erratic and high cortisol values.

PTSD. Post Traumatic Stress Disorder (PTSD, Yehuda, Teicher, Trestman, Levengood, & Siever, 1996; Strickland, Morriss, Wearden, & Deakin, 1998) has been linked to low morning and low evening cortisol.

Sleep. Sleep disturbance was associated with low morning and high evening cortisol in one study (S. E. Sephton et al., 2000). Sleep loss often results in an elevation of cortisol levels the next evening (Leproult, Copinschi, Buxton, & Van Cauter, 1997).

Perceived stress. Perceived stress is associated with higher cortisol across the day, a well established finding (Abercrombie et al., 2004; McEwan, 1998).

Work stress. Among stressed workers, Dahlgren, Kecklund, and Akerstedt (2005) found a flatter slope, due to higher evening cortisol, in more stressful working weeks.

Exhaustion. Vital exhaustion is linked to flatter slopes (Nicolson & Van Diest, 2000).

This is not an exhaustive review of the cortisol-health literature, but it clearly illustrates that cortisol profiles are disrupted in a wide range of illnesses. The studies involve different cortisol profiles, different methodology, and different patient groups. However, reflecting on the ongoing work in this area, I noticed that several researchers have proposed a specific hypothesis about the cortisol-health relationship. Could cortisol profiles function as a biomarker for *general health status*, as well as stress and HPA activity? I found four locations in the literature where this hypothesis was eluded to; if not always explicitly:

Expert meeting. At a meeting of experts, it was proposed that either cortisol rhythm dysfunction or total cortisol output, might relate to health. The report said

"This is an area of significant controversy"[3]. They concluded that the notion is a provisional hypothesis, and requires further research (e.g., Kudielka & Kirschbaum, 2003; Kudielka, Buske-Kirschbaum, Hellhammer, & Kirschbaum, 2004). Currently, cortisol is accepted as a biomarker for stress (Clow et al., 2004).

CAR. Kudielka and Kirschbaum (2003) discussed the CAR as a biomarker for health status. They conclude that the evidence is equivocal: "While it is tempting to speculate that an increased awakening cortisol level might be a biological marker for general health problems, the present data are derived from small samples and should thus be viewed as very preliminary" (p.43).

Regulation. It has been proposed that the DP might reflect *general regulatory competence* of several bodily/psychosomatic systems. This seems likely, given that cortisol is involved in glucose production, fat metabolism, inflammatory responses, vascular responsiveness, central nervous system and immune functioning (Spiegel & Sephton, 2001; S. Sephton & Spiegel, 2003; A. A. Stone et al., 2001).

Fatigue. The diurnal cortisol profile correlates with illnesses that are "characterized by fatigue" (Bower et al., 2005, p. 97). Fatigue represents a combination of mental and physical limitations in "what people are able to do and how they feel", and is an indicator of (low) functional health status (Ware et al., 1993, p. 9:1).

I return to this issue below, and in the next chapter, where I test for an association between cortisol and the SF-36 scales.

9.4.1 Mechanisms linking cortisol with health

What are the underlying mechanisms linking many different illnesses to the DP? Cortisol is known to be "cortisol secretion as a response to perceived stress is a powerful factor regulating disease-generating events in the periphery." (Rosmond, Dallman, & Bjorntorp, 1998, p. 1859). That is, it is associated with diseases even

[3] `http://tinyurl.com/39atln`

outside the endocrine systems (Rosmond et al., 1998). However, the mechanisms underlying individual differences in diurnal variation are not fully understood (A. A. Stone et al., 2001). Observed cross-sectionally, a flat DP could be a cause or a consequence of illness. If the associations continue to appear in cross-sectional studies, then the hope is that large scale longitudinal studies will begin to incorporate measures of cortisol. Building on arguments from Spiegel and Sephton (2001); S. Sephton and Spiegel (2003); A. A. Stone et al. (2001), that diurnal variation reflects a general loss of regulatory competence in other systems, one hypothesis is a "common cause". That is, some underlying mechanism is responsible for both cortisol and the health problem, regulating the quality or integrity of bodily systems:

Genetic. Shared genes might control both the circadian clock and these illnesses characterized by fatigue.

Allostatic load. Allostatic load (Korte et al., 2005; McEwan, 1998). Flattened cortisol rhythms reflect the "burden" of stressors on the HPA axis (Van Itallie, 2002) and on anthropometric, endocrine, and hemodynamic factors (Rosmond et al., 1998) such as low testosterone secretion, insulin resistance. Over-activation of other systems may also be involved.

Activity. Chronic inflammatory processes and associated behaviours (e.g. fatigue) might disrupt diurnal variation, via rest and activity patterns (Hatfield, Herbert, Van Someren, Hodges, & Hastings, 2004).

Immunity. Adrenal cortex derived steroids, such as cortisol, have potent effects on various aspects of the immune system (Spiegel & Sephton, 2001).

Stress. Perceived stress might influence illnesses, and diurnal cortisol profiles, simultaneously.

Hippocampus. Depression might cause hippocampal atrophy, and flattening of the DP (Praag et al., 2004), at the same time.

Neuroticism. Neuroticism plays a critical role in stress appraisal (Bolger & Schilling, 1991; Kendler et al., 2004; T. Smith, 2006; Van Os & Jones, 1999). Therefore, cortisol could act as a mediator between personality traits and illness. It is a biologically plausible variable that could explain why Neuroticism

is sometimes associated with worse health outcomes. This is discussed in more detail below.

9.4.2 Rationale for selecting the DP in the cortisol study

My decision to focus on the DP, rather than the CAR, was one of the most important decisions in the research. There were four principal reasons for the decision to focus the DP:

Reliability. The correlation across two days is slightly higher for the DP than for the CAR. For example, correlations across both days were .34 for mean increase (MnInc), .50 for AUC, .45 for the sample provided three hours after waking, and .65 for the 12 hour sample (S. Edwards et al., 2001). The DP is therefore more reliable than the CAR.

Health criteria. The correlates of the CAR tend to be short term, severe stressors, such as post-traumatic stress disorder (Yehuda et al., 1996; Strickland et al., 1998). In contrast, the criteria associated with the DP are longer term and more wide-ranging. If Neuroticism is involved, the association is likely to occur over longer time periods.

Depression. The DP is strongly related to negative affectivity (including depression). Neuroticism might underlie these associations.

Compliance. Deviations from sampling times are less important for the DP. This is because samples for the CAR are timed to minutes, whereas samples for the DP are timed in hours.

9.5 Neuroticism and cortisol

Neuroticism is the most researched trait in relation to illness, and has emerged as a predictor or mortality and morbidity. However, the findings are not consistent (see chapter two). The different populations studied may partly explain the different findings (Shipley et al., 2007). However, there are methodological

issues to consider. Many studies report correlations between subjective measures of health and Neuroticism, but Neuroticism is known to increase the likelihood of symptom reporting (Costa & McCrae, 1987). Stronger studies link Neuroticism with many objective illnesses, including asthma (Huovinen, Kaprio, & Koskenvuo, 2001), gastrointestinal disorders (Drossman et al., 2000), immune system functioning (Daruna, 1996; O'Cleirigh et al., 2007) and prognosis and CHD death risk after first heart attack (Denollet, Sys, & Brutsaert, 1995; Murberg, Bru, & Aarsland, 2001). However, the literature on possible candidate mechanisms, linking Neuroticism with illness and mortality, is relatively sparse. Cortisol is one such candidate. There are three reasons why it is important to clarify whether Neuroticism is associated with the DP. As shown below, the literature has produced mixed findings, which suggests the need for clarification and further study. Second, there are substantive theoretical reasons to suspect that Neuroticism, rather than other big five traits, is associated with cortisol. Prior research has shown that cortisol is often associated with general health status (E. O. Johnson et al., 1992; McEwan, 1998) and specific pathologies (Goodyer et al., 2001; Kunz-Ebrecht et al., 2004; McEwen et al., 1997; Walker et al., 1998). It is also associated with sensitivity to stress(Bolger & Schilling, 1991; Kendler et al., 2004; Tyrka et al., 2006; Van Os & Jones, 1999) but see (Ormel, Rosmalen, & Farmer, 2004). This is because high Neuroticism scorers consistently appraise life events as more stressful (Costa & McCrae, 1987; Matthews, Deary, & Whiteman, 2003; Watson & Pennebaker, 1989). In the long term, over-activation of the HPA axis could raise cortisol output, preventing it from returning to baseline levels.

9.5.1 How the association might work

If an association between cortisol and Neuroticism does exist, then it will become more important to develop explanatory accounts of how this association might work. A systematic review may be appropriate, even if a suitable theory is lacking. Indeed, "the importance or unimportance of theory is unlikely to emerge unless review activity is structured to cross problem/outcome areas" Oakley (1999), cited in Higgins and Green (2006). However, at this early stage

in the literature it is important to state the ways in which an association *might* work. There are five possible explanations (adapted from Singh-Manoux, 2005):

Mediator. A third variable (e.g. depression, perceived stress) could mediate the association between Neuroticism and cortisol. Neuroticism may increase perceived stress, which may raise cortisol.

Moderator. The size of the association may vary, depending on the level of a covariate (e.g. depression, gender).

Common cause. A common cause might underlie Neuroticism and cortisol, such as shared genes (Munafo et al., 2003; Phillips, Douglas, Burns, & Mark, 2005; Portella, Harmer, Flint, Cowen, & Goodwin, 2005). An association between related several constructs and variables, that is under genetic control, is referred to as an endophenotype (Gottesman & Gould, 2003). For example, a genetic predisposition to experience high levels of psychological distress, Neuroticism, and cortisol production: collectively, these could form an endophenotype.

Reverse causation. Cortisol is upstream of Neuroticism, and is therefore a cause of personality change. Cortisol is known to elevate clinical levels of depression. It has even been suggested that antidepressants work by normalizing HPA axis activity (for a discussion, see Praag et al., 2004, p. 148). It is also plausible that cortisol might elevate Neuroticism, although this hypothesis has not been tested.

It is beyond the scope of this thesis to evaluate these nascent explanatory accounts. They are all tenable as multiple working hypotheses (Chamberlin, 1965). First, it is important to draw on the existing literature, address methodological shortcomings and sampling inconsistencies, and attempt to provide a descriptive account.

9.5.2 Previous research

Portella, Harmer, Flint, Cowen, and Goodwin (2005) contacted a small sub-sample of volunteers with very high (N = 15, 7 male) and very low (N = 15, 7 male) N scores on the Eysenck Personality Inventory (H. Eysenck & Eysenck, 1964) as part of a larger study. The participants were all adults, and were free from Axis I disorders (e.g. depression, anxiety) and for females, premenstrual week was avoided. They provided five saliva samples every 15 minutes starting at waking time; then at 12:00, 18:00 and 22:00. Coefficients of intra- and inter-assay variation were not reported. Differences between high and low Neuroticism groups were not observed for the diurnal profile. For the waking response, higher cortisol (assessed by a group $*$ time interaction in ANOVA) was observed for high Neuroticism scorers (ps = .03, .01 and .01 for 15, 30 and 45 minutes after waking, respectively). This also applied to area under the curve, which was significantly higher for high N scorers (p =.02), and among them, the correlation was .52. There was not a significant correlation between Neuroticism and Beck Depression Inventory scores in this group. The correlation between Neuroticism and Extraversion was not reported.

LeBlanc and Ducharme (2005) studied a small sample of 20 (11 of whom were male) white volunteers "in good physical and mental health" (p. 677). The age was fairly young (25 to 35 years of age). Smoking and alcohol were not permitted on the day of the study. Cortisol was measured from a blood sample at 13:00 in a laboratory setting. A light meal was provided at 11:30. Intra- and inter-assay variation coefficients were not reported. Personality was measured using the Big-Five Inventory (BFI). There was a large negative correlation (.52, p = .03) between Neuroticism and cortisol. Higher Neuroticism scores were associated with lower cortisol values. They suggest that this may be because the samples were provided at different times of the day, and urge researchers to investigate time of day effects in future studies. The strengths of this study include its control of the influence of four key confounding health behaviours (nutrition, activity, alcohol and smoking) on the day of the study. The study provides an insight

into afternoon levels of cortisol, and the possibility that high cortisol from the CAR is included in the sample can be discounted. However, only one sample was provided. Although this sample can be described as part of the diurnal profile, it was not possible to calculate an area under the curve, or consider the contribution of late afternoon and evening samples. A weekday sample was assumed, given that laboratories tend not to open at weekends, although this was not clarified in the paper.

In Vedhara, Stra, Miles, Sanderman, and Ranchor (2006), 85 female breast cancer patients and 59 controls provided saliva samples at waking, 30 minutes later, 11:00-13:00 before lunch, and then 20:00-22:00, at least 2 hours after evening meal. This was repeated across two days. A measure of both the CAR and the DP was therefore available. Participants were asked not to eat or drink 30 minutes prior to all of the samples. Intra-assay variation ranged from 4.1 to 8.2%. Inter-assay variation ranged from 5.6 to 12.6%. Personality traits were measured using the Eysenck Personality Questionnaire-Revised (EPQR-S, H. J. Eysenck & Eysenck, 1985) translated into Dutch. There are 12 items for each trait. The correlation between Neuroticism and Extraversion was not reported. Only one of the author's multiple regression models resulted in a significant association between a psychosocial variable. Neuroticism was negatively associated with cortisol in "the model examining the role of trait measures in predicting early morning peak in patients" (p. 309). The parameter estimate was β = -.33. The early morning peak measures HPA axis reactivity to waking, and was calculated simply as the second sample subtracted from the first. This suggests that among patients, but not controls, higher Neuroticism is associated with lower cortisol peak in the CAR. The strengths of the study are that a fairly large sample size was used. Several samples were provided for both the CAR and the DP, and these were repeated across two days. Participants were asked to refrain from eating and drinking before samples. This controls for the influence of food and drink on the samples. There are some weaknesses to the study. The authors of the study do not report whether participants were asked not to smoke or drink alcohol. However, they

were asked not to drink any fluid 30 minutes before each sample. It is not clear from the paper whether weekend samples were not permitted.

In contrast to Portella et al. (2005), who hypothesized that Neuroticism and depression were risk factors for cortisol production, Zobel et al. (2004) conceptualize Neuroticism and HPA axis dysregulation as risk factors for affective disorders such as depression. Neuroticism and cortisol are upstream and depression is downstream. There is theoretical support for this notion, given that "neuroticism and HPA dysregulation are both thought to reflect the tendency to cope less efficiently with stress and other challenges" (Portella et al., 2005, p. 393). This illustrates heterogeneity in the theoretical basis for models of Neuroticism and cortisol in the literature. In both studies, Neuroticism is an independent variable. However, cortisol and depression are variously positioned as dependent variables. This shows that researchers can model cortisol, or depression, as the outcome measure of interest. The participants in the study were physically healthy, and consisted of 94 people (39 males). Some of these participants had mental symptoms however, and "reported previous or current symptoms asked in probing questions: depressed mood for several days (N = 8), intensive feelings of anxiety without sufficient reason (N = 8), which, however, did not qualify for a diagnosis of mental illness" (p. 394). This is a weakness of the study, because Neuroticism might be confounded with depressed mood, negative affects, and subclinical depression. The dexamethasone (Dex) suppression test, in which HPA activity is monitored after administration of dexamethasone, was administered at 23:00. Samples were provided the following day 15:00, 15:30, 15:45, 16:00 and 16:15. A significant positive correlation was reported between Neuroticism and maximum cortisol output (.36, ps $< .001$). Therefore, higher Neuroticism was associated with higher cortisol output during the afternoon to later afternoon. Although the Dex test measures reactivity rather than circadian rhythmicity, it is interesting to note that the afternoon to early evening profile did produce a significant association with Neuroticism.

(Walsh, Wilding, Eysenck, & Valentine, 1997) recruited a sample of lecturers (N = 32, 24 males), who had a lecturing and a non-lecturing week. Traits were measured using the Eysenck Personality Questionnaire (H. J. Eysenck & Eysenck, 1985). Salivary cortisol samples were provided on Monday, Wednesday and Friday, between 10:00 and 12:00. After excluding participants with more than two missing data points, then taking the average of available days for remaining missing points, 25 participants remained. The fact that missing data patterns were reported can be considered a strength of this study. No significant association was observed between "linear trend" of cortisol and Neuroticism. This study has the strength that the samples were repeated over three days. It was possible to calculate a linear trend across the week. However, a key weakness is that only one sample was provided. Although this was after the CAR, excluding the possibility of residual waking cortisol, it is not possible to calculate area under the curve or amount of diurnal decline. If cortisol levels were similar on each day, the restricted variance could prevent a significant association with Neuroticism from being detected. The authors did not test for an association between Neuroticism and cortisol across the day, or for mean cortisol levels. A second weakness of the study was that methodological controls and confounding factors were not reported (e.g. eating and drinking before samples). Finally, the sample size was small.

In Westrin, Engstom, Ekman, and Traskman-Bendz (1998), suicidal patients were compared with 38 health controls. In contrast to Zobel et al. (2004), the patients were screened for mental illness. They were also free of medication, including oral contraceptives. The authors administered the Dex test, which is a measure of the reactivity profile. However, because the samples were taken across the day, the profile can loosely be described as "diurnal". The first sample was at 15:00 on the first day, before the dexamethasone was provided at 22:00. On the second day, cortisol was measured at 08:00 and 15:00. One weakness here is that the 08:00 sample could be confounded with the CAR. The authors report an inter assay coefficient below 5% and 7% for intra- and inter-assay coefficients of

variation, respectively. The results showed a positive correlation between Neuroticism and cortisol, but only in patients (.17), and this was not statistically significant. There was no association in controls. This same pattern was observed in the study of breast cancer patients, in which the association was limited to the breast cancer group, not to controls (Vedhara et al., 2006).

Hanson, Isacsson, Janzon, and Lindell (1990) studied 104 employees in two companies. Trait negative affect was measured using the Well-Being Questionnaire. The protocol was different from most other studies, because participants were asked to provide samples throughout the day, using a method called Ecological Momentary Assessments (Csikszentmihalyi, 2002). A laptop computer alerted participants at "semirandom intervals" to provide six samples throughout the day, over two days. This not only provides the diurnal profile, but unlike many studies in the literature, provides a two day retest. However, one of the two days was a workday, and the other a rest day. Given given evidence that cortisol profiles differ at weekends (Thorn et al., 2006), this may represent a weakness of the study. Ideally, two of the same type of day should be recorded. The results showed that only mood (state negative affect) was associated with cortisol. Trait negative affect did not, although this was after controlling for time of day. The authors did not test whether the association was stronger at certain times of the day, or if it was associated with diurnal decline — the slope as cortisol decreases across the day. The authors also controlled for food consumption and for smoking. In the discussion section, the authors make an interesting link to the literature on Neuroticism and health. They propose that cortisol could be a mechanism linking negative affect (state or trait) with ill health. If state negative affect can raise cortisol throughout the day, then sustained negative affects over time might result in chronic over-activation of the HPA axis. Neuroticism becomes, in effect, a chronic stressor.

Allen, Batty, and Dodd (1985) measured serum cortisol, rather than salivary cortisol, measured in medical students in the week preceding an important annual

examination. There was more than one sampling day, and the samples were collected at 10:00 and 16:00 for one study (N = 32); and at 09:00, and 16:00 for the second study, one year later (N = 20). In study 1, samples were provided in the first week of term but the examination was not until the end of term. In study 2, samples were provided leading right up to the examination itself. This is the largest test-retest interval I have encountered in the literature (one year). Neuroticism was measured using the EPI (H. Eysenck & Eysenck, 1964), but was not associated with cortisol in either study. As expected, there was an increase in cortisol for each measure, leading up to the examination itself. The study has a number of strengths. By focusing on a group of males, this controls for oral contraceptive use. The fact that all the students are preparing for the same stressor, likely controls for the influence of minor life events, hassles and uplifts. Individual differences in cortisol reactivity to the *same* life stressor might be better observed than individual reactions to individual life events. The assumption in many naturalistic studies, is that the life events can be ignored since they average out across the sampling days. Researchers often choose experimental reactivity studies in order to study this part of the profile. However, they cannot access the waking and diurnal profiles easily in this way.

Schommer, Kudielka, Hellhammer, and Kirschbaum (1999) studied both the reactivity and the diurnal profiles, collecting samples every 30 minutes from 09:00 to 21:00 for 81 volunteers. The number of samples is a clear strength of this study, but the plot they present illustrates a potential weakness for those other studies that begin sampling at 09:00. There is a clear peak, particularly for participants high in Extraversion, in cortisol during the first four. This might represent residual cortisol from the waking response. Ideally, the first diurnal sample should be three hours after waking (Wust et al., 2000). No association was found between Neuroticism and mean cortisol levels. However, the authors dichotomized personality traits into upper and lower quartiles, which can result in a loss of information (Streiner, 2002). Neither did they consider the slope of the cortisol decline. It is possible that the rate of decline, not the mean levels across the day, are associated with personality traits. Although the report was a short one, the

authors are rather dismissive of the potential for a Type II error (a false negative). It is possible that an association does exist, but that their study did not detect one. They do not evaluate the weaknesses of their study, or suggest ways in which the methodology could be improved. Therefore, their rejection of the hypothesis that "personality traits like Extraversion or Neuroticism are closely related to basal or stimulated free cortisol concentrations" (p. 841) is premature.

A study by Londen et al. (1997) contained a sample of depressed patients, and a control sample (N = 37; 17 male), similar to other studies (Westrin et al., 1998; Vedhara et al., 2006). The controls were recruited from newspaper advertisements, and were excluded if they reported mental or physical illness, recent stressful life events, or shift work. The mean age was 41.2, showing that the sample was comprised of adults rather than students. Cortisol was measured in serum, with samples collected at 08:00, 16:00 and 23:00. This potentially confounds the waking with the diurnal profile. Furthermore, it cannot be considered naturalistic because blood samples were taken in the laboratory. Cortisol was not associated with Neuroticism, as measured by the EPQ (H. Eysenck & Eysenck, 1964), "after controlling for the number of comparisons ($p < .006$)" (Londen et al., 1997, p. 289). The authors do not report what the associations were, before controlling for multiple comparisons.

If an association between Neuroticism and the DP exists, it may exist in primates. In rhesus macaques, the construct of Excitability is equivalent to the human trait of Neuroticism. In a study of 16 animals, lower cortisol reactivity was associated with Excitability during a series of afternoon blood samples (Capitanio, Mendoza, & Bentson, 2004). The authors suggest that Neurotic animals are less likely to show a strong reactivity profile, because homeostasis has been disrupted. Although this was a test of the reactivity profile, not the DP, it is noteworthy because the findings were not significant for the morning samples. It is possible that reactivity interacts with the diurnal components of cortisol production, a topic worthy of further study. Poor reactivity could result in lower basal levels

of cortisol, over time[4]. The authors concluded that diurnal rhythmicity should become a research priority for trait researchers who are interested in cortisol. Reactivity might have stronger relationships with traits at certain points in the diurnal profile. The authors note the lack of research in humans, but the potential of the diurnal profile to explain the health consequences of HPA axis dysregulation. The starting question should be, what proportion of variance in cortisol output can be explained by personality traits?

9.6 Research priorities

The evidence supporting an association between cortisol and the DP is equivocal (Vedhara et al., 2006). It is clear from the accounts above that the nomological network supporting an association between Neuroticism and the diurnal profile is not well established. A study which addressed the methodological shortcomings of previous research would be a useful addition to the literature. It could clarify the size and specificity of the association, if it exists at all. Ultimately, if the association were replicated, it could be used to improve models of personality and health, and the literature on the biological basis of personality traits. There are five priorities for a new cortisol study:

Methodology. None of studies included in this review meet the gold standard criteria for salivary cortisol sampling of the diurnal profile, outlined in chapter nine. This should be addressed in future research. There is considerable heterogeneity in methodology.

Study quality. Only one study reported a positive association between cortisol (Zobel et al., 2004) and Neuroticism for the diurnal profile. One study reported a negative association, which was unexpected (LeBlanc & Ducharme, 2005). Eight studies reported no association. However, many of the other studies had low methodological quality, making it impossible to draw firm conclusions. The fact that an association appears several times in patient groups,

[4] Arguably, it could also lead to higher basal levels, if the HPA axis is unable to regulate homeostasis.

but not healthy controls, suggests that physical illness may moderate the association, or act as a confounding factor. It is therefore important to carefully describe the nature of a sample carefully, and distinguish healthy volunteers from those with physical illnesses.

Test-retest. Very few studies repeated sampling on a second day. It is important to improve the reliability of circadian measures in future studies with a test-retest design, in future research.

Assay reliability. Intra- and inter-assay variation was not always reported. These should be reported in any future research.

Diurnal focus. Modelling strategies are highly inconsistent across studies. The diurnal nature of cortisol presents modelling challenges. Ideally, a method should be chosen that can incorporate time, test-retest measures, covariates, diurnal variation, and individual differences. Diurnal rhythmicity is an urgent priority for trait research (Capitanio et al., 2004).

Modelling. None of the studies have, so far, utilized the generalized latent variable modelling framework (Muthén, 2002), to model the trajectory of cortisol across the day (e.g., Willoughby et al., 2007). This will be addressed in my own research, presented in the next chapter.

9.6.1 The need for a new study

Taken together, these priorities, and the methodological criteria outlined in the previous chapter, provide a clear starting point for investigation the association between Neuroticism and the diurnal cortisol profile. There is a clear need to design a study that can integrate the strengths and weaknesses of previous research, clarify whether there is an association between Neuroticism and the DP, and specify whether this occurs in a sample of healthy volunteers.

CHAPTER 10

Growth modelling: Cortisol and Neuroticism

> *How we spend our days is, of course, how we spend our lives* (Dillard, 1990, p. 32-33).

10.1 Introduction

This chapter presents results from the cortisol study, designed to address the methodological shortcomings of previous research. The central research question concerns whether Neuroticism is cross-sectionally associated with the diurnal cortisol profile (DP). However, in line with the overall theme of this thesis, the chapter also concerns a modelling strategy. Latent growth curve modelling, part of the generalized latent variable modelling framework (Muthén, 2002), can model the DP. Towards the end of this chapter, I also test a secondary hypothesis that functional health status and DP are related. Evidence that cortisol is a biomarker for functional health status, in addition to a marker for stress, would be provided if it were shown that the SF-36 correlated with DP.

10.1.1 Modelling cortisol data: Strategies adopted by previous researchers

There are at least five ways to model cortisol data. The first three are typical of the literature, and use forms of general linear modelling such as correlation and regression. Multilevel modelling and latent growth curve modelling are more

advanced techniques, and have not been widely adopted in the cortisol literature.

SUM and DIF. These refer to the sum of all four cortisol measurements (SUM) and the difference between sample 1 and sample 4 (DIF), following Wust et al. (2000) and S. Edwards et al. (2001). These approaches usually take the mean from each sampling day.

AUC. Area under the curve can be calculated for the 12-hour day (DAUC) and the regression (slope) coefficients (BSLOPE) from 3 to 12 hours after waking for each subject on each day (e.g., S. Edwards et al., 2003). Area under the curve can include or exclude the baseline, which is referred to as the intercept (Vedhara et al., 2003).

Time. Time can be coded as a continuous variable (Vedhara et al., 2003). For example, the sample times are treated as variables, rather than calculating the amount of cortisol output across the day.

Multilevel. Multilevel modelling treats time as a continuous variable, but models explicitly between-individual (individual differences), within-individual and measurement (method) error. Hruschka, Kohrt, and Worthman (2005) describe how multilevel models can estimate between-person differences in the means and diurnal slopes of momentary cortisol assessments in naturalistic settings. Measurement error is generally accepted as contributing little (5%) to cortisol models (Kirschbaum et al., 1990). However, between-individual differences, across days, are rarely modeled. This is an important flaw because it limits the size of the correlations that are possible to obtain between cortisol values and psychosocial variables (Hruschka et al., 2005). When between-individual differences share 30% of the variance in cortisol values, then the correlation with any between-individual trait cannot exceed .30 (Hruschka et al., 2005).

LCM. Latent growth curve modelling (hereafter, LCM) is part of the generalized latent variable model framework (Muthén, 2002). This method is very similar to multilevel modelling, but models the data as a latent variable model

(Willoughby et al., 2007). However, it is particularly well suited to the cortisol data (Willoughby et al., 2007), and since it de-attenuates error from repeated measurements (K. A. Bollen & Curran, 2005), it is likely to maximize the size of any association found with personality traits. This was one of the key research issues identified in my introduction (see Pruessner et al., 1997). The modelling technique has existed for some time, but has only recently been implemented in models of personality-health (e.g., Hampson, Andrews, & Barckley, 2007; Hampson, Andrews, Barckley, & Severson, 2006).

Latent growth curve modelling was adopted as the method for the present research.

10.2 Latent growth curve modelling

It is clear from Figure 10.1 that there are individual differences in the cortisol intercepts, and to a lesser degree, in the slopes. Some people start with higher cortisol levels than others, and some people have a sharper rate of decline across the day than others. A model of this data could estimate four parameters: mean of the intercepts, mean of the slopes, variance of the intercepts and variance of the slopes. It is the variances that capture individual differences. This is achieved in LCM by creating two latent variables, or factors: latent intercept and latent slope. These are given a mean structure, so that there is a mean (called the "fixed" part) and a variance (called the "random" part) for each. The latent variables capture the two major sources of variation in the data: intercepts and slopes.

Illustration. To illustrate, Figure 10.1 shows that the lines of best fit connecting the cortisol samples together are not perfect. There are residuals — deviations from the line of best fit (error). When a trajectory model has imperfect fit, this allows us to test a more parsimonious model. The researcher can "smooth over"

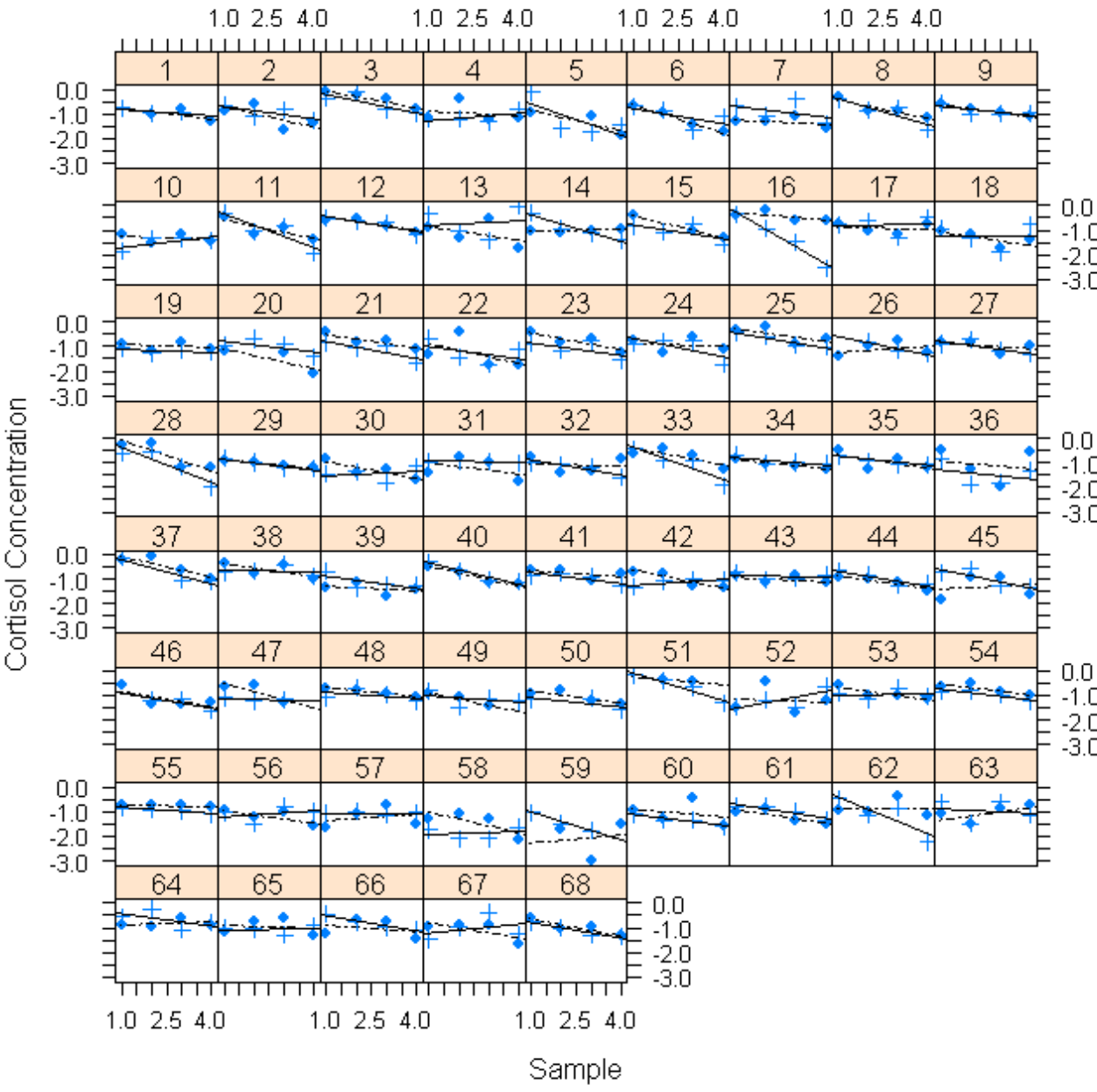


Figure 10.1: Line graphs illustrating a line of best fit for each of the 68 participants' cortisol values, showing diurnal decline across both days.

the data and hypothesize that a single latent variable (such as a straight line) explains most of the variation in each of the 68 slopes (K. A. Bollen & Curran, 2005). A single underlying process might account for the straight line. In a simple LCM, an ordinary least squares (OLS) regression line can be fitted to changes in variables over time, where the intercept represents the model implied value at the initial time period (K. A. Bollen & Curran, 2005). For example, cortisol measured at 3, 6, 9 and 12 hours after waking could be modelled by a straight line, smoothing over the decline in cortisol that usually occurs throughout the day. The time point at three hours after waking is the intercept. At least three time-points are required in order to identify a LCM model (K. A. Bollen & Curran, 2005). The decline in cortisol is linear for the most part, best described by a single straight line (with some error). It would be far less parsimonious to draw four separate lines, joining the time-points together[1].

10.2.1 Advantages of LCM

LCM can model patterns of change, or *trajectories*, over three or more time-points. It represents a great improvement over two time-point data. Two time-points are informative for particular situations (for example, to provide test-retest reliability statistics, or when only a single followup is required). However, when the aim is to model a pattern of change, and only two time-points are available, the researcher can only connect the values of a variable together. This produces a perfect fit to the data (DATA = MODEL) which means researchers cannot create a more parsimonious model than the data (K. A. Bollen & Curran, 2005). Nor can the goodness of fit be tested. When connecting three or more values of a variable together, the trajectory (line of best fit) is necessarily imperfect. This produces error, and the goodness of fit can therefore be tested (Miles & Shevlin, 2000).

History of LCM. LCM is a fairly recent development, building on research techniques used earlier in history. K. A. Bollen and Curran (2005) explain that interest

[1]Other types of lines might be appropriate for other situations. For example, a quadratic curve might best represent other patterns of change. This requires at least four time points (K. A. Bollen & Curran, 2005).

in LCM has arisen out of growing recognition of the limitations of cross-sectional data, the growing availability of longitudinal data, and growing dissatisfaction with traditional approaches to modelling of change/stability. They argue that interest in the process of change can be seen as far back as Aristotle's thinking. They trace the development of LCM through different techniques designed to model change: polynomial mortality tables (19th Century), nonlinear polynomials and logistic curves (20th Century), individual and group analysis of variance (ANOVA) (1938-1950), factor analysis (1950-84) and structural equation modelling. From 1984, advanced techniques were developed for LCM which allowed researchers to use multiple indicators, multiple groups, mixture modelling and more than two explanatory variables (K. A. Bollen & Curran, 2005). These developments are relatively recent, and few have been approached by personality researchers.

10.2.2 Equivalence to multi-level modelling.

LCM is mathematically equivalent to multi-level modelling (hereafter, MLM), at least "under a broad set of conditions" (Curran, 2003, p. 529). A standard MLM (Gelman & Hill, 2006) is a very restrictive CFA (Muthén, 2002). For an example of cortisol data modelled using MLM, see (Hruschka et al., 2005). In LCM, random effects are simply the repeated measures. The time variable is implied by the factor loadings 3, 2, 1, 0 (a slope increasing upwards and to the right would be coded 0, 1 , 2, 3). These imply a linear, declining slope (see above). LCM is often preferred over MLM because there is no need for the second "level" (K. A. Bollen & Curran, 2005). The second level is the random part, allowing each participant to have their own intercept and slope. In LCM, this is implied by the variance, the individual differences, or random part. The pathways leading from personality to intercepts and slopes would correspond to the second level in the MLM framework.

Reason for adopting the LCM approach

I adopted LCM for three reasons:

Framework. LCM is part of the generalized latent variable modelling framework (Muthén, 2002). The central claim of this thesis is that associations between traits and health are best communicated within this framework.

Model fit. Model fit is more readily available (at the time of writing) from LCM models than MLM models

Processes. Parallel growth processes, where there is more than one trajectory, can be modelled in LCM. These are also called multivariate latent curve models (K. A. Bollen & Curran, 2005), and model "simultaneous changes in two or more sets of repeated measures data" (Willoughby et al., 2007, p. 141). For example, it is possible to model the growth pattern for cortisol on more than one day. This is more difficult to model in MLM frameworks (K. A. Bollen & Curran, 2005).

10.3 Methods

10.3.1 Protocol and questionnaire

The protocol for the cortisol study was approved by the NHS Lothian Regional Ethics Committee. The inclusion criteria were that volunteers were healthy, non-smoking adults, not taking any prescription medications, normal day-active individuals not involved in shift work.

Consent form. The consent form, Participant Information Sheet, and questionnaire, are included in the Appendix.

Personality trait measures. Participants completed the NEO-PI-R U.K. edition (Rust & Lord, 2006), which has comparable or better internal consistencies with the NEO-PI-R US edition (Costa & McCrae, 1992a).

Functional health status measures. The SF-36 version 1 (Ware et al., 1993) was described in chapter four. It consists of 36 items, producing eight scales which measure different aspects of physical and mental functional health status. PCS and MCS scores were derived from these scales using the U.K. factor score coefficients (Jenkinson et al., 1999). Histograms showing the frequency distributions for these scores are shown in Figures 11.2 and 11.3.

Perceived stress measures. The Perceived Stress Scale (PSS) was included as a measure of global perceived stress (S. Wright, Johnston, & Weinman, 1995). Examples items include "In the last month, how often have you been upset because of something that happened unexpectedly?" and "In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?". Reliability and validity information are described in (S. Cohen, Kamarck, & Mermelstein, 1983). Responses are recorded on a Likert scale ranging from 0 to 4 (never, almost never, sometimes, fairly often, very often)

Life event measures. The Hassles and Uplifts Questionnaire is a 117 (hassles) and 136 item (uplifts) scale (S. Wright et al., 1995). Participants are first asked to circle the hassles that have happened to them in the past month, then indicated the severity of the hassle by circling one of three response options, ranging from 1 to 3 (somewhat severe, moderately sever, extremely severe). Examples items include "Displacing or losing things" and "Pollution". For uplifts, participants circle events "that makes you feel good" indicating that they occurred in the past month, then a number ranging from 1 to 3 to indicate the frequency of the uplift (somewhat often, moderately often, extremely often). Examples items include "Getting enough sleep" and "Flirting". This measure was excluded from further analysis because participants used the response scales inconsistently, and appeared not to understand the instructions. For example, some rated all events, perhaps assuming that all items should be completed. This caused ambiguity in the score totals. Others responded only to those events which had happened to them.

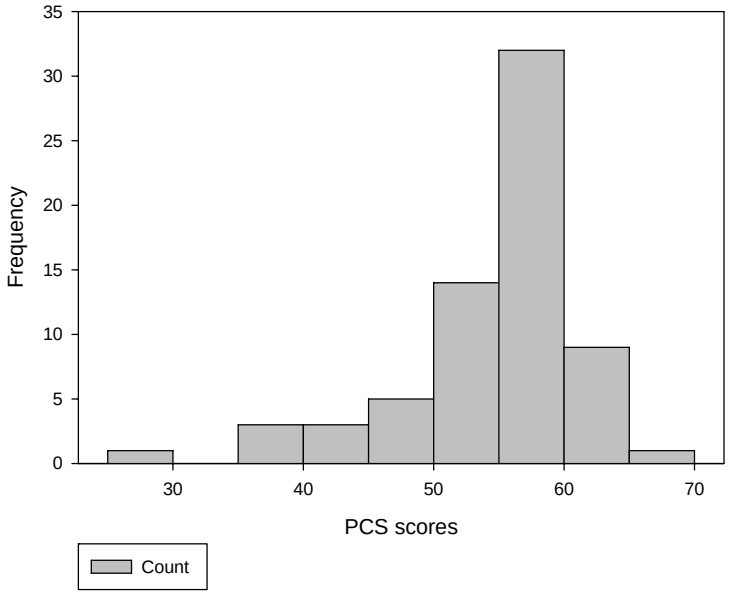


Figure 10.2: Histogram of SF-36 Physical Component Summary T scores: cortisol study

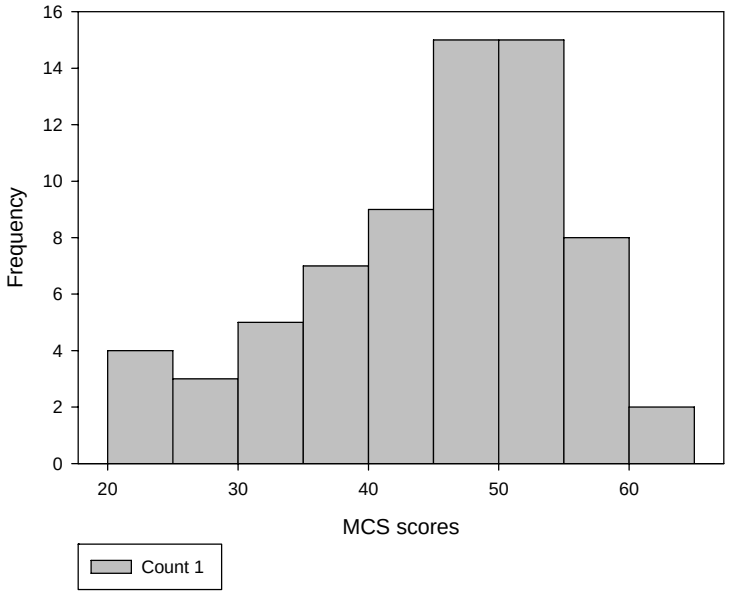


Figure 10.3: Histogram of SF-36 Mental Component Summary T scores: cortisol study

Health behaviours. Confounding from health behaviours was addressed by attempting to exclude, rather than control for, relevant behaviours. Nonsmokers were sampled, and participants were asked not to drink alcohol before and between samples. Averaging the samples across two days (to control for activity) and taking samples across the whole day (to control for nutrition and the influence of heavy meals on cortisol production) should control for daily levels of nutrition and exercise.

10.3.2 Participants

Participants (N = 68, 66.2% female) were non-smoking students recruited from the University of Edinburgh (mean age = 22.11, SD = 7.97). All participants gave informed consent which included the advice that they were free to discontinue the study at any time. All received full verbal and written instructions concerning the study procedures. Table 10.1 shows the demographic characteristics of the sample.

10.3.3 Sample collection

Participants provided saliva samples in plastic vials over two consecutive weekdays. They were instructed, verbally and with written instructions (see appendix) to provide samples 3, 6, 9 and 12 hours after waking, the diurnal profile, over two consecutive weekdays, and to refrain from eating and drinking (except water) 30 minutes prior to sample collection. Participants were asked not to drink alcohol before and between giving samples, nor to brush their teeth before giving a sample. They were instructed to wash out their mouth with water ten minutes prior to giving a sample. Samples were provided by allowing saliva to collect passively in the mouth for two minutes and then transferred using a plastic straw into the vial. Participants delivered their vials to the Department of Psychology where they were frozen in the laboratory freezer within three days. All samples were frozen at -22C for a minimum of 48 hours before cortisol assay to precipitate mucins (to encourage the proteins in the saliva to solidify). Saliva samples

Table 10.1: Demographic characteristics of participants in the cortisol study

	Total	Males	Females
Sample size	68	23	45
Age	21.22 (4.14)	20.35 (3.78)	21.67 (4.28)
White	35 (51.5)	13 (56.5)	22 (48.8)
Mixed	2 (2.9)	0 (0)	2 (4.4)
Asian or Asian British	7 (10.3)	2 (8.7)	5 (11.1)
Black or Black British	1 (1.5)	0 (0)	1 (2.2)
Chinese	23 (33.8)	8 (34.8)	15 (33.3)
Education	19.57 (2.17)	19.57 (2.21)	19.58 (2.17)

were analyzed using the Salimetrics Expanded Range High Sensitivity Salivary Cortisol Enzyme Immunoassay Kit. Averaged intra- and inter- assay variation on the controls was 7.2% and 7.1% respectively[2].

Compliance. As S. Edwards et al. (2001) reiterate, it has been demonstrated that participants from academic communities can reliably and accurately follow these types of instructions and the samples are valid in the assessment of diurnal cortisol activity (Schmidt-Reinwald et al., 1999). The reason for requesting that the samples are provided on weekdays is that previous studies show the CAR is attenuated on weekdays (see Thorn et al., 2006). I am not aware of a study showing attenuation of the DP. However, given that one explanation of the CAR attenuation is compliance may be poorer at weekends, it was safer to restrict the analysis to weekdays.

10.3.4 Cortisol assay

Saliva samples were analyzed using the Expanded Range High Sensitivity Salivary Cortisol Enzyme Immunoassay Kit. The Salimetrics kit is an immunoassay designed for the quantitative measurement of salivary cortisol in research settings. It is highly reliable, correlates with free cortisol in serum, and has a built-in Ph indicator. It is designed to be resilient to the effects of interference caused by collection techniques that affect Ph, such as consumption of food or drink. It is a popular tool for cortisol researchers. The kit contains a microtiter plate. This is a plate, coated with a known quantity of cortisol antibodies. The saliva samples are pipetted into the wells on the plate. The kit measures the reaction between the antigen (in the saliva) and the antibodies (to cortisol). It does this in terms of colormetric signals. That is, antigen-antibody reactions using colour changes. The stronger the colour change, the higher the amount of cortisol in the saliva sample (Chard, 1995). The kit therefore follows the general principle of Enzyme-Linked ImmunoSorbent Assays (ELISA).

[2]Values of 10% or less are considered acceptable.

10.3.5 Procedure

- Saliva samples in plastic vials were thawed fully, then vortexed. Vortexing ensures an even distribution of matter in the tube.
- Samples were then centrifuged at 1500 x g (@3000 rotations per minute, RPM) for 15 minutes. Centrifuging ensures that all the matter in the tube moves to the bottom of the vial.
- The plate layout was determined. The layout determines how the samples are arranged in the wells. Each sample is duplicated as an additional measure of reliability.
- 24ml of assay dilutent was pipetted into a disposable tube (a spare sampling tube) and put aside.
- 25μl of standards, controls and unknowns were pipetted into the appropriate wells. All three of these measures are duplicated.
- 25μl of assay dilutent was assayed into 2 wells, these measures perform as "the zero", or control.
- 25μl of assay dilutent was also pipetted into each non-specific binding (NSB) well.
- A 1:1600 dilution of the conjugate was made, by adding 15 μl of the conjugate to the 24ml of assay dilutent. The diluted conjugate solution was immediately mixed and pipetted 200 μl into each well using a multichannel pipetter.
- This was mixed on a plate rotator for 5 minutes at 500 RPM and incubating at room temperature for an additional 55 minutes.
- The plate was washed four times with 1x wash buffer. A plate washer was not available due to restrictions on resources (it is not essential). I pipetted 300μl of wash buffer into each well, and discarded the liquid by inverting the plate over a sink. The plate was thoroughly blotted on lint free paper towels before being turned upright. The wash buffer was added using the multichannel pipetter, and the order in which I added this was the same throughout all these steps.

- 200μl of tetramethylbenzidine (TMB) solution was added to each well using the multichannel pipette. TMB is a chemical that causes the colour change process.
- The plate was rotated, to mix, on the plate rotator for 5 minutes at 500 RPM and incubated in the dark at room temperature for an additional 25 minutes.
- 50μl of stop solution was added to each well with the multichannel pipette. This stops the colour changing process.
- Mixed on a plate rotator for 3 minutes at 500 RPM. The bottom of the plate was wiped dry with water-moistened lint-free paper.
- Finally, the plate was read in the plate reader at 450 nm within 10 minutes of adding the stop solution.

10.4 The cortisol data

The cortisol values were calculated as instructed in the kit, using KC Junior software. KC Junior is software for plate a reader which has flexible data manipulation and curve-fitting for basic endpoint, kinetic, multi wavelength, and spectral scan requirements. There are seven curve-fit types and extrapolation, standard averaging, and interpolation options. Data is stored in a single MS Access file, with the option of exporting to Excel. The use interface is intuitive, making it attractive for use in research. The software: (1) Calculates the average optical density (OD) for all duplicate wells; (2) Subtracts the average OD for the NSB wells from the average OD of the zero, standards, controls and unknowns; (3) Calculates the percent bound (B/Bo) for each standard, control and unknown by dividing the average OD (B) by the average OD for the zero (Bo); and (4) Determines the concentrations of the controls and unknowns by interpolation a logistic regression curve. This allows the expected and observed cortisol concentrations to be compared, and there should be very high levels of agreement. An R-square of .9 or higher is acceptable.

Missing data. There were three missing data points for samples 2 and 4 on day 1 (4.4%) and two missing data points for sample 1 on day 1, sample 1, 2 and 4 on day 2. When taking the mean of both days, this resulted in no missing data.

Transformation of data. As expected, cortisol values were positively skewed. A logarithmic transformation (base 10, no constant added) was performed so that the scores approximated a normal distribution.

Sex differences. There were no significant sex differences for the samples 3, 6, 9 or 12 hours after waking (all $ps > .05$). Table 10.2 shows the descriptive statistics for the cortisol data, separated by gender.

10.5 Stages of latent growth curve model building

The process used to create a growth curve model for the cortisol data is now described. After this introductory description, the results from the cortisol study will be presented in similar stages.

10.5.1 Step 1: Calculating means and slopes.

The latent means and variances are estimated by fixing the factor loadings of the pathways leading from the latent variables to the observed variables. For the intercept, the factor loadings are fixed at one. For the slope, the factor loadings are fixed at 3, 2, 1, 0. These numbers "reflect the time course under study" (Willoughby et al., 2007, p. 129) and imply three things: (1) there are equally space intervals of sampling; (2) the shape of the slope is linear; (3) that the intercept is the time point where the slope is zero. The effect is that the mean structure is passed from the measurement model to the structural model (Willoughby et al., 2007, p. 128). The slope factor "explains" the linear decrease in cortisol across the four time points. Applying this model to the cortisol data, the following values are obtained for the intercept and slope parameters:

Table 10.2: Descriptive statistics for the cortisol values, separated by gender

Variable	Mean	SD	Mean (males, N = 23)	SD	Mean (females, N=45)	SD
Cortisol[a] 3 hours after waking	0.2090	0.1454	0.2465	0.1462	0.1898	0.1428
Cortisol 6 hours after waking	0.1600	0.1469	0.1553	0.1363	0.1624	0.1534
Cortisol 9 hours after waking	0.1153	0.0724	0.1147	0.0660	0.1155	0.0761
Cortisol 12 hours after waking	0.0764	0.0731	0.0872	0.1100	0.0709	0.0446

[a]Micrograms per decilitre

Intercept. The mean intercept: -.81 (taking the antilog, this means that the average participant has 0.16μl of cortisol three hours after waking).

Slope. The mean slope: -.15 (taking the antilog, every three hours, the average participant's cortisol drops by .71μl).

Intercept variance. Variance of intercept factor: .04

Slope variance. Variance of slope: .00

The mean values are within the expected normal range for healthy adult volunteers. The four values, or parameters, show that the LCM provides a parsimonious model of the data presented in Figure 10.1. All relevant features of the data are captured by these parameters.

10.5.2 Step 2: Is there any variation in the intercepts and slopes?

The second step in LCM is to determine if there is any interesting variation in the intercept and slope. That is, latent variance (Willoughby et al., 2007). If people had similar cortisol slopes and variances, this implies that people have the same slope (K. A. Bollen & Curran, 2005). However, if there are individual differences in the amount of cortisol produced across the day, or in the relative decline in cortisol across the day, this would be interesting. However, the first step is to determine is there is any variation at all in the intercepts and slopes. Visually, Figure 10.1 suggests that there might be. The intercepts are quite different for each participant, although the slopes look quite similar in shape. If there is little variance in intercepts or slopes, this can lead to problems in identifying the model. The maximum likelihood estimator may have difficulty minimizing the discrepancy function between the variation in slopes implied by the model, and the variation actually observed.

Negative variance estimates

When fitting the basic LCM to my data, the latent variable covariance matrix was nonpositive definite, which means that there was a negative eigenvalue. As reported above, the variance of the slopes is near zero (.00). Nonpositive

definite matrices can occur for many different reasons, including small sample sizes, or when running secondary analysis of a published covariance matrix, or even a mistake in the matrix (Wothke, 1993). Another problem that occurred when attempting to fit the model to the cortisol data was the existence of negative variance estimates or Heywood cases (K. A. Bollen & Curran, 2005). Negative variances are statistically impossible in the population, but it is mathematically possible to encounter them in a sample. One strategy, advocated by Chen, Bollen, Paxton, Curran, and Kirby (2001), is to constrain the parameter to a fixed value (discussed below).

The solution. In the cortisol data, the negative variance estimate of -.01 implies that the variance is in fact zero, or that negative estimates are within sampling fluctuations of zero (K. A. Bollen & Curran, 2005, p. 49). Therefore, the slopes can be constrained (fixed) to be zero. When parameters are fixed to zero, the maximum likelihood estimates should be treated with caution in this situation, and chi-square derived significance tests become problematic. It is wise to use these only as "heuristic tools" (K. A. Bollen & Curran, 2005, p. 49). As explained by R. B. Kline (2005, p. 102), the "inequality constraint forces the value of a parameter estimate to be less than or greater than a specified value". Knowledge about what the value should be is often not possible, but in this instance it is clear that the reason for the problem is almost certainly the lack of variance in the slopes. In short, the solution is to simply re-run the model, fixing the variance of the slopes to zero. The participants show individual differences in intercepts, but not slopes.

10.5.3 Step 3: Is there a relationship between intercept and slope?

If variance in the intercept and slope can be estimated, then it makes sense to ask if these latent variables co-vary — latent covariance (Willoughby et al., 2007). Do the intercepts correlate with the slopes? Do people with higher levels of cortisol across the day have less steep rates of decline, for example? Are the intercepts

and slopes uncorrelated, suggesting two separate processes? Unfortunately, because I fixed the cortisol slope variance to zero, these questions could not be explored. If the variance of the slopes is zero, then there can be no covariance with other variables in the model. It is not possible to leave the slopes unfixed, even for other models, because this causes problems with the model estimation.

10.5.4 Step 4: Are individual differences in intercepts or slopes related to covariates?

LCMs containing covariates (variously termed "explanatory variables", "independent variables") are called conditional LCMs (K. A. Bollen & Curran, 2005). Pathways from covariates to intercepts and slopes ask, do these variables predict individual differences in intercepts and slopes? That is, do covariates explain the variance in intercepts? They essentially have the same role here as they do in other models in the generalized latent variable modelling framework (Muthén, 2002). In the cortisol study, gender and personality traits are covariates. Do gender and personality "explain interindividual differences in intraindividual change" (Willoughby et al., 2007, p. 128)? Unfortunately, because the slope variance is fixed at zero, only the intercepts can be regressed on the covariates.

10.5.5 Step 5: Are there indirect effects of the covariates on the intercepts or slopes?

The possibility that covariates interact with time points, is an exciting and relatively new development in LCM. It is likely to have important implications for cortisol research, and I hope to explore this technique in the future. The question asked here is, do the covariates predict individual differences in slopes at certain points of the day? Unfortunately this model is not possible in the cortisol data here, because there is no slope variance.

10.5.6 Step 6: Possible moderation effects

Interactions in regression equations imply that a different regression line is needed for different groups (e.g. males and females). In the cortisol data, this would imply that different models are required for males and females. As another exam-

ple, Willoughby et al. (2007) tested two models for their cortisol data. In one, the parameters for males and females did not change (they were estimated simultaneously), in the other, they were free to vary. A chi-square different test can be used to compare two models. The question asked is, does forcing the groups to be equal significantly worsen the fit of the model to the data? In Willoughby et al. (2007), it did, suggesting that gender was a moderator. However, Oswald et al. (2006) found interaction effects between gender and Neuroticism, and between gender and Extraversion. This led them to suggest that separate models would be required for males and females, and implies that interactions with gender should be considered in cortisol-personality models.

10.5.7 Step 7: Interpretation of conditional model estimates.

The next step in LCM is to regresses the intercept and slope latent variables on relevant covariates. The question asked is, what covariates explain individual differences in cortisol intercepts and slopes? Whereas the unconditional model has no covariates, and is defined by the means of intercepts and slopes (and deviations from those means), the conditional model considers variables that predict intercepts of slopes (K. A. Bollen & Curran, 2005). The covariance coefficients are interpreted in the same way as regression coefficients in that they provide expected difference in the outcome for a one-unit difference in the explanatory variable, net of the other explanatory variables (K. A. Bollen & Curran, 2005).

10.5.8 Covariates in LCM.

Covariates can be categorical (e.g. male, female) or continuous (e.g. personality traits). Because the maximum likelihood estimator is used, dummy coding or non-normality on the covariates are permitted, as long as the observed variables have multivariate normal distributions, no excess multivariate kurtosis (K. A. Bollen & Curran, 2005, p. 126-161). If these assumptions are not met, the ML estimator might still work, but significance tests may be inaccurate (K. A. Bollen & Curran, 2005). In my cortisol data, females are coded 0 and

males coded 1. The intercept of the regression equation is the mean of the intercept (mean cortisol three hours after waking) when all the predictors equal 0. Mplus will provide estimates of the mean intercept and slope for the female group, and differences in mean intercepts and slopes of latent trajectories for the male group compared to the female group. As an example, an intercept of .2 would suggest that males have higher mean cortisol daily output. A slope of 0 would means there are no differences in slopes (K. A. Bollen & Curran, 2005, p. 126-161). When covariates are continuous, interpretation of the model estimates depends on the scale used. It is "the expected change in α_i/β_i resulting from 1-unit change in x", where x is the covariate.

10.5.9 Identification issues

It is necessary to identify a LCM before estimating intercepts and slopes. I explained the process of identification of models in chapter five, so I will limit the discussion to those features that are different about identification for LCMs. Identification refers to the ability of the model to estimate unknown model parameters from the information available. In an LCM, the identified variables are "means, variance, covariances" of the DV, a population mean for each time point, and population variances and covariances for each time point and DV. There are "$\frac{1}{2}T(T+3)$ identified parameters with which to work" (K. A. Bollen & Curran, 2005, p. 23). Regardless of the number of waves of data, the $\frac{1}{2}T(T+3)$ formula "provides the number of means, variances, and covariances available no matter the number of waves of data" (K. A. Bollen & Curran, 2005, p. 23). The rule helps to explain why two time points are not sufficient for LCM or multi-level modelling. There are seven unknowns and five knowns, resulting in more unknowns than knowns (Miles & Shevlin, 2000). If there are more means, variances and covariances than parameters, this is over-identified. If the unknown ones are a function of the known ones, then the unknowns can be identified (K. A. Bollen & Curran, 2005, p. 23).

Four time points are available. K. A. Bollen and Curran (2005) stresses the importance of having as many time points as possible — preferably four or more. Fortunately, the cortisol data has four time points, which is one more than is necessary to identify a simple LCM. Two additional assumptions can be made by the researcher to identify the basic LCM model: (1) the trend values are known (i.e. assume a linear decline represented by factor loadings of 3, 2, 1, 0 for the slope); (2) each case has the same error variance in the same time period, although the variances differ over time (K. A. Bollen & Curran, 2005). The nature of diurnal cortisol decline is well documented, which is fortunate because it allows me to assume a linear decline from three hours after waking to 12 hours after waking. Researchers should know what kind of growth to expect, prior to modelling. When modelling less well known trajectories, such as personality change, several different possible shapes would need to be considered (linear, splines, quadratic, cubic, piecewise etc.).

10.6 Results

It is important to note, given the small sample size, that the findings presented in the models below may occur by chance (type II errors), and that significant associations may have been missed (type I errors). However, as Oswald et al. (2006) note, "since examination of the relationship between cortisol dynamics and personality is still exploratory, we believe that false-negative results may be as important as false-positive results" (p. 1589). In other words, the literature is at a very early stage. Therefore, all results should be reported. The sample size also meant that chi-square derived goodness-of-fit indices (e.g. the chi-square test of exact fit, CFI, TLI and RMSEA) were problematic, because they are not very accurate in small samples. The standardized root mean residual (SRMR), in contrast, is a good indicator of model fit when sample sizes are modest. It provides standardized differences between observed and predicted covariances, is not based on chi-square, and is most sensitive to mis-specified covariances (Hoyle, 1999; L. Hu & Bentler, 1998). Acceptable thresholds for SRMR are not

agreed, but range from .06 to .10 (see chapter five). The latent growth curve models for the cortisol and Neuroticism data were assessed one at a time. There were seven models in total. The descriptive statistics for the Neuroticism data is shown in Table 10.3, and the model fit statistics are shown in Table 10.4:

Model 1. First, the plausibility of a separate LCM for each day was tested. This is called a multivariate or parallel process model. There is one intercept and one slope for each day: four latent variables in total. The model did not converge. Negative slope variance implied that there was little or no slope variance. Therefore, slope variance was fixed to zero.

Model 2. In a model where the slope variance was fixed to zero, the covariance between the intercepts was greater than one, suggesting that mean daytime cortisol output is very similar across both days. Therefore, the covariance between the slopes was fixed to one.

Model 3. Fixing the covariance of the intercepts to one resulted in a very similar fit values. Comparing unconstrained (where the means are freely estimated in each day) with constrained models (where the means are constrained equal) had little effect on the SRMR value (both constrained = .11, both unconstrained = .10, constrained slope unconstrained intercept = .11, unconstrained slope constrained intercept = .10). Because the fit of the more restricted model is not significantly worse than the freely-estimated version, this implies that the average true rate of change in cortisol levels is the same in each day, and therefore using the average level at each time point across both days is acceptable and more parsimonious than a parallel process LCM.

Model 4. A single process model using the average cortisol values for each timepoint across both days, with slopes fixed to zero, showed good fit to the data (SRMR = .05). It was not possible to fit a quadratic growth factor (an additional slope with a time coding of 0, 1, 4, 9). A quadratic growth factor requires a linear intercept and a random slope factor, which was not permissible with these data. The linear term is nested within the quadratic term, such that the latter requires the former (Curran, 2003). The linear

Table 10.3: Descriptive statistics for cortisol study: NEO-PI-R traits and Neuroticism facets

Variable (T-scored)	Mean	SD
N	52.05	10.34
E	47.79	10.77
O	52.03	12.18
A	44.67	11.74
C	46.53	11.72
N1	50.58	11.56
N2	50.35	8.77
N3	52.74	9.69
N4	52.23	10.01
N5	50.55	9.82
N6	52.53	10.44

growth model is most consistent with prior research and with the visual representation of the diurnal decline in Figure 10.1. Therefore, the single process linear growth model was used subsequently, for testing the contribution of covariates (explanatory variables).

Model 5. The big five (N, E, O, A and C) traits, gender, and interactions with gender (N*male, E*male, O*male, A*male and C*male), were added as covariates. The rationale for including interactions with gender followed Oswald et al. (2006), who found that the big five traits interacted with gender. Only Neuroticism was a significant predictor of intercept variance (β = .47).

Model 6. A model containing gender and the Neuroticism facets was tested. This showed that the association with Neuroticism detected by model 5, was accounted for by the N2 Angry Hostility facet (β = .44). One T-score increase on N2 was associated with a standard deviation increase in (log) cortisol intercept values. N6 Vulnerability (vulnerability to general stress, Costa & McCrae, 1992a) approached significance (β = .28, p = .08).

Model 7. Potential confounding factors were added to the model, to check if the association found in model 6 was attenuated: PSS, PCS and MCS scores were included. The parameter estimate for N2 was slightly attenuated (β = .42), but increased slightly for N6 (β = .30, p = .06). PSS, PCS and MCS were not associated with cortisol intercepts, although MCS approached significance (β = -.32, p = .08). Therefore, the pathway from N2 to cortisol intercept variance was not mediated by perceived stress, physical health component scores, or mental health component scores. The results from this model are shown in Table 10.5 and Figure 10.4. Only significant pathways are shown.

A note on the statistical power of model 7. Given the small sample size, a Monte Carlo simulation study was conducted in order to determine statistical power. The procedure is described in Muthén and Muthén (2002). The example provided, referring to a growth model with normally distributed continuous outcomes without missing data, with a covariate that has a regression coefficient of

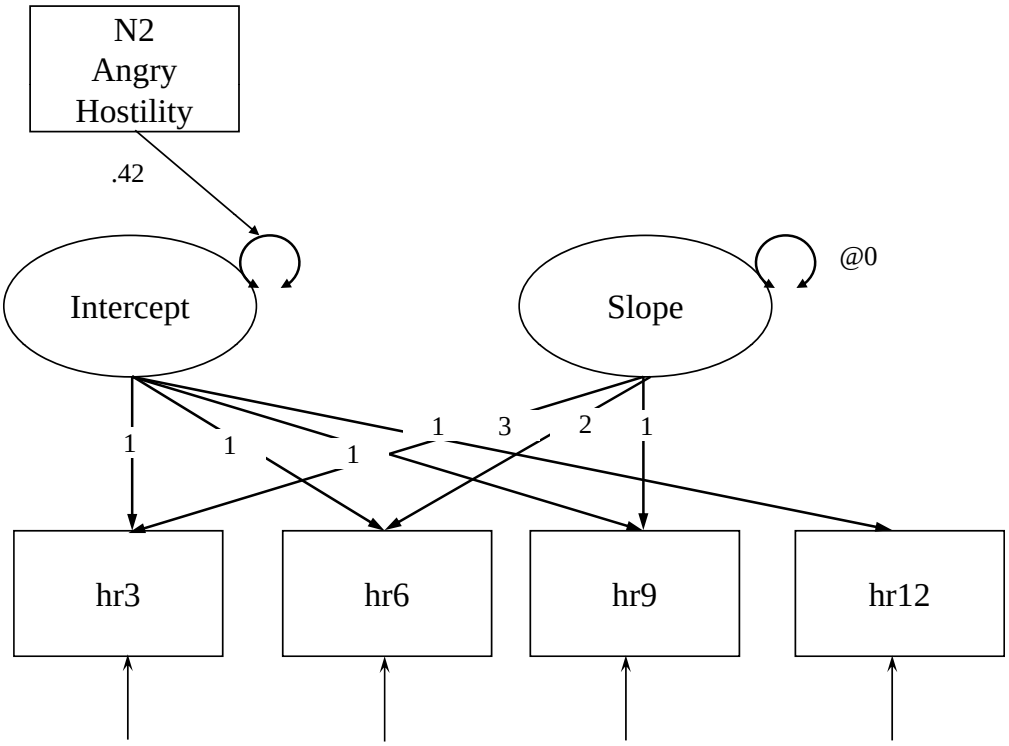


Figure 10.4: Path diagram for latent growth curve model of cortisol and Neuroticism facets, showing a significant pathway, from the NEO-PI-R N2 Angry Hostility facet, to cortisol intercept variance

Table 10.4: Goodness-of-fit statistics for latent growth curve models of cortisol data with Neuroticism facets as covariates

Model number	Chi square	d.f.	p	CFI	TLI	AIC	BIC[a]	RMSEA	SRMR[b]
1	30.51	22	.11	.93	.91	-157.99	-142.74	.08	.09
2	43.12	29	.04	.88	.88	-164.30	-142.74	.09	.10
3	44.99	32	.06	.89	.90	-165.23	-142.74	.08	.10
4	4.51	7	.72	1.00	1.04	-47.20	-44.95	.00	.05
5	44.94	40	.27	.93	.91	-3144.21	-3121.74	.04	.05
6	28.87	31	.58	1.00	1.04	-1870.82	-1856.39	.00	.05
7	34.48	37	.59	1.00	1.04	-2173.35	-2156.11	.00	.05

[a]Sample size adjusted
[b]This measure of fit is not derived from chi-square, and is therefore more appropriate for this data.

Table 10.5: Parameter estimates for latent growth curve models of cortisol data with Neuroticism facets as covariates

Variable	Estimate	Standardized estimate[a]	Standard error	p
Male	-.02	-.07	.12	.78
N1 Anxiety	.00	.18	.19	.34
N2 Angry Hostility	.01	.42	.14	.00
N3 Depression	.00	.09	.17	.61
N4 Self Consciousness	.01	.30	.16	.06
N5 Immoderation	.00	.00	.14	1.00
N6 Vulnerability	-.01	-.29	.19	.12
Perceived Stress	-.01	-.30	.21	.16
SF-36 Physical Component Summary	.00	.10	.12	.39
SF-36 Mental Component Summary	-.01	-.32	.18	.08

[a]Because male is a categorical covariate, the standardized estimate STDY is reported. This is interpreted as units of standard deviation change when the covariate changes from 0 (female) to 1 (male)

.2 for the slope growth factor, was modified (p. 16). I adapted this example because it is most similar to model 7, presented in Figure 10.4. For the purposes of this analysis, the slope variance was fixed to zero, the covariate regression coefficient was changed to .4, and the sample size was set at 68. In 10,000 replications, each simulation tests the null hypothesis that the regression coefficient parameter (the pathway from the covariate to the random intercepts) is equal to zero. Given that the population model set the coefficient to .4, the simulations should reject this hypothesis at least 80% of the time, to meet the standard criterion of 80% power. The results showed that this occurred 77.8% of the time. Therefore, the model is slightly under-powered. One limitation of this simulation, is that only one covariate was considered. There is unlikely to be sufficient power to model several covariates at the same time, and their intercorrelations.

10.7 Discussion

This chapter illustrates that cortisol must be modelled in a way that is appropriate to the data. Cortisol intercept variance, which indicates cortisol output across the day, was positively associated with N2 Angry Hostility. The methodology provides a way to reveal, and maximize (Pruessner et al., 1997), correlations between cortisol, traits, and other measures relevant to personality-health research. The results reinforce the important of using the more precise NEO-PI-R (Costa & McCrae, 1992a), rather than the shorter NEO-FFI because associations appear at the facet level, which are not available in the NEO-FFI. The results build on the recent report that facets of Neuroticism and Extraversion are differentially related to cortisol responses to stress in an experimental setting, the only report to consider facets so far (Oswald et al., 2006). This is the first report, to my knowledge, of such a pattern in daytime cortisol output in naturalistic real-life settings. The results are interesting, given the very recent findings published by Wasserman et al. (2007). Their results showed an association between N2 Angry Hostility and a gene region called TBC19, which is involved in HPA axis activity. The sample was restricted to patients who had attempted suicide, and has not

yet been replicated. However, it provides a testable hypothesis that the N2 facet may be involved in diurnal cortisol output.

10.7.1 Strengths and weaknesses

There were two key strengths to this study. First, the saliva samples began three hours after waking, were synchronized to waking time, not clock time, and repeated over two days. This approach ensures that the analysis is specific to the diurnal profile: high levels of cortisol from the waking profile are excluded. It also allows for individual differences in waking time, avoids measurement problems associated with weekends, and improves reliability (and here, missing data) by re-testing participants on the second day. The samples were provided in naturalistic, ambulatory settings by participants during their normal daily routine, enhancing ecological validity. Diurnal decline was observed for nearly all participants, showing that compliance with the sampling protocol was observed. I also observed small yet significantly higher cortisol output on the first day (see chapter four), which suggests that the sampling protocol elevates cortisol on the first day. Second, the study was among the first to explore associations at the facet level of personality (e.g. N2 Angry Hostility). To date, most studies have reported association at the trait level (e.g. Neuroticism and Extraversion). This can be misleading, if facets are differentially related to cortisol. The data reported here not only provide additional criterion validity for the NEO-PI-R facets (O'Connor & Paunonen, 2007), which is very important, but suggest that cortisol research into personality may bear more fruit at the facet level. It is noteworthy that the facets provided connect with constructs that are well researched in the cortisol literature — depression, general vulnerability to stress, and positive emotions. The NEO framework, as part of the big five, provides a benchmark (de Raad & Goldberg, 2002) with which to integrate personality research with other cortisol studies, and can provide clues about which constructs to study more precisely in the future.

Depression in the participants

Clinical depression was not measured in this study, leaving open the possibility of confounding by latent depression. However, the MCS scale captures variance in mental health in the population. Furthermore, MH and MCS cut-off scores of 52 and 42, respectively, can detect depression (Ware et al., 1993; Silveira et al., 2005). This suggests that a large proportion of the sample were indeed depressed. According to the MCS criterion, 22 (32.4%) were depressed. According to the MH criterion, 43 (63.2%). The study was advertised as a "Personality and Stress Study" which may have attracted depressed volunteers. In future research, it would be advisable to measure clinical depression and exclude these participants. Given the results from the HAPPLE study, which showed low MCS scores compared to the general population, it may be useful to explore participants' motivations for wanting to participate in a personality-health study. However, the limitation can be addressed in these data, because including MCS scores in the model controlled for the influence of variation in mental health status. The association with N2 was not attenuated when MCS scores were controlled for.

Lack of variance in the slopes

One feature of the data was that very little variation in cortisol slopes (morning to evening decline) was observed. This meant that predictors of individual differences in slopes could not be explored. However, disruption to diurnal cortisol rhythms is associated with specific illnesses, which would lead to individual variability in daily rates of change on cortisol. Lack of individual differences in slopes may reflect the healthy characteristics of the sample. A second weakness was the relatively small sample size. Popular measures of model fit (e.g. CFI, RMSEA) are not readily interpretable with very small sample sizes (Hoyle, 1999). However, the sample size is as large as (Oswald et al., 2006) and larger than many existing published studies of cortisol and personality. Significant associations were detected, even with the modest sample size. However, false-negative results may have occurred, which are equally as important as false-positive results

(Oswald et al., 2006, p. 1589). As the cost of cortisol assays continues to fall, personality researchers may find that they have the resources available to increase sample size, and the number of days of sampling, in order to improve the quality of data sets available. Finally, there were several measures which I did not control for, such as sleep quality, clinical depression, exercise, daytime activity, meal times, and body mass index. These could be incorporated into future research designs.

10.7.2 Strengths and weaknesses in relation to other studies

As discussed in chapter nine, evidence that Neuroticism is associated with the diurnal profile is mixed, at best. My results may provide clues about why the literature has been so inconsistent. First, the associations may be specific to the facet level. When associations are reported with Neuroticism, it could be that variance from depression or facets, such as N2 Angry Hostility, are involved. Third, the associations may change depending on the cortisol profile being studied. Whereas I studied the diurnal profile and mean daytime cortisol output, other researchers Oswald et al. (2006) considered the state profile — cortisol responses to experimentally induced stressors. In that study, Extraversion was significant for males (E1 Warmth, E4 Activity and E6 Positive Emotions), and Neuroticism for females (N3 Depression and N4 Self Consciousness). The association with N4 Self Consciousness is understandable given that the Trier Social Stress Test (TSST) involves being observed while performing mental tasks. However, there is a clear need for replication and the development of theories to explain any associations that are repeated in the literature. The fact that gender was a moderating variable, but was not in my study, suggests that associations may depend on which cortisol profile is being studied (e.g. waking, diurnal, or reactivity). Furthermore, personality associations have sometimes been limited to specific groups, such as breast cancer patients (Vedhara et al., 2006). In that study, Neuroticism was associated with cortisol morning peak (not diurnal rhythm) for

patients (not for controls). Researchers should therefore clearly describe the cortisol profile being studied, and look for interaction effects with illness status, as well as gender.

10.7.3 Meaning of the study

The associations reported here between daytime cortisol output, and facets of the NEO-PI-R, are the first available, to my knowledge. They should be regarded as exploratory, in need of replication, before strong conclusions are made about the meaning of this association. Considered alongside (Oswald et al., 2006), the results suggest that Neuroticism and Extraversion are the most relevant traits for future research. Sexual dimorphism in HPA activity (Kudielka et al., 2004; Traustadóttir, Bosch, & Matt, 2003) might help explain why different traits mediate cortisol output for males and females in laboratory settings, but not in the naturalistic settings considered here. Finally, it would be important to compare the N2 Angry Hostility facet with other measures of aggression and hostility from different inventories, because there are important differences in the way that different inventories measure these constructs (see M. C. Whiteman et al., 2001). I turn now to the question of self-reported functional health status, and the possibility that the model described above can provide any insights into the possibility that SF-36 scores are related to cortisol.

10.8 Cortisol output and the SF-36 scales

In this section, I return to an unanswered question. Could circadian cortisol profiles operate as biomarkers for general health status? In this chapter, an appropriate method for modelling the diurnal cortisol profile has been presented. Therefore, the SF-36 can now be considered in relation to individual differences in cortisol intercepts. This section briefly addresses this outstanding issue of convergent validity — the extent to which these two measures overlap.

10.8.1 Results

The decision was made to focus on the scales, rather than the PCS or MCS, so that a more precise estimate could be obtained about how cortisol might be associated with functional health status. As described in chapter four, there are eight SF-36 scales: Physical Functioning (PF), Role-Physical (RP), Bodily Pain (BP) and General Health (GH),Vitality (VT), Social Functioning (SF), Role-Emotional (RE) and Mental Health (MH). These are converted into T-scores, with a mean of 50 and a standard deviation of 10. The model results are shown in Table 10.6:

Model 1. Model 4 from the analyses presented above was adopted as a starting point.

Model 2. The SF-36 scales and gender, were included as covariates. The General Health (GH) scale of the SF-36 was the only statistically significant pathway. GH was negatively associated with cortisol intercept variance (β = -.36). One T-score increase in GH scores was associated with a .36 standard deviation reduction in (log) cortisol intercept values.

Model 3. Perceived stress, a potential confounding factor, was added to the model. GH remained negatively associated with cortisol intercept variance, and the parameter estimate was not attenuated (β = -.36). This shows that the GH pathway is not confounded by PSS scores. The path diagram for this model is shown in Figure 10.5, and the parameter estimates are shown in Table 10.7. Only significant pathways are shown.

10.8.2 Discussion

These results demonstrate that in addition to N2 Angry Hostility, the SF-36 GH scale and cortisol intercepts are related. The GH scale measures general health perception (mental and physical). The correlation is not very high, suggesting that the SF-36 and cortisol measures different aspects of functional health status. The GH association could be a type I error (false positive), given the small sample

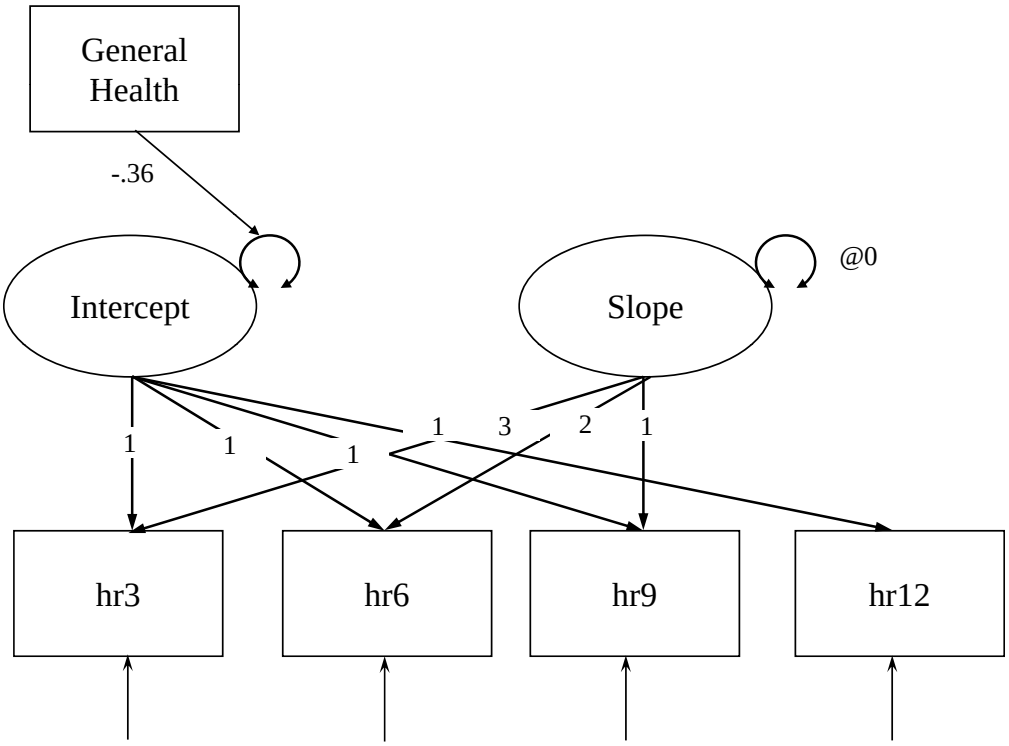


Figure 10.5: Path diagram for latent growth curve model of cortisol and SF-36 scales, showing a significant pathway, from the SF-36 General Health scale, to cortisol intercept variance

Table 10.6: Goodness-of-fit statistics for latent growth curve models of cortisol data with SF-36 scales as covariates

Model number	Chi square	d.f.	*p*	CFI	TLI	AIC	BIC[a]	RMSEA	SRMR[b]
1[c]	4.51	7	0.72	1.00	1.04	-47.20	-44.95	0 (.00 to .11)	.05
2	37.55	34	.31	.95	.94	-1939.88	-1921.11	.04 (.00 to .10)	.06
3	42.39	37	.26	.92	.90	-2160.34	-2139.14	.05 (.00 to .10)	.05

[a]Sample size adjusted
[b]This measure of fit is not derived from chi-square, and is therefore more reliable than the other measures reported.
[c]This model is that same as model four in Table 10.4

Table 10.7: Parameter estimates for latent growth curve models of cortisol data with SF-36 scales as covariates

Variable	Estimate	Standardized estimate[a]	Standard error	p
Male	.04	.21	.13	.46
Physical Functioning	.00	.17	.14	.21
Role Physical	.00	.08	.16	.60
Role Emotional	.00	-.18	.17	.29
Energy Vitality	.00	-.01	.17	.98
Mental Health	.00	-.16	.22	.46
Social Functioning	.00	.15	.18	.39
Pain	.00	-.04	.15	.78
General Health	-.01	-.36	.14	.01
Perceived Stress	.00	-.01	.20	.96

[a]Note. Because male is a categorical covariate, the standardized estimate STDY is reported. This is interpreted as units of standard deviation change when the covariate changes from 0 (female) to 1 (male)

size. Associations for the other scales could have been missed, particularly if they are small (type II error, false negative). Cortisol may be associated with illness, but not with self-reported functional health status. Restricting the sample to healthy volunteers may have resulted in restricted range on the cortisol variable, and on covariates. Inevitably, there are several unmeasured variables that might have resulted in additional insights. Confounding factors such as sleep quality, caffeine withdrawal, body mass index (BMI), and morningness or eveningness traits could be measured in future research studies. Several mechanisms could underlie the association between GH and cortisol. Additional variables could mediate the association. For example, increases in GH could encourage activity, positive affects, and social engagement, which could lower cortisol. Reverse causality is very plausible in models of cortisol, especially given that HPA axis activity is cyclical in nature. Another mechanism is the "common cause" hypothesis that a "third variable" could influence both GH and cortisol. There might be shared genes that influence both HPA axis activity and subjective experience of health and illness, such circadian rhythmicity. Finally, an antecedent variable could occur earlier in time to changes in GH and cortisol, such as an illness with persistent and long lasting symptoms. One limitation of this analysis is that, like the HAPPLE study, the SF-36 scale scores are significantly different from the general population, as shown in Table 10.8. General Health scores are higher, for males and females. However, there was enough variability on the GH scale to detect an association with cortisol intercept variance.

10.8.3 Implications

The results have several important implications. First and foremost, they provides a point of connection between studies of cortisol and health outcome and quality of life research. These separate literatures are concerned with both general and specific aspects of health. For example, the SF-36 provides general (PCS, MCS) component summaries, as well as specific scales with their own criterion validity (GH). This five-item scale may be particularly useful to add to future cortisol studies where there is limited scope to include the full SF-36 or its shorter

Table 10.8: Comparison of cortisol study SF-36 scale scores with the general population.

	U.K. males[a]		U.K. females		Cortisol study males				Cortisol study females				
Variable	Mean	SD	Mean	SD	Mean	SD	t	p^b	Mean	SD	t	p	
PF	89.76	18.78	86.66	20.15	95.65	13.76	-3.47	$< .0001$	89.44	21.75	-1.05	.30	
RP	89.01	21.09	85.83	22.52	85.87	28.03	.92	.36	88.89	24.75	-1.01	.32	
BP	81.25	22.21	76.97	23.44	88.26	12.42	-4.53	$< .0001$	87.44	15.16	-5.60	$< .0001$	
GH	70.86	20.29	71.28	20.54	74.57	15.29	-1.97	.05	74.67	13.46	-2.04	.04	
EV	60.81	18.93	55.91	19.85	63.26	12.93	-1.53	.13	61.00	15.98	-2.60	.01	
SF	84.71	22.56	81.33	23.62	82.07	15.91	1.34	.18	78.33	19.47	1.26	.21	
RE	88.08	19.91	84.07	21.79	82.61	33.14	1.36	.18	67.41	38.59	3.55	$< .0001$	
MH	74.32	17.24	70.05	18.65	76.17	12.39	-1.22	.22	67.82	16.29	1.12	.27	
PCS	50.63	9.41	49.54	10.40	53.79	6.63	-3.86	$< .0001$	54.24	7.05	-5.41	$< .0001$	
MCS	51.16	9.34	49.17	10.39	47.87	8.36	3.21	$< .0001$	43.82	10.44	4.19	$< .0001$	

[a]See Jenkinson et al. (1999).
[b]Two-tailed.

versions. Cortisol has been association with general health and specific illnesses (see chapter nine). The hypothesis that cortisol profiles are a marker for "health", not just for stress, is worth pursuing. A more precise association should be sought, in a larger sample, and the SF-36 provides a useful departure point for further study. Its strength lies in its reliability and validity. The nature of the association could be studied in specific illnesses, specific populations, and discovered at specific parts of the SF-36 hierarchy. It would be particularly informative to include the SF-36 in longitudinal studies of cortisol, given its sensitivity to change. In conclusion, subjective instruments such as the SF-36 "measure something more — and something less — than objective medical ratings" (Maddox & Douglass, 1973). These results show that the same statement applies to cortisol intercepts.

Part IV

Conclusion

CHAPTER 11

Conclusion

> *[A]nalytic choices can mask significant effects* (Wiebe & Fortenberry, 2007, p. 150).

11.1 Introduction

In this thesis I have made two key claims. First, there are a number of important gaps in the personality-health literature. Second, that these gaps can be addressed using the latent variable modelling framework (Muthén & Muthén, 1998–2007) and its extensions (Muthén, 2002). This concluding chapter begins by summarizing the previous work that had been done in the area. Following that, I show how my work relates to this and what I intend to add to the conversation. The intended readership for the research findings is personality-health researchers and students of differential and health psychology. However, the results may have wider appeal. The notion that psychological traits could influence cortisol, which could then influence physical and mental health outcomes (Praag et al., 2004; Tyrka et al., 2008), is only beginning to become accepted by the wider scientific community. My work underscores the point that "it is imperative to link personality to unambiguous measures of physical health if we are to make lasting contributions beyond the disciplinary bounds of personality and

health psychology" (Wiebe & Fortenberry, 2007, p. 142). Interdisciplinary work is required in order to make these links.

11.1.1 Relation to the broader field

The research described in this thesis can be defined broadly as biopsychosocial. However, it also fits into the health behaviour and stress moderation models in personality-health research (T. Smith, 2006). There may be scope for integrating health behaviours with cortisol, into both these models, in future research.

The stress moderation model. This model assumes "that stress can cause illness, and personality acts by making one more or less vulnerable to its deleterious effects" (Wiebe & Fortenberry, 2007, p. 138). Chronic over-activation of the HPA axis, stress-induced immunosuppression, and chronic inflammation; are among the negative outcomes that can be responsible for illness (E. O. Johnson et al., 1992). Cortisol can also influence mental health outcomes, such as depression (Praag et al., 2004) and anxiety (Tyrka et al., 2008). High Neuroticism scorers appraise life events as stressful, and are more likely to experience negative life events. This could raise cortisol output, and since Neuroticism is a stable trait, could chronically over-activate the HPA axis. When measured correctly (see chapter nine), cortisol is an unambiguous and objective measure that can be used to measure specific aspects of stress processes.

The health behaviour model. In the health behaviour model "personality is hypothesized to affect health by influencing one's engagement in health-enhancing or health-damaging behaviours" (Wiebe & Fortenberry, 2007, p. 139). This is not mutually exclusive from the stress moderation model, but is often approached separately. Many studies find that traits influence behaviours (e.g., Booth-Kewley & Vickers, 1994; Bogg & Roberts, 2004). It is well known that these health behaviours influence health in turn (World Health Organization, 2002). However, fewer studies have included social and demographic variables (cf. Hampson,

Goldberg, et al., 2007; Monden & Kraaykamp, 2006). Health behaviours are socially patterned (Jarvis & Wardle, 2006), so that socio-economic variables may mediate trait-health behaviours relationships (Hampson, Goldberg, et al., 2007) or even moderate them. Socio-economic status, and education in particular, are important variables that need further study in relation to the health behaviour model.

11.1.2 Relation to theory

My thesis results have relevance for theories of personality and health. First, cortisol has been identified as a clear contender for a mechanism linking Neuroticism with illness, and it deserves a place in the stress moderation model (T. Smith, 2006; Wiebe & Fortenberry, 2007). However, not all studies find an association between Neuroticism and cortisol, Neuroticism and illness, nor between Neuroticism and mortality. If the stress moderation theory is correct, all three of these associations should be robust. Having said that, it is not possible to discount the possibility that differences in sampling and methodology explain the inconsistencies in the literature. As I showed, the sampling protocol and modelling strategy for cortisol are crucial. Regarding Neuroticism and illness, Shipley et al. (2007) observed that "length of follow-up time, sample size, mean age of the subjects, educational achievement, and initial health" (p. 923) differ between studies, and only two studies had variance on age (Christensen et al., 2002; Huppert & Whittington, 1995). This may explain why Neuroticism is not always associated with mortality. My study showed that methods and modelling techniques may explain why the association between Neuroticism and cortisol is not always observed. They may imply, if replicated, that the N2 Angry Hostility facet is involved in pathways to illness, via cortisol output. The second relevance of my findings to theory concern the health behaviour model (T. Smith, 2006; Wiebe & Fortenberry, 2007). Following Hampson, Goldberg, et al. (2007), the results suggest that health behaviours are not the only mediator of the relationship between Conscientiousness and health outcomes. They also suggest

that mental, as well as physical, health outcomes, can be considered simultaneously. Although not replicated in the HAPPLE data set, SES should feature prominently in future models that contain traits, health behaviours, and health outcomes. This is discussed in more detail below.

11.1.3 Relation to specific studies

Having described where my research fits into wider theoretical frameworks, I will now describe how it extends prior personality-health research, and helps the field progress. I address health behaviour studies first, then studies involving traits and cortisol.

Studies involving health behaviours

The Terman Study. In the Terman Study, conscientious children died later, which was only partly explained by health behaviours and by the effect of other traits (Friedman et al., 1993, 1995). Cheerful children died younger, but this was explained by unhealthy behaviours (smoking, alcohol use and risk-taking propensity). Personality traits were based on six dimensions created from teacher's ratings of the children. This study highlighted the need to include Conscientiousness, because it was associated with survival and health behaviours, making the case for using the big five model of personality rather than the Eysenck model (H. Eysenck & Eysenck, 1964). Furthermore, it illustrated the need to include facets, because these were not available. The HAPPLE study was designed to address this limitation, by using the IPIP NEO, which contains facets. In addition, a wider range of health behaviours were included.

The Edinburgh Artery Study. In this study, hostility predicted cigarette smoking, and submissiveness predicted non-fatal myocardial infarction. The trait submissiveness, for example, was associated with lowered risk of non-fatal myocardial infarction (MI, M. Whiteman, Deary, et al., 1997). Hostile thoughts predicted alcohol consumption, and hostile behaviours predicted cigarette smoking (M. Whiteman, Fowkes, et al., 1997). This study demonstrated the benefit

of including clear, unambiguous measures (Wiebe & Fortenberry, 2007) of health behaviours and health outcomes. The big five was later added to this study, but this was the NEO-FFI not the NEO-PI-R (M. C. Whiteman et al., 2001). Again, the IPIP NEO was used in the HAPPLE study so that traits and facets would be available. In HALS data set, I selected cause specific mortality because, like M. Whiteman, Deary, et al. (1997), the outcome measure is unambiguous.

The Dunedin Multidisciplinary Health and Development Study. Several health behaviours (smoking, alcohol use, violent crime, sexual behaviour and driving habits) were measured in this study, based in New Zealand (Caspi et al., 1997). Traits were measured using the Multidimensional Personality Questionnaire, which provides 10 scales, at aged 18, but temperament was assessed from age 3. Several time points were included, so that the researchers could model the onset of risky health behaviours by age 21. This study highlighted the need to include a wide range of health behaviours, including sexual behaviour and traffic related risk behaviour. The results were analyzed using MANOVA, and were complex. These might have been better communicated in the generalized latent variable modelling framework. The HAPPLE study included a similarly wide range of health behaviours, and although it does not have more than two waves, latent variable modelling was used to model the complex health criteria. MANOVA is a special case of latent variable modelling (Miles, 2003).

The Hawaii Cohort. Several reports have emerged from this longitudinal study (Digman, 1989; Hampson et al., 2001; Hampson & Goldberg, 2006). Hampson, Goldberg, et al. (2007) showed that educational attainment and health behaviours mediate the association between traits and health outcomes. Health outcomes were defined by body mass index and items from the SF-36 measuring general health and functional status. One major strength of this study was that childhood traits (although not facets) and adult health behaviours were available. It also demonstrated that social variables can be modelled alongside traits, health behaviours and health outcomes. The HAPPLE study was designed to record a wider range of SES indicators. Although an insufficient number of participants

reported their education, occupation and postcode; it was possible to test the role of education as a potential mediator in a similar way. The results showed that Conscientiousness was associated with mental, but not physical, health. These differences in results are likely due to the different demographic characteristics of the samples, rather than the modelling strategy, which was very similar.

The Heidelberg Study. Matthews, Yousfi, Schmidt-Rathjens, and Amelang (2003) found that clusters of health outcomes were related to personality traits but not single health outcomes and there was little difference between healthy and non healthy participants in terms of personality. This German data set (Amelang, Schmidt-Rathjens, & Matthews, 1996; Amelang, 1997) is extremely valuable because it will provide data on many different objectively measured illnesses (Matthews, Deary, & Whiteman, 2003). However, it was the possibility of latent clusters of illnesses that encouraged me to explore the possibility of latent class analysis of the HAPPLE SF-36 data. There were three latent classes (healthy, mental healthy, and physical healthy), which raise the possibility of health "types". This type of model could also be tested for health behaviours, or for personality traits (e.g. B. P. Chapman et al., 2007).

The Health and Lifestyle Survey (HALS). As described in chapter eight, the HALS study contains data on deaths up to 2005, in addition to Neuroticism and Extraversion (Shipley et al., 2007). Neuroticism was associated with all-cause mortality and death from CVD, while Extraversion was associated with protection from respiratory disease only, and was a risk factor for CVD in adults aged 40 to 59. The authors suggest that this could represent a chance finding (Shipley et al., 2007). These results showed that traits may have broad and specific relationship with mortality, and morbidity. Several health behaviours were controlled for, but reference was not made to latent variables of health behaviour (Boniface & Tefft, 1997). These results also prompted me to consider CVD mortality (chapter eight) rather than all cause mortality, since the big four are specifically related to CVD risk.

Studies involving traits and cortisol

The studies described above make important contributions to the field. Although each has strengths and weaknesses, the findings are easier to assimilate than those from the stress moderation model. Under this model, stress is operationalized in many different ways throughout the literature. Even cortisol, a specific measure, can refer to the waking profile, diurnal profile, or reactivity profile. The Neuroticism-diurnal profile literature, which I chose to focus on, was inconclusive. Pruessner et al. (1997) were concerned that correlations between cortisol and personality traits were not being maximized; a concern also implied by Li, Chiou, and Shen (2007); Hruschka et al. (2005). Aggregation was clearly required, but aggregation can mean many things to many people. Aggregation is necessary to "uncover" (Pruessner et al., 1997, p. 616), that is, reveal and maximize, the true association between traits and cortisol. Or, as these authors put it, "a personality trait related endocrine response disposition" (p. 616).

Neuroticism facets and the reactivity profile. In contrast, cortisol reactivity had been consideration in relation to the NEO-PI-R in a well designed study (Oswald et al., 2006). Gender moderated the association, because higher cortisol was associated with high Neuroticism in women (and the facets Anxiety, Anger, Hostility, Depression, and Self-Consciousness), and lower Extraversion in men (facets of Warmth, Activity, and Positive Emotion). In fact, several studies have considered Neuroticism in relation to cortisol reactivity (e.g., Schommer et al., 1999; Vedhara et al., 2006). The reactivity literature is much larger than the diurnal literature. Some research has taken the findings further downstream, showing that Neuroticism and cortisol can influence response to the hepatitis B vaccination (Phillips et al., 2005). The Oswald et al. (2006) study was highly informative but left open the question of the contribution of traits to circadian cortisol rhythms. Latent growth curve modelling of the reactivity profile (Willoughby et al., 2007) had been demonstrated, but this was again applied to reactivity data. I chose not to study reactivity, seeing a clearly need for more work on the circadian aspects

of cortisol. This was the motivation for undertaking the research, and for conducting the systematic review. There was a set of gold standard methodological criteria available and waiting.

11.1.4 Relation to the research questions in the introduction

The research began with five research questions (see introduction). Each was designed to address a gap in the personality-health literature, and ensure that the research was related to what had been studied before. I now return to these questions:

Facets. First, can traits and facets be modelled at the same time, and related to health behaviours? This would contribute to the resolution of the bandwidth-fidelity dilemma (J. Hogan & Roberts, 1996a; Ones & Viswesvaran, 1996; Paunonen, 1998; Paunonen & Ashton, 2001; Paunonen, 2003). This dilemma is a central problem for personality research, emphasizing that facets offer predictive validity and utility. The results from chapter seven illustrate that it is possible to combine facets and traits in the same model. Using parcelling, which reduces the sample size required, confirmatory factor analysis can model pathways from facets, to broad health criteria, after controlling for general trait variance.

SES. Second, does socio-economic status (and other social/demographic variables) mediate the relationship between Conscientiousness, health behaviours and health outcomes, following Hampson, Goldberg, et al. (2007)? The HAPPLE study was designed to address these issues. Unfortunately, it was not possible to model SES as a latent variable. Few respondents reported all three of the indicators (education, occupation and deprivation). Education was used as a proxy, but this was not shown to mediate an association between Conscientiousness and health outcomes. This likely reflects the lack of association between Conscientiousness and SF-36 PCS score in the HAPPLE data set, rather than the absence of an association in the population. The results of this chapter provided evidence that Conscientiousness was

associated with higher mental health scores and that this was not mediated by a general health behaviour latent variable, or by education. The study also contributed to the field by providing new evidence of test-retest stability. The IPIP-NEO can now be regarded as a reliable and valid measure of the big five, since it shows test-retest stability in addition to criteria of high reliably and factor structure that had been previously demonstrated (Buchanan, Johnson, & Goldberg, 2005). The study also provided evidence that Internet data collection provides variance on traits, demographic variables and self-reported health outcomes.

Cortisol. Third, are traits such as Neuroticism, related to the diurnal cortisol profile? The systematic review, and then the stress study, were designed to answer this question. It extends the Oswald et al. (2006) analysis of the cortisol reactivity profile, by showing that Neuroticism, and N2 Anxiety in particular, are associated with the diurnal profile. This association remains after controlling for the other Neuroticism facets, sex, perceived stress and self reported physical and mental health status. It is not explained by smoker status, alcohol use, medications, or age (these variables were controlled for by the inclusion criteria for the study). The methodology required in order to test this association in the diurnal profile was carefully designed. By using the NEO-PI-R (Costa & McCrae, 1992a; Rust & Lord, 2006), this allowed direct comparison with their reactivity findings. As an additional benefit, the stress study allowed me to show that self reports of general health *are* related to cortisol output across the day. Testing the hypothesis that cortisol is a marker for general health status, is an important priority for future research.

Aggregation. Fourth, can an appropriate model be found that aggregates cortisol data? The results from chapter 10 extended those of K. A. Bollen and Curran (2005); Willoughby et al. (2007). Latent growth curve modelling models the diurnal decline appropriately, so that an association can be detected and maximized between Neuroticism and cortisol. No prior study, to my knowledge, had used appropriate methods, and appropriate models, to

detect a reliable association. Wiebe and Fortenberry (2007) recently argued that "analytic choices can mask significant effects" (p. 150). My results confirm that analytic choices can *reveal* significant effects. The strongest and most important statement that can be made from the stress study observations is that diurnal cortisol should be measured and modelled correctly, before attempting to find its association with traits.

Behaviours. Fifth, should multiple health behaviour covariance be considered in its own right, rather than "controlling" for the effects of other health behaviours, as is traditional. So-called "ultimate" (O'Toole & Stankov, 1992) validity, was indeed shown by the predictive validity of a lifestyle latent variable on cause specific mortality. Clearly, more research is needed to find explanatory variables further "upstream" of the "lifestyle" factor. What causes the covariance? This chapter raised additional questions for future research, such as the possibility of latent classes. Different combinations of scores on health behaviours (and health outcomes) may cluster into latent classes in the population. Traits, health behaviours and health outcomes may not have the same relationships in each class.

These five questions are distinct, but related, areas of concern. The five research questions are related because they all concern issues of modelling. Arguably, emphasis on specific modelling strategies may have contributed to these problems in the literature, because it had de-emphasised the similarities between models. The generalized latent variable modelling framework allows for multiple working hypotheses to be explored and suggests avenues for future research. Overall, there was a good match between the research questions and the approach taken.

11.1.5 Relation to the generalized latent variable modelling framework

Each research questions was addressed by choosing an appropriate set of measures, and an appropriate model. The theme that united each chapter was its relationship to the general LVM framework. Each model described is a "special case" of this framework. For example, survival analysis is an ellipsis B model

(Muthén & Muthén, 1998–2007; Muthén, 2002) because it contains a categorical latent variable (alive, dead, or censored). Latent growth curve modelling is an ellipsis A model because it contains a continuous latent variable. The generalized latent variable modelling framework has benefits over and beyond maximizing associations: it provides a common language. Using the terms "continuous covariates", "categorical covariates", "continuous latent variable", "categorical latent variable", "continuous outcome", "categorical outcome" (Muthén & Muthén, 1998–2007) provides the necessary language to communicate the specificities of a model, how it differs from "sibling" and "parent" models in the family; at the same time as the generality of the features it shares with the general framework. The models, like theories, represent only one possible representation of reality. They show plausible relationships, but leave open the possibility that other models within the framework fit the data equally well (Chamberlin, 1965).

11.2 Critical reflection

Although Internet mediated research can broaden a sampling frame beyond students, it raises a new set of problems. The HAPPLE study contained a large proportion of students, and was substantially different from the general population even among non-students. Coupled with the missing data on demographic variables, the data was limited in its capacity to model important relationships in the models. Gender is a substantive variable in health research, but no associations were observed in the model containing Conscientiousness and health outcomes. This may reflect the small sample size available for those participants who reported their gender. Missing data was also problematic for the SES latent variable, which was not possible to fit. In retrospect, requesting postal sector (e.g. EH8 9XX) rather than full postal codes (e.g. EH8 9JZ) may have increased the proportion of available data for Carstairs scores. The choice of Internet data collection was governed by constraints on resources, but produces a sample that was difficult to describe in terms of a sample frame from a larger population. In future, demographic variables could be made compulsory and the sampling

frame could be defined more clearly, perhaps by targeting the questionnaire at mailing lists or particular web sites. It is regrettable that measures of SES were inadequate in the HAPPLE study, and unavailable in the cortisol study. In future, online data collection could be combined with saliva samples in populations that encompass a population with more variance on SES. This would require additional resources, but cortisol is a candidate mechanism linking traits to health. SES is a candidate moderator of this pathway. The fact that gender did not emerge as a significant pathway, in the HAPPLE or cortisol studies, is a limitation of these data sets. Gender is known to be important variable in personality-health models. Furthermore, the lifestyle factor in the HALS data only emerged for males. This suggests further work is needed to clarify the role of gender in latent variable models that contain traits and health.

11.2.1 Anticipated objections

SES. The research had two key limitations in relation to SES. The HAPPLE study, although it was designed to maximize sample size and minimize missing data, did not produce the type of data expected. Few participants reported their postal code, and far greater proportion of economically inactive people took part than expected. However, there was variance on education, occupation and deprivation scores. Despite this, the intended latent variable model, in which SES was defined by three indicators, did not converge. Instead, years of education was used as a proxy for SES in the economically active sample. This was permissible, and produced interesting results, but unfortunate in that it was not the original aim of the study. In the cortisol study, all of the participants were students. Therefore, it was not possible to model SES.

Causality. First, none of the models I can presented can imply causality. They are based on cross-sectional data, with the exception of cortisol which was measured across two days. The diurnal profile is measured across the day, but once operationalized, it is a cross-sectional variable. It is not possible,

from these findings, to discern whether traits precede cortisol, or if cortisol influences personality traits. Latent growth modelling too, does "not presume a causal influence but merely an association, similar to that of a correlation, but within a longitudinal perspective" (Hampson, Goldberg, et al., 2007, p. 295). This objection is not specific to the findings, but to latent variable models in general. Growth curve modelling is still important because it reduces measurement errors, and can reveal associations between traits and health criteria.

Casual primacy. Second, the position of the variables in the models could be changed. Health outcomes could operate as upstream variables, influencing health behaviours and personality traits. These associations were not tested in this thesis because the models were informed by existing theories. Traits have causal primacy in personality-health research (Matthews, Deary, & Whiteman, 2003). This applies to models containing health behaviours, and stress moderation models (T. Smith, 2006). It is sensible to treat traits as independent variables, because they have roots in childhood temperament (Cloninger, Svrakic, & Przybeck, 1993; McCrae, Costa, Hrebickova, et al., 2000), stability increases throughout adolescence (McCrae et al., 2002), and becomes, fairly linearly, more stable over time into adulthood (Hampson & Goldberg, 2006; Hampson, Andrews, Barckley, Lichtenstein, & Lee, 2006; Caspi, Roberts, & Shiner, 2005), particularly after 30 years of age (Terracciano, Costa, & McCrae, 2006). The "set like plaster" (Terracciano et al., 2006; Srivastava, John, Gosling, & Potter, 2003) conceptualization probably over-simplifies a complex relationship, because trait scores were not fully stable in any of these studies, but traits are stable enough to treat them as predictors of external criteria (Costa & McCrae, 1992a). Education is appropriate as a mediator because traits and other demographic variables influence educational attainment (Digman, 1989; Hampson, Goldberg, et al., 2007; Shiner, Masten, & Roberts, 2003). Health outcomes are valuable to individuals and to society, and inequalities in health outcomes are key public health issues (World

Health Organization, 2002). The position of the variables was therefore informed by prior research, existing theory, and public health relevance.

Replication. Third, the models presenting require replication in other samples. Taken alone, a model is a hypothesis (Cliff, 1983), and I have made no claims that these models cannot be falsified, improved, or they do not need replicating. As Muthén and Muthén (1998–2007) argue, "The purpose of modeling data is to describe the structure of data in a simple way so that it is understandable and interpretable. Essentially, the modeling of data amounts to specifying a set of relationships between variables" (p.1-2). The relationship are unique *to the data*, until replicated on other data, but they have offered important avenues for future research.

Other variables. Fourth, there are many variables which intervene between the variables included in the models. For example, coping strategies might mediate the relationship between Neuroticism and cortisol (Houtman & Bakker, 1991; Iglesias et al., 2005). The relationship between cortisol and general health perception might be mediated by fatigue (Bower et al., 2005; Demitrack et al., 1991; Sluiter, Frings-Dresen, & Van Der Beek, 2000). The relationship between the big four health behaviours and mortality is likely to do the development of cardiovascular risk factors (atherosclerosis and high blood pressure, Bartley et al., 2000; Boniface & Tefft, 1997). When mechanisms linking variables are well understood, and they are biologically plausible, model validity could be enhanced by adding additional variables. However, there is always a compromise between parsimony and explanatory power (Crawley, 2007), in any model. I turn now to the anticipated criticisms that are specific to the HAPPLE study:

Income. Fifth, income was not measured in the HAPPLE study. Future studies could measure income alongside the other three indicators. This decision was made in the HAPPLE study for two reasons. First, only three indicators are required in order to identify a latent variable (Loehlin, 2004). Second, it was anticipated that items about income would increase the refusal rate: "income information is considered to be sensitive and is not measured

in many studies" (Braveman et al., 2005, p. 2881). This might have led to missing data, or higher levels of drop-out. Unfortunately, missing data was problematic for job description and for postcodes. Therefore, this concern may not have been founded. However, given the high refusal rate for postcodes, there are strong reasons to suspect that a similarly high refusal rate would have occurred for questions about income.

Missing data. Sixth, missing data on covariates. A weakness of the HAPPLE study was that demographic variables were not measured on the first page at the beginning of the study. When it became apparent that large numbers of respondents were not reaching the demographics page, this was changed. Some missing data was modelled or imputed by inferring covariates from second wave responses. This is not detrimental to the study, because the sample size was moderately large when these participants were excluded. The data could be reanalyzed in the future if methods for modelling missing data on covariates become more widely available.

Validity scales. Seventh, the lack of a socially desirable responding correction, or so-called validity scale. This approach is not recommended (Costa & McCrae, 1992a; Piedmont et al., 2000) because trait items are not a perfect way to measure traits, "and validity scales will not make them so" (Piedmont et al., 2000, p. 591). Items designed to control for socially desirable responding are self-reports, and are not the best method for correcting other self-reports. Instead, the emphasis should be on improving validity and ensuring that participants understand the instructions and purpose of the inventory (Costa & McCrae, 1992a), and on clinical judgement (Costa & McCrae, 1992b). The concern is perhaps more applicable in occupational psychology, where there is a clear evidence, and plausible motivation factors, that participants fake their responses to NEO items (Winkelspecht, Lewis, & Thomas, 2006). Arguably, where research is voluntary, is for research purposes, and has no obvious negative consequences, there is no motivation for participants to distort their answers (Roberts et al., 2005).

Representativeness. Eighth, limitations of the HAPPLE study more generally. For example, the choice of outcome measure (SF-36) could be challenged, as could the reliance on self-reports which can inflate correlations due to shared method variance. The criticism that HAPPLE is not representative of the general population is the most difficult to address.

There are two responses to the criticism that the HAPPLE study is not representative of the general population. First, the health outcomes did not differ significantly from the SF-36 UK norms. Therefore, it is the mean and variance of the predictor variables that raise concern. The fact that the demographic variables have different mean and variance from those in the population, could attenuate the associations found. As Deary (2000, p. 112) explains, narrow standard deviations will attenuate correlations from the true population correlation. High mean scores can result in ceiling effects, which will also reduce correlations. Second, I argue that it is better to measure a demographic variable with restricted range, than not to measure it at all. Controlling for demographic variables, even if imperfectly, can only take place if they are measured. By collecting data on a wide range of demographic variables, I have gone further than many researchers would. Most (70%) social/personality psychology research is conducted on student populations (Kimmel, 1996; Sieber & Saks, 1989). Students have high means and narrow standard deviations on many variables that influence health; such as mental ability (Deary, 2000), SES background and education. Even among students, there are personality differences in those who take part and when they take part in psychological research, such as Conscientiousness (Aviv, Zelenski, Rallo, & Larsen, 2002; Stevens & Ash, 2001).

Representativeness of the cortisol sample. The cortisol study was designed in order to minimize variance on SES, reducing the number of demographic variables in the data. This helps control for confounding factors, and is consistent with many other cortisol studies. However, Pruessner et al. (1997) acknowledged that it would be desirable to expand this focus in the future: "A final call for cautious interpretation of the data refers to the small sample size and the highly selected

group examined. Only healthy, nonsmoking, male students with a narrow age range were studied" (p. 623).

11.2.2 Unanswered questions

I have explained that I did the research because there was a clear need to explore candidate psychobiological mechanism (e.g. cortisol) and social mechanisms (e.g. SES) in personality-health models. These key variables were studied separately, but there remains a need to consider both in the same model. This leads to a future research question. What is the relationship between cortisol and socio-economic status, as mediators of trait-health associations? I propose a future study, which could help answer that question.

11.2.3 Relation to further study

A limitation of the HAPPLE and cortisol studies is that they do not allow the effect of SES on any of the relationships observed, to be tested. As noted by Kristenson et al. (2004, p. 1515), "There are few population-based human studies examining the role of SES on cortisol levels". SES is variously "positioned" in models, and the researcher's choice seems to predominate. Socio-economic variables can be associated in at least four ways (Singh-Manoux, 2005). Below, I propose four ways in which traits, SES and cortisol might be related. Each of these could be tested in future research. In the descriptions that follow, the subsequent association between cortisol and health outcomes is not featured, but can be assumed. There is already a large literature showing that HPA axis dysregulation is associated with poor health outcomes (see E. O. Johnson et al., 1992; McEwan, 1998):

As mediator. In this model, traits influence SES, and these variables raise cortisol in turn. This is conceptually similar to the model proposed by Hampson, Goldberg, et al. (2007), where traits influence educational attainment, which both influence health in turn.

As moderator. In this model, the relationship between traits and cortisol changes, depending on the level of SES. This model is compatible with findings from several studies which show the cortisol varies according to SES. However, Kunz-Ebrecht et al. (2004) did not find this relationship. Instead, they found that higher perceived demands raised cortisol (in high SES civil servants); but in low SES servants, less perceived control raised cortisol. It was therefore inappropriate to use the construct "job strain" (high demand and low control). Furthermore, men's cortisol was raised by low control, while women's cortisol was raised by high demand (Kunz-Ebrecht et al., 2004). It would therefore be important to include stress appraisal measures, and gender, alongside measures of SES.

As confounder. In this model, SES is related to both to traits and to cortisol, for other reasons. The result is that traits and cortisol appear to be related, but they do not form a causal chain. For example, low SES might raise Neuroticism (Monden & Kraaykamp, 2006). From a large literature, we already know that cortisol is higher in low SES groups (S. Cohen, Doyle, & Baum, 2006; C. E. Wright & Steptoe, 2005; Dowd & Goldman, 2006; Kristenson et al., 2004; Kunz-Ebrecht et al., 2004). Therefore, this might produce a spurious association between Neuroticism and cortisol.

As antecedent. In this model, SES is an antecedent of cortisol and Neuroticism: SES becomes an upstream variable. This model is conceptually similar to Dowd and Goldman (2006). It is known that cortisol output is higher in low SES groups (see above), and SES could clearly precede cortisol output. There is also evidence that low education is cross-sectionally associated with Neuroticism, although this might be explained by cigarette smoking (Monden & Kraaykamp, 2006).

The four associations described above could only be compared in a longitudinal study, ideally in children or adolescents, followed into early adulthood and beyond. One important avenue for future research is therefore to measure diurnal cortisol longitudinally in early adulthood, when education has been completed, SES established and health behaviours begin to exert their influence on SF-36

scores over time. The most sensible strategy for conducting this study would be to find an existing data set where many of the key variables are already measured: traits, health behaviours, SES, perceived demands and control at work, health outcomes. This would be a complex undertaking, but achievable, if an existing longitudinal study could be found. It may be possible to discern a point at which cortisol output begins to result in chronic over-activation of the HPA axis, leading to fatigue, tendencies to self-medicate with unhealthy behaviours, and poor health outcomes. The saliva sampling protocol could be administered at relatively low cost, and participants could provide the samples in everyday naturalistic settings. There is very little research on cortisol profiles over months, years and across the lifespan. In summary, I argue that SES should become a substantive variable in future personality-health research (health behaviours and cortisol). There is substantial theoretical support for such a model.

11.2.4 Relation to practice

The results also have practical relevant to personality-health models in general, with specific recommendations possible for future models that contain traits, health behaviours and health outcomes. The results of my thesis suggest that:

Facets. It is clear that the NEO-PI-R rather than the NEO-FFI should be used in cortisol-personality research, because it provides facet level detail.

SES. Sociological variables (e.g. socio-economic status), and a wide range of demographic variables, can be measured alongside traits and health behaviours and health outcomes. Although it was not possible using the HAPPLE data set, a latent variable can be used to model SES. It is important to ask about economic inactivity and treat these as a separate group (students, retired, carers, permanently sick or disabled) because the traditional indicators of SES (education, occupation, social deprivation, income) may not apply to these people. In addition, strategies should be sought than minimize missing data on these indicators.

Retest. Cortisol should indeed be measured over (at least) two days, to improve reliably and reduce missing data. It should be distinguished from the waking response and the reactivity profile (see also methodological criteria in chapter nine).

SF-36. The SF-36 is an efficient way to obtain mental and physical health outcomes, and the GH scale shares variance with cortisol output, showing that both measure "something more, and something less" (Maddox & Douglass, 1973, p. 92) than each other in relation to health. Combining objective and subjective measures of health is a useful strategy. when investigating personality and health relationships.

Combination. Where resources allow, cortisol and SES should be measured and modelled at the same time. Research has shown that the impact of cortisol on health is not the same at different levels of SES. This is referred to as an interaction, or in epidemiology, effect modification.

Latent classes. Latent class and latent profile analysis should be used to check for the existence of two or more groups, in health outcome or health behaviour data.

Survival. Latent variables (e.g. of health behaviours) can be combined with survival analysis (Larsen, 2005). The may appeal to the "consumers" of trait research (see below) in the health sciences, and it demonstrates "ultimate" validity (O'Toole & Stankov, 1992).

Framework. Results can be reported in a common modelling "tongue" (c.f. Saucier & Goldberg, 2002), the generalized latent variable modelling framework (Muthén, 2002).

11.3 So what? Application of personality-health research

The most promising benefits of future research in this area are twofold. First, it has the potential link personality-health research with epidemiology, psychoneuroendocrinology and ultimately, with consumers who can implement interventions. Krueger et al. (2000) had warned trait researchers that they must think epidemiologically, and are worth quoting at length: "Personality researchers can

build on the accomplishments of the last 2 decades by sharing their accomplishments with the broader social science community. To move epidemiological personology forward, personality psychologists need to make contact with researchers in sociology, epidemiology, and public health, where we predict that structural models of personality will prove their 'added value' to both basic science and social policy." (p. 994). These health sciences are the "consumers" (R. Hogan, 2005) of personality-health research, but they have met the field with scepticism.

11.3.1 Pioneers

Choice of model is determined, at least partly, by intended audience (Tabachnick & Fidell, 2000). Who are the intended for personality-health research? Deary and Hettema (1993) have previously argued that trait researchers have devoted too much time to internal, factorial validity of tests, and not enough to external, criterion validity. Personality-health researchers have perhaps swung the pendulum in the opposite direction, with a great deal of research demonstrating predictive validity for traits on health. Perhaps even to the extent that health has provided traits with an "anchor" (Bowers, 1987, p. 344) to validate their existence. Relatively little discussion takes place about how personality-health research might applied. There are two notable exceptions. One is to envisage traits being measured in medical settings (Matthews, Deary, & Whiteman, 2003). Another is to target or tailor health promotion interventions to individual differences (e.g. Donohew et al., 2000; Noar, Benac, & Harris, 2007). In general however, trait researchers have tended not to discuss applications of personality-health research, because the most important goal for them is to first show that traits do influence health. Demonstrating a true (rather than spurious) causal link between traits and health is needed before any interventions based on traits can be developed. The "groundwork" of personality-health research is concerned with demonstrating associations between traits and health, and this is where trait researchers focus their efforts. This process takes time, because no one study can include all

of the possible variables that might be involved. However, not making *any* reference to potential applications might weaken the potential for trait researchers to "sell" their work to "consumers". R. Hogan (2005) recently used the terms "pioneer" and "consumer" to refer to the distinction between occupational trait researchers and those who might implement trait research in occupational settings. These terms are useful when thinking about traits and health.

11.3.2 Consumers

Consumers of personality-health research might exist within the discipline of psychology, such as clinical-, health- or neuro-psychologists. They will also exist outside the discipline, including epidemiology (Krueger et al., 2000), health promotion and public health. The consumers will look to trait research not only because of its empirical findings, but in terms of how it can help them design evidence based interventions. If there is no substantive content relating to interventions, then the potential implications are few. This might impact trait research in the long term, because there is not a one directional relationship between pioneers and consumers. Consumers are needed in order to improve the groundwork. It is the consumers who provide access to study populations, resources, feedback, and working relationships with pioneers. Even the groundwork of trait research requires collaboration with these disciplines. Therefore, attending to the potential applications of trait researchers might accelerate the groundwork, by encouraging more consumers to take part. In the section below, I introduce a typology of potential applications, which help to categorize the types of interventions that personality-health researchers might consider. Before that, a brief discussion of the groundwork is required.

11.3.3 The groundwork

A brief discussion of the groundwork, where associations between traits and health are reported, is required (see also chapters two, three and four of this thesis). In health sciences, causality is assessed using the criteria for judgement

(see chapter two). As this thesis has shown, there are several plausible models of how personality and health could be related. As I described in chapter two, there are challenges to measurement, which has slowed the pace at which such associations can be understood. Understanding of personality-health associations is hampered when personality measures are incomparable across studies. The adoption of the Five Factor Model has accelerated the pace of research, and the inclusion of facets in its hierarchy has been useful because components of traits and be compared across studies (Jonassaint et al., 2007; T. W. Smith & Williams, 1992; Weiss & Costa, 2005). For example, cardiovascular disease has been linked to hostility, which is a component of Agreeableness (Miller, Smith, Turner, Guijarro, & Hallet, 1996), and survival (mortality) to social dependability, which is a component of Conscientiousness, also associated with mortality (Friedman et al., 1993, 1995; L. R. Martin, Friedman, & Schwartz, 2007; O'Cleirigh, Ironson, Weiss, & Costa, 2007; Shipley, Weiss, Der, Taylor, & Deary, 2007). The pace of personality-health research has increased since the Big Five allowed researchers to be consistent in their measures of traits (Friedman et al., 1993, 1995; Friedman, 2000; L. R. Martin et al., 2007)

Skepticism

Consumers have met the groundwork with scepticism. This scepticism also occurs in response to findings in from psychosomatic medicine, another discipline concerned with the relationship between psychological traits and health. The responses included arguments that the effect size (the strength of the association) "is so low it has as yet no practical meaning for prevention and prediction purposes" in medical settings (Myrtek, 2001, p. 245). This response is moot — the effect sizes are comparable to other risk factors for illness (Bogg & Roberts, 2004), and are large in several cases (Du et al., 2002; Piedmont, 2001). Another argument is that trait-health associations have "too much face validity" (Stansfeld, 2002b, p. 113), meaning that researchers are too quick to accept personality as a predictor of health outcomes, without the supporting evidence base. A

third argument is that trait measures "were not designed with specific knowledge of brain/behavior relationships in mind, and thus have had little direct applicability". This comments came from a neuropsychologist, who developed a trait inventory with neurological processes in mind (Nelson, Drebing, Satz, & Uchiyama, 1998, p. 550).

11.4 Moving forward: The Five T's

The Five T's classify different kinds of personality-health applications. They emerged after I reflected on the groundwork, and on the possibility of engaging with different audiences for personality-health models. I noticed that only a small number of personality-health models have reached the stage of application. The purpose of this section is to illustrate the ways in which personality-health models could reach the stage of application, and at the same time, engage the consumers so that trait measures become more widely accepted.

11.4.1 Targeting

The personality traits mostly strongly related to health behaviours (e.g. Conscientiousness) could be targeted in health promotion campaigns. Targeted campaigns take into account the characteristics of a group of people, or a defined population. It is known that "different types of individuals may attend to, comprehend, accept, and retain different types of messages" (Caspi et al., 1997, p. 1061). Traits influence all four of these steps in the processing of health promotion messages. Sensation seeking (SS, corresponding approximately to high Extraversion and low Conscientiousness in the Big Five) is particularly important, because it is known to influence risky health behaviours and interact with communication media. High sensation seekers (HSS) require novelty, intensity, sound, drama, stimulation, suspense, fast pace, emotionality, complexity, ambiguity, unconventionality and movement. If a televised health promotion campaign lacks these features, HSS will consider it boring and ignore it. Conversely,

low sensation seekers (LSS) prefer familiar and less sensational delivery of information. The SS targeting campaign SENTAR (Palmgreen, Donohew, Lorch, Hoyle, & Stephenson, 2001) showed that HSS adolescents reduced their cannabis use after watching targeted, televised advertisements. Furthermore, 72% of calls to the supporting telephone hot line were from HSS not LSS. The principles developed by the researchers might fruitfully be used with other traits: (1) pick a trait to target; (2) conduct focus group research with high and low scorers; (3) design messages appropriate to high and low scorers; (4) place messages in contexts appropriate for high and low scorers. A weakness of targeting personality traits is that any group watching or reading the message will contain a full range of individual differences on the trait of interest. A researcher can define a *group* of HSS, but without individual assessments it is difficult to find the individuals. It is often impractical to create more than two or three different versions of a campaign, so studies to date have dichotomized traits into *high* and *low*. However, recent developments in computer technology have allowed researchers to take more variables into account, including continuous ones, which could allow messages to be "tailored" to the individual, rather than targeting groups (Noar et al., 2007).

11.4.2 Tailoring

Tailored and targeted interventions are not the same thing, although these terms have been used interchangeably by some health promoters (Kreuter & Skinner, 2000). Tailoring interventions are modified based on the assessment of an individual, "whereas targeted messages are developed to be effective with an entire segment of the population. Tailored messages, however, do require individualized assessments of members of the population to develop such communications" (Noar et al., 2007, p. 674). There are many psychological measures which can be used in tailored health promotion. These include personality, mental ability, reading ability and cognitive style. Tailored interventions could take many forms: printed materials (leaflets, booklets, letters), videotapes, audiotapes, web sites, kiosks, CD-ROMs or other interactive multimedia programs (Kreuter &

Skinner, 2000). It is perhaps surprising that no research, to my knowledge, has explored the potential utility for tailoring health interventions to personality traits. Personality assessment is an individualized assessment. The theory behind tailoring is that an individual's engagement with a health message will depend on its personal relevance to them, and that if engagement is higher then motivation to act on the message will also be higher. For example, a letter from a clinic could be modified for someone who scores low on Conscientiousness (e.g., "You may want to set yourself a reminder about your appointment next week"). The theory would predict that the reader will be more motivated to change their behaviour because the letter is tailored to their tendency for disorganization. There is evidence that tailored messages are more successful than non-tailored or targeted messages at changing health behaviours. For example, Skinner, Strecher, and Hospers (1994) used computer technology to assemble individualized letters tailored according to where women were in the "stage of readiness" for mammography: pre-contemplators, contemplators, actors or maintainers. A recent systematic review and meta analysis showed that printed tailored interventions were more effective than targeted ones (Noar et al., 2007). This was particularly true for studies which tailored the intervention to health behaviour, demographic variables, and used a health behaviour change theory (e.g. stage of change). The authors did not find any studies that tailored according to personality traits, highlighting the need for tailoring research in relation to personality. However several studies in the review included constructs that overlap considerably (e.g. self-efficacy, which is a facet of Conscientiousness). Tailoring is the least researched application of personality-health research but it is potentially the most fruitful. If health promotion can tailor to personality traits more effectively than targeting personality traits, then this is a strong incentive to study the utility of tailoring to traits. This would provide consumers with evidence that traits can be measured for "prevention and prediction purposes" (Myrtek, 2001, p. 245).

11.4.3 Training

Although personality traits are largely stable, there is some preliminary evidence that basic (Big Five) traits might be modifiable by psychological therapies such as psychotherapy or cognitive behavioural therapy. For example, a six week program of psychotherapy, totalling 180 hours of "training", was designed to treat drug addiction. The aim was to intervene on basic traits, improving vocational skills (Conscientiousness), coping ability (Neuroticism), spiritual development (Openness to Experience), and social skills (Extraversion and Agreeableness). Surprisingly, changes were observed on all the traits, particularly Neuroticism which was lowered by one half a standard deviation – considered a large effect size (Piedmont, 2001). Traits, then, may not be stable if there is input from clinicians aimed at changing them. The investigators did note that these changes might have occurred due to reductions in psychological distress. Another recent trial used training to modify personality traits. The Penn Resiliency program (see Freres et al., 2002; J. Gillham & Reivich, 2004; "Preventing Depression Among Early Adolescents in the Primary Care Setting: A Randomized Controlled Study of the Penn Resiliency Program", 2006) was designed to cultivate optimistic traits in children. Optimism corresponds to Cheerfulness / Positive Emotions in the Big Five. The program used cognitive behavioural therapy (CBT) to discourage children from interpreting events as internal, stable and global (e.g. my fault, always my fault, applies to every situation). This explanatory style is a risk factor for depression and anxiety, particularly when coupled with catastrophic beliefs about the future and the belief that small problems are insurmountable. Children were encouraged to identify these depressive styles of thinking and were taught techniques such as "putting it into perspective" and "one step at a time". The program provided encouraging results. Two year after the program ended, children who participated in it showed fewer depressive symptoms than controls not in the program. Those in the program were half as likely to report symptoms in the moderate to severe range, with the prevention effect growing stronger over time. A recent follow-up showed the training reduced depression, anxiety,

and adjustment disorders (when combined) among high-symptom participants. Both of these studies show that it is possible to cultivate traits, traits which may protect against illness. Training might be developed to modify Big Five traits in the future. However, the possibility that traits can be changed raises ethical questions about autonomy and authority – who should decide? This question is relevant to prescribed medications that may change traits, and is therefore discussed below.

11.4.4 Treatment

Arguably, many people temporarily "treat" or modify their traits using drugs such as caffeine and alcohol. However, the "new neurotechnologies" (P. Martin & Ashcroft, 2005) have the potential to make longer term changes to personality traits, in contrast to temporary changes in mood or behaviour, which are states. Widely prescribed psychotropic medications such as Methylphenidate (MPH, e.g. Ritalin) and selective serotonin reuptake inhibitors (SSRIs, such as Prozac and Seroxat) change molecular events that underlie cognition, emotions, identity — and perhaps personality. Commentators have already raised the possibility of "other possibilities" for their use, including personality trait change. This is a strong claim. Farah et al. (2004) noted that changing brains actually changes people. What are the individual, ethical and societal implications of medications that can change basic traits? Drugs prescribed for psychological problems can fundamentally change the way people feel about and represent themselves, often in unintended ways (P. Martin & Ashcroft, 2005). At least five studies have demonstrated changes in Neuroticism or Extraversion in response to SSRIs. For example, SSRIs have been shown to lower scores on Neuroticism (Andrews, Parker, & Barrett, 1998; Bagby, Levitan, Kennedy, Levitt, & Joffe, 1999; De Fruyt, Van Leeuwen, Bagby, Rolland, & Rouillon, 2006; Du, Bakish, Ravindran, & Hrdina, 2002; Knutson et al., 1998). This is only a preliminary association, and note that a baseline measures of personality is often not available. Lowered Neuroticism scores could reflect removal of depressive symptomology (the observed score might represent a return to normal). The effect size

is about one half a standard deviation. Interestingly, this is about the same effect as that observed during psychotherapy training (above). Some commentators have argued that cognitive enhancements undermine the value and dignity of hard work and learned self-insight. D. W. Brock (1998) argued that "altering a fundamental character trait or psychological feature by a 'quick fix' of 'popping a pill' seems to some people too easy and less admirable than changing that same trait or feature through hard-earned insight psychotherapy" (p. 58).

The distinction between enhancement and treatment. Enhancements differ from treatments, in that they improve human performance, appearance or behaviour without medical need. For example, if lead poisoning lowered a child's IQ from 100 to 80, medication that raised it back to 100 would be considered a treatment. Without poisoning, medication that raised IQ from 80 to 100 would be considered an enhancement (Schwartz, 2005, p. 18). Many aspects of psychological functioning are potential targets for enhancement: memory, mental ability, mood, appetite, libido, sleep and personality traits (Farah, 2005). The use of cognitive enhancements is not hypothetical. US soldiers in Iraq take Modafinil to reduce the need for sleep. Ritalin is used in US colleges to improve concentration for exams (Babcock & Byrne, 2000). Sales of Ritalin are widely disproportionate to the prevalence of ADHD, suggesting that children and/or parents are using Ritalin as an enhancement (Wolpe, 2002). Treatment is clearly the most controversial application of personality-health research. As the use of "cosmetic psychopharmacology" (Turner & Sahakian, 2006) increases, dialogue between trait researchers and bioethicists will be required.

11.4.5 Transformation

Personality change (transformation) is a symptom of many illnesses, such as multiple sclerosis, dementia, Alzheimers, Parkinsons or stroke. Personality trait transformation is different from the other four T's in that it is something observed, rather than something applied. Trait change has important implications for patients and their families, and is a key clinical problem. Changes can cause

distress, and in some cases it is more distress for caregivers than for patients. Patients may lack insight into their personality change, which may result from what J. Stone et al. (2004) called "emotional agnosia". Personality change has received far less attention for neuropsychologists than emotional disorders and cognitive impairment. A key challenge for this application of personality research is to obtain a premorbid measure of personality traits, since these are not usually measured in medical settings. As a result, little is known about how illnesses can change basic traits. The instruments that are used in clinical and medical settings tend be those designed as measures of psychopathology (e.g. Anxiety and depression) or as measures of neurological disorders or brain injury (Nelson et al., 1998). For example, the Neuropsychology Behaviour and Affect Profile (NBAP) provides a premorbid and current description of personality. It has been used successfully with stroke, dementia and closed head injury patients. The scales include Indifference (to one's injury), Inappropriateness, Depression, Mania, Pragnosis (social/pragmatic communication style). However, neuro-specialists may be reluctant to measure the Big Five, since it was not designed with brain structures in mind. Nonetheless, trait researchers can collaborate with specialists observing personality change after illness, designing personality measures that are both descriptive of basic traits and are informed by illness aetiology. Measures that assess adjustment or coping after illness will also be of high value.

11.4.6 Conclusion

The Five T's are a framework which might help orient the models of personality-health research better to the consumers. Personality-health researchers often have trouble persuading their colleagues in psychology (in health-, clinical- and neuro-psychology) to "get on board the train", let alone those working in other disciplines. Arguably, the applied utility of personality for health improvement initiatives has not been made clear, and this is part of the reason. The emphasis

on showing that personality is related to health, the groundwork of personality-health research, has distracted from a discussion about the possible applied end-points or "destinations" of this work. Like the circle line on the London Underground, basic empirical work is always ongoing and does not necessarily have a final destination. There are at least five possible destinations (targeting, tailoring, training, treatment, transformation). In my view, by pointing out possible end points for this research, it more likely that "increased collaboration between personality psychologists and researchers in fields such as public health, epidemiology, and sociology" (Krueger et al., 2000, p. 968) will actually happen. This dialogue will not only encourage others to get on board the train, but will improve the groundwork. Personality researchers need to engage with passengers from other disciplines, to move beyond the groundwork that traits influence health, to what can or should be done about it.

11.4.7 Using the Five T's to select a modelling strategy

B. P. Chapman et al. (2007) argue that categorical latent variables "may have untapped screening and heuristic value in understanding general physical health and physical and psychosocial functioning" (p. 925). When evidence is available for types, they argue, brief measures "might feasibly be used in busy medical settings, pending typological validation in general adult samples" (p. 925). They also argue that "most people who are not psychologists tend to think of personality in typological terms. Costa, Herbst, McCrae, Samuels, and Ozer (2002) have argued that five factor model dimensions are better than types in terms of predictive validity "but in the day-to-day work of the clinician it is possible that, because of evolved cognition, thinking in terms of types may come more naturally and lead to better understanding of clients" (p. S84). This may be particularly true among the health scientists, public health specialists, and health care providers who are trained to "find patterns" and think categorically. If the consumers of personality-health research are more accepting of types, then model audience will become an important area for the field in the near future. Similarly, continuous latent variables, such as "lifestyle" (chapter eight) might help

persuade epidemiologists to consider the criterion validity of latent variables in explaining variance in health, over and above additive or unidimensional scales (Belloc & Breslow, 1972; B. M. Brock et al., 1988; Wingard et al., 1982; Dean & Salem, 1998), scales which have good content validity but are lengthy (Christo et al., 2003) or describe substance use poorly (R. R. Vickers et al., 1990; Vickers & Hervig, 1984).

11.4.8 Improving communication of results to consumers

How can models be communicated better to consumers? As Vollrath and Torgersen (2002) argued, there is evidence that clinicians and the general public use notions of types, not personality traits, in their thinking about personality. Costa et al. (2002) previously argues that types "would make communication with clinicians easier" (p. S84). Vollrath and Torgersen (2002) adds that types are "graphic heuristics", helping people to understand the relations between traits and health behaviours. By "grouping similar people together into prototypic categories" (B. P. Chapman et al., 2007, p. 912), using a combination of traits, health behaviours, and health outcomes, this may help communicate to other disciplines. There is a clear future role for multidimensional scaling, visual approaches to data, and latent class analysis. Beyond communication, it may be easier to target types with health promotion or public health interventions (Palmgreen et al., 2001), than to individually tailor interventions to individuals (Kreuter & Skinner, 2000). From a public health perspective, it is important to prioritize "well known, common, substantial and widespread" risk factors (World Health Organization, 2002, p. 11). Traits and health behaviours meet these criteria. As I eluded to above, this suggests trait researchers should demonstrate that interventions are first *possible*, before their findings will be accepted in the wider scientific and practitioner communities.

11.5 Closing comments

My research took place at a time when personality trait research began to move from a focus on measurement models, to structural models. The literature about the measurement models for traits was once complicated, with a proliferation of jingles and jangles for trait terms. Now that these have been organized into higher order frameworks (usually the big five, although some researchers do not adopt this paradigm), the great debates in the literature have changed. Within personality-health research specifically, debate has recently focused on the existence or size of the association between traits and health (arguably for too long). Typically, this was the size of a correlation, or the size of R-square, the predictive validity of a multiple regression model. Associations are no longer a means to an end, to validate traits (Bowers, 1987), but are studied in their own right. It is now a multitude of models, not measures, that have proliferated the literature. As I have shown in my results and data, the generalized latent variable modelling framework, should go a long way to integrating modelling strategies, and how they are reported across disciplines. There is no perfect model that can communicate the true relationships between traits and health. Models, like theories, should be tested empirically (Gholson & Barker, 1985) and multiple working models can be be allowed to compete (Chamberlin, 1965). A common language that can emphasize similarities between models, rather than differences, will accelerate the testing of these multiple working hypotheses. This research adds an important part to that conversation, by opening up several new avenues for further study of the relationship between traits and health.

Health and Personality Processes: Links Explored (HAPPLE)

Participant Information Sheet

You are being invited to take part in a research study.

Before you decide whether you would like to take part, we would like to explain why the research is being done and what we will ask you to do. Please take the time to read the following information carefully. If you would like more information or if anything is unclear, please do not hesitate to ask (contact information is at the bottom of this form). Take time to decide whether or not you wish to take part.

What is the purpose of this study?

Differences between people seem to affect how healthy their lifestyles are. We are interested in how personality might affect different kinds of health behaviours and if it might be possible to use a questionnaire to provide clues about this. If personality affects health, we might be able to improve the way in which online health information is provided. Until now, most research in this area has used students and we would like to find out more about people across the U.K., using the Internet to reach people.

Why have I been chosen?

We would like any adults in the U.K. who use the Internet, to take part, except those with a long-term illness, health problem or disability, which might limit ability to participate.

Do I have to take part?

No - it is up to you to decide whether or not you want to take part. If you decide to take part, you a still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive anywhere else.

What will happen to me if I take part?

We would like you to fill in two questionnaires, on detailed one and a shorter one two months later. We would e-mail you to tell you when to complete the second questionnaire, with a maximum of two e-mail reminders.

What would I have to do?

Each questionnaire will seek information about your background, personality, health (including sexual health) and symptoms. We realize that some people find health questions quite personal. You do not have to answer any questions you find uncomfortable. The information you provide will help us improve our understanding of how to target health information in the future.

How long will it take?

The amount of time taken may vary, but for most people it should take between 40 and 60 minutes. If you are concerned that staying online for this length of time will be costly, we can post a paper version to you. Our contact details are at the bottom of this page.

What are the possible disadvantages and risks of taking part?

You may find that the questionnaire takes a long time to complete, in which case we recommend that you do this at a time convenient to you. You may also find that some of the questions make you feel uncomfortable, in which case you can simply leave them blank.

What are the possible benefits of taking part?

At the end of the questionnaire, there is information about health including useful telephone numbers and web site addresses. There is also a space to make any comments you might have about the study.

The questionnaire will also provide feedback about your responses to the personality questionnaire, which you might find interesting or useful.

Will my taking part in this study be kept confidential?

All information that is collected during the course of the research will be kept strictly confidential. The information sent to us from the online questionnaire is encrypted using 128bit encryption technology, similar to that used by online banking services.

Are there any more ways to keep my information safe?

Running a firewall and anti-spy software on your PC is a good way to protect your information further.
Zone Alarm is a free personal firewall that blocks dangerous Internet threats, guarding your PC from many of the tactics used by hackers and data thieves. CLICK HERE to download Zone Alarm.
Anonymizer is anti-spy software. Anonymizer 2004 shields your IP address and protects you from online tracking, SPAM harvesting, hackers and snoops. CLICK HERE to download Anonymizer 2004.

I am using a shared PC. What if another user infiltrated my PC and saw my information?

If you used a shared PC we can post a paper version to you. This will include a pre-paid envelope.

What if something goes wrong?

If you experience any technical difficulties with the questionnaire or have any other problems, feel free to contact us and we will do our best to resolve them. We can post a paper version to you. If there are problems with the personality questionnaire feedback, we can also post this to you.

What will happen to the results of the research study?

Once the research study is successfully completed, the overall results will be written up in a research report. If you wish, we can arrange for you to obtain a copy of the findings.

Who is organizing and funding the research?

The data collected in this research project will be used to compile my report for my MSc by Research in Psychology at the University of Edinburgh. I will not be paid or funded for the study, and examiners and my dissertation supervisor will view the report. The study is being funded by the Economic and Social Research Council. These people will not know your identity.

Who has reviewed the study?

This study was reviewed and approved by the NHS Multi Region Ethics Committee for Scotland.

Who do I contact for further information about the study?

If you wish to receive further information about the study, please feel free to contact my supervisor:

Dr. Martha Whiteman
Department of Psychology
The University of Edinburgh
7 George Square
EDINBURGH EH8 9JZ
0131 650 3317
M.Whiteman@ed.ac.uk

Or you can e-mail me:
G.E.Hagger-Johnson@sms.ed.ac.uk

Who do I contact for further information about health?

If you are concerned about your health, you should consult your G. P. in the first instance. Information is also available from NHS Direct Online:

http://www.nhsdirect.nhs.uk/

Thank you

Thank you very much for taking the time to read through this information, and for taking part in the study. Your participation will be greatly appreciated and with your help we can gain more information about how to improve online health information in the U. K.

Gareth Hagger-Johnson

Department of Psychology
The University of Edinburgh
7 George Square
EDINBURGH EH8 9JZ
0131 650 3317
G.E.Hagger-Johnson@sms.ed.ac.uk

How to take part

Please answer these questions and click on the button below

I have a long-term illness, health problem or disability, which limits my daily activities or the work I can do

Yes No

I live in the United Kingdom (U.K.)

No Yes

I would like to participate in this study

Yes No

I have read the Participant Information Sheet

No Yes

I understand that I can leave the study at any time by e-mailing the researcher

Yes No

> I have read the above statements and w ould like to participate in this study

---

| recommend this study to a friend or colleague | ask a question |

Instructions

Please use the rating scale below to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Please read each statement carefully, and then fill in the appropriate bubble.

The first questions are about your health, and should take around 10 minutes.

What is your e-mail address?

We need your e-mail address so we can tell you when to complete the second questionnaire in three months time.

How did you find out about this study?

If web site, please enter web site address
If search engine, please enter name of the search engine

I eat a balanced diet.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I get enough sleep.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I keep emergency numbers near the phone.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I choose my spare time activities to help me relax.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I take chances when crossing the street.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I have a first aid kit in my home.	Very like me	Like me	Unsure	Unlike me	Very unlike	select one

					me	
I destroy old or unused medicines.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I see a doctor for regular checkups.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I pray or meditate.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I avoid getting chilled.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I watch my weight.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
If you do drive a car, do you travel within the speed limit?	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I watch for possible signs of major health problems (e.g., cancer, hypertension, heart disease).	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I exercise to stay healthy.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I cross the street against the stop light.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I avoid high crime areas.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one

I don't take chemical substances which might injure my health (e.g. drugs, stimulants).	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I check the condition of electrical appliances, the car, etc., to avoid accidents.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I stay away from places where I might be exposed to germs.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I fix broken things around the home right away.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I see a dentist for regular checkups.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I limit my intake of tea/coffee/cola.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I limit my intake of sugar.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I limit my intake of saturated fats.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I avoid over-the-counter medicines.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
I take vitamins.	Very like me	Like me	Unsure	Unlike me	Very unlike me	select one
When driving or riding in the seat of a car do	Very like	Like me	Unsure	Unlike	Very	select one

you wear a seat belt?	me			me	unlike me	
I cross busy streets only at a crossing (e.g. zebra, pelican).	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☉ select one
I avoid areas with high pollution.	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☉ select one
I discuss health with friends, neighbours and relatives.	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☉ select one
I gather information on things that affect my health (e.g. by watching television and reading books, newspapers, Internet, or magazine articles).	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☉ select one
I use dental floss regularly.	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☉ select one
Drivers only. I speed while driving.	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☉ select one
I brush my teeth regularly.	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☉ select one
I take health food supplements (e.g. protein additives, wheat germ, bran, lecithin).	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☉ select one
I learn first aid techniques.	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☉ select one
I get vaccinations to prevent illness.	☐ Very like	☐ Like me	☐ Unsure	☐ Unlike	☐ Very	☉ select one

	me			me	unlike me	
I take more chances doing things than the average person.	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☑ select one
Drivers only. I drive after drinking alcohol.	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☑ select one
I engage in activities or hobbies where accidents are possible (e.g. motorcycle riding, skiing, using power tools, sky or skin diving, hand gliding, etc.).	☐ Very like me	☐ Like me	☐ Unsure	☐ Unlike me	☐ Very unlike me	☑ select one
Women only. Do you know how to examine your own breasts for lumps ?	☐ Yes	☐ No				☑ select one
If 'YES', about how many times a year do you examine your breasts for lumps?	☐ Never	☐ 1-2 times per year	☐ 3-10 times per year	☐ More than 10 times		☑ select one
Women only. How long has it been since you had a cervical (Pap) smear test?	☐ I have never had a smear test	☐ Less than one year	☐ 1 - 3 years	☐ More than 3 years		☑ select one
Men only. Do you know how to examine your own testicles for lumps?	☐ Yes	☐ No				☑ select one
If 'YES', about how many times a year do you examine your testicles for lumps?	☐ Never	☐ 1-2 times per year	☐ 3-10 times per year	☐ More than 10 times		☑ select one
Have you gone to a GP, hospital or clinic for a test for an HIV test?	☐ No	☐ Yes, negative result	☐ Yes, positive result			☑ select one

Do you use condoms with your main partner? By 'main partner', we mean someone that you have a long-term sexual relationship with. For example, boyfriend or girlfriend, husband or wife, a serious relationship or significant other.	☐ Always	☐ Usually	☐ Sometimes	☐ Usually not	☐ Never	☉ I do not have a main partner or not applicable
Do you use condoms with casual partners? Examples of what we mean by 'casual partners': one-night stands, sex that is 'no strings attached', sex with someone you don't know, sex without a relationship.	☐ Always	☐ Usually	☐ Sometimes	☐ Usually not	☐ Never	☉ I do not have casual partners or not applicable
How many casual sexual partners have you had in the last six months?						
SMOKING						
Do you smoke?	☐ Yes	☐ No				☉ select one
How many cigarettes do you smoke per day?						
How long have you been a smoker? YEARS						
How long have you been a smoker? MONTHS						
If 'NO: 'have you ever smoked cigarettes for more than one month?	☐ Yes	☐ No				☉ select one
If 'YES': how many cigarettes per day did you used to smoke?						
How long were you a smoker? YEARS						

How long were you a smoker? MONTHS

DRINKING ALCOHOL

Would you describe yourself as

- A non-drinker
- A very occasional drinker (special occasions only)
- An occasional drinker
- A regular drinker

select one

On how many days over the past two weeks did you have a drink?

days

On the days that you did drink, how many units did you have, on average?

units (approx.)

What is a unit?

A half pint of standard strength beer/lager(3.5% ABV)

A small glass of table wine 125ml (8% ABV)

A small pub measure of spirits (25ml)

Would you like to reduce the amount that you drink?

- Yes
- No

select one

If you would like to say more about your responses to the questions above, please use this space:

Next page >

Please note. We are unable to prevent occasional malfunctions in the questionnaire. If you experience any difficulties, please e-mail us because we want to know about them. It is likely that your information has not been lost.

The next questions are about your personality.

Please use the rating scale below to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Please read each statement carefully, and then fill in the appropriate bubble.

1.	I worry about things.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
2.	I make friends easily.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
3.	I have a vivid imagination.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
4.	I trust others.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
5.	I complete tasks successfully.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
6.	I get angry easily.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
7.	I love large parties.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
8.	I believe in the importance of art.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
9.	I would never cheat on my taxes.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
10.	I like order.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○

11.	I often feel blue.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
12.	I take charge.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
13.	I experience my emotions intensely.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
14.	I make people feel welcome.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
15.	I try to follow the rules.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
16.	I am easily intimidated.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
17.	I am always busy.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
18.	I prefer variety to routine.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
19.	I am easy to satisfy.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
20.	I go straight for the goal.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
21.	I often eat too much.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
22.	I love excitement.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
23.	I like to solve complex problems.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐

24.	I dislike being the center of attention.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
25.	I get chores done right away.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
26.	I panic easily.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
27.	I radiate joy.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
28.	I tend to vote for liberal political candidates.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
29.	I sympathize with the homeless.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
30.	I avoid mistakes.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
31.	I fear for the worst.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
32.	I warm up quickly to others.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
33.	I enjoy wild flights of fantasy.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
34.	I believe that others have good intentions.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
35.	I excel in what I do.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
36.	I get irritated easily.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐

37.	I talk to a lot of different people at parties.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
38.	I like music.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
39.	I stick to the rules.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
40.	I like to tidy up.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
41.	I dislike myself.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
42.	I try to lead others.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
43.	I feel others' emotions.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
44.	I anticipate the needs of others.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
45.	I keep my promises.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
46.	I am afraid that I will do the wrong thing.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
47.	I am always on the go.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
48.	I like to visit new places.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
49.	I can't stand confrontations.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐

50.	I work hard.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
51.	I don't know why I do some of the things I do.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
52.	I seek adventure.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
53.	I love to read challenging material.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
54.	I dislike talking about myself.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
55.	I am always prepared.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
56.	I become overwhelmed by events.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
57.	I have a lot of fun.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
58.	I believe that there is no absolute right or wrong.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
59.	I feel sympathy for those who are worse off than myself.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
60.	I choose my words with care.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
61.	I am afraid of many things.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
62.	I feel comfortable around people.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○

63.	I love to daydream.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
64.	I trust what people say.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
65.	I handle tasks smoothly.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
66.	I get upset easily.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
67.	I enjoy being part of a group.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
68.	I see beauty in things that others might not notice.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
69.	I use flattery to get ahead.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
70.	I want everything to be "just right."	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
71.	I am often down in the dumps.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
72.	I can talk others into doing things.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
73.	I am passionate about causes.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
74.	I love to help others.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
75.	I pay my bills on time.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐

76.	I find it difficult to approach others.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
77.	I do a lot in my spare time.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
78.	I am interested in many things.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
79.	I hate to seem pushy.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
80.	I turn plans into actions.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
81.	I do things I later regret.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
82.	I love action.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
83.	I have a rich vocabulary.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
84.	I consider myself an average person.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
85.	I start tasks right away.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
86.	I feel that I'm unable to deal with things.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
87.	I express childlike joy.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
88.	I believe that criminals should receive help rather than punishment.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐

89.	I value cooperation over competition.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
90.	I stick to my chosen path.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
91.	I get stressed out easily.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
92.	I act comfortably with others.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
93.	I like to get lost in thought.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
94.	I believe that people are basically moral.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
95.	I am sure of my ground.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
96.	I am often in a bad mood.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
97.	I involve others in what I am doing.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
98.	I love flowers.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
99.	I use others for my own ends.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐
100.	I love order and regularity.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree ☐

101.	I have a low opinion of myself.	Strongly disagree ☐	Disagree ☐	Unsure ☐	Agree ☐	Strongly agree

102.	I seek to influence others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
103.	I enjoy examining myself and my life.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
104.	I am concerned about others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
105.	I tell the truth.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
106.	I am afraid to draw attention to myself.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
107.	I can manage many things at the same time.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
108.	I like to begin new things.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
109.	I have a sharp tongue.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
110.	I plunge into tasks with all my heart.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
111.	I go on binges.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
112.	I enjoy being part of a loud crowd.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
113.	I can handle a lot of information.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
114.	I seldom toot my own horn.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
115.	I get to work at once.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree

116.	I can't make up my mind.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
117.	I laugh my way through life.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
118.	I believe in one true religion.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
119.	I suffer from others' sorrows.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
120.	I jump into things without thinking.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
121.	I get caught up in my problems.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
122.	I cheer people up.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
123.	I indulge in my fantasies.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
124.	I believe in human goodness.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
125.	I come up with good solutions.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
126.	I lose my temper.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
127.	I love surprise parties.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
128.	I enjoy the beauty of nature.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
129.	I know how to get around the rules.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
130.	I do things according to a plan.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree

131.	I have frequent mood swings.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
132.	I take control of things.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
133.	I try to understand myself.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
134.	I have a good word for everyone.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
135.	I listen to my conscience.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
136.	I only feel comfortable with friends.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
137.	I react quickly.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
138.	I prefer to stick with things that I know.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
139.	I contradict others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
140.	I do more than what's expected of me.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
141.	I love to eat.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
142.	I enjoy being reckless.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
143.	I enjoy thinking about things.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
144.	I believe that I am better than others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
145.	I carry out my plans.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree

146.	I get overwhelmed by emotions.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
147.	I love life.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
148.	I tend to vote for conservative political candidates.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
149.	I am not interested in other people's problems.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
150.	I make rash decisions.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
151.	I am not easily bothered by things.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
152.	I am hard to get to know.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
153.	I spend time reflecting on things.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
154.	I think that all will be well.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
155.	I know how to get things done.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
156.	I rarely get irritated.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
157.	I prefer to be alone.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
158.	I do not like art.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
159.	I cheat to get ahead.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
160.	I often forget to put things back in their proper place.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree

161.	I feel desperate.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
162.	I wait for others to lead the way.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
163.	I seldom get emotional.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
164.	I look down on others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
165.	I break rules.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
166.	I stumble over my words.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
167.	I like to take it easy.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
168.	I dislike changes.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
169.	I love a good fight.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
170.	I set high standards for myself and others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
171.	I rarely overindulge.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
172.	I act wild and crazy.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
173.	I am not interested in abstract ideas.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
174.	I think highly of myself.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
175.	I find it difficult to get down to work.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree

176.	I remain calm under pressure.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
177.	I look at the bright side of life.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
178.	I believe that too much tax money goes to support artists.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
179.	I tend to dislike soft-hearted people.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
180.	I like to act on a whim.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
181.	I am relaxed most of the time.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
182.	I often feel uncomfortable around others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
183.	I seldom daydream.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
184.	I distrust people.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
185.	I misjudge situations.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
186.	I seldom get mad.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
187.	I want to be left alone.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
188.	I do not like poetry.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
189.	I put people under pressure.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
190.	I leave a mess in my room.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree

191.	I feel that my life lacks direction.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
192.	I keep in the background.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
193.	I am not easily affected by my emotions.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
194.	I am indifferent to the feelings of others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
195.	I break my promises.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
196.	I am not embarrassed easily.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
197.	I like to take my time.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
198.	I don't like the idea of change.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
199.	I yell at people.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
200.	I demand quality.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
201.	I easily resist temptations.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
202.	I willing to try anything once.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
203.	I avoid philosophical discussions.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
204.	I have a high opinion of myself.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree

205.	I waste my time.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
206.	I can handle complex problems.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
207.	I laugh aloud.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
208.	I believe laws should be strictly enforced.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
209.	I believe in an eye for an eye.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
210.	I rush into things.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
211.	I am not easily disturbed by events.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
212.	I avoid contacts with others.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
213.	I do not have a good imagination.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
214.	I suspect hidden motives in others.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
215.	I don't understand things.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
216.	I am not easily annoyed.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
217.	I don't like crowded events.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○

218.	I do not enjoy going to art museums.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
219.	I pretend to be concerned for others.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
220.	I leave my belongings around.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
221.	I seldom feel blue.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
222.	I have little to say.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
223.	I rarely notice my emotional reactions.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
224.	I make people feel uncomfortable.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
225.	I get others to do my duties.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
226.	I am comfortable in unfamiliar situations.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
227.	I like a leisurely lifestyle.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
228.	I am a creature of habit.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
229.	I insult people.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
230.	I am not highly motivated to succeed.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○

231.	I am able to control my cravings.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
232.	I seek danger.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
233.	I have difficulty understanding abstract ideas.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
234.	I know the answers to many questions.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
235.	I need a push to get started.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
236.	I know how to cope.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
237.	I amuse my friends.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
238.	I believe that we coddle criminals too much.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
239.	I try not to think about the needy.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
240.	I do crazy things.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
241.	I don't worry about things that have already happened.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
242.	I am not really interested in others.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
243.	I seldom get lost in thought.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○

244.	I am wary of others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
245.	I have little to contribute.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
246.	I keep my cool.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
247.	I avoid crowds.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
248.	I do not like concerts.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
249.	I take advantage of others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
250.	I am not bothered by messy people.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
251.	I feel comfortable with myself.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
252.	I don't like to draw attention to myself.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
253.	I experience very few emotional highs and lows.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
254.	I turn my back on others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
255.	I do the opposite of what is asked.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
256.	I am not bothered by difficult social situations.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree

257.	I let things proceed at their own pace.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
258.	I dislike new foods.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
259.	I get back at others.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
260.	I do just enough work to get by.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
261.	I never spend more than I can afford.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
262.	I would never go hang gliding or bungee jumping.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
263.	I am not interested in theoretical discussions.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
264.	I boast about my virtues.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
265.	I have difficulty starting tasks.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
266.	I readily overcome setbacks.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
267.	I am not easily amused.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
268.	I believe that we should be tough on crime.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
269.	I believe people should fend for themselves.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○

270.	I act without thinking.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
271.	I adapt easily to new situations.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
272.	I keep others at a distance.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
273.	I have difficulty imagining things.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
274.	I believe that people are essentially evil.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
275.	I don't see the consequences of things.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
276.	I rarely complain.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
277.	I seek quiet.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
278.	I do not enjoy watching dance performances.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
279.	I obstruct others' plans.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
280.	I am not bothered by disorder.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
281.	I am very pleased with myself.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
282.	I hold back my opinions.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○

283.	I don't understand people who get emotional.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
284.	I take no time for others.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
285.	I misrepresent the facts.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
286.	I am able to stand up for myself.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
287.	I react slowly.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
288.	I am attached to conventional ways.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
289.	I hold a grudge.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
290.	I put little time and effort into my work.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
291.	I never splurge.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
292.	I dislike loud music.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
293.	I avoid difficult reading material.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
294.	I make myself the center of attention.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
295.	I postpone decisions.	Strongly disagree	Disagree	Unsure	Agree	Strongly agree

296.	I am calm even in tense situations.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
297.	I seldom joke around.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
298.	I like to stand during the national anthem.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
299.	I can't stand weak people.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○
300.	I often make last-minute plans.	Strongly disagree ○	Disagree ○	Unsure ◉	Agree ○	Strongly agree ○

If you would like to say more about your responses to the questions above, please use this space:

Next page >

Please note. We are unable to prevent occasional malfunctions in the questionnaire. If you experience any difficulties, please e-mail us because we want to know about them. It is likely that your information has not been lost.

This part of the survey asks for your views about your health, and should take around five minutes.

In general, would you say your health is:

- ☐ Excellent
- ☐ Very good
- ☐ Good
- ☐ Fair
- ☐ Poor
- ☉ select one

Compared to one year ago, how would you rate your health in general now?

- ☐ Much better now than one year ago
- ☐ Somewhat better now than one year ago
- ☐ About the same as one year ago
- ☐ Somewhat worse than one year ago
- ☐ Much worse than one year ago
- ☉ select one

The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? Is so, how much?

Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports

- ☐ Yes, limited a lot
- ☐ Yes, limited a little
- ☐ No, not limited at all
- ☉ select one

Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf

- ☐ Yes, limited a lot
- ☐ Yes, limited a little
- ☐ No, not limited at all
- ☒ select one

Lifting or carrying groceries

- ☐ Yes, limited a lot
- ☐ Yes, limited a little
- ☐ No, not limited at all
- ☒ select one

Climbing **several** flights of stairs

- ☐ Yes, limited a lot
- ☐ Yes, limited a little
- ☐ No, not limited at all
- ☒ select one

Climbing **one** flight of stairs

- ☐ Yes, limited a lot
- ☐ Yes, limited a little
- ☐ No, not limited at all
- ☒ select one

Bending, kneeling or stooping

- ☐ Yes, limited a lot
- ☐ Yes, limited a little
- ☐ No, not limited at all
- ☒ select one

Walking **more than a mile**

- Yes, limited a lot
- Yes, limited a little
- No, not limited at all
- select one

Walking **half a mile**

- Yes, limited a lot
- Yes, limited a little
- No, not limited at all
- select one

Walking one hundred yards

- Yes, limited a lot
- Yes, limited a little
- No, not limited at all
- select one

Bathing or dressing yourself

- Yes, limited a lot
- Yes, limited a little
- No, not limited at all
- select one

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

Cut down on the **amount of time** you spent on work or other activities

- Yes
- No
- select one

Accomplished less than you would like

- Yes
- No
- select one

Were limited in the **kind** of work or other activities

- Yes
- No
- select one

Had **difficulty** performing the work or other activities (for example, it took extra effort)

- Yes
- No
- select one

During the past 4 weeks, have you has any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

Cut down on the **amount of time** you spent on work or other activities

- Yes
- No
- select one

Accomplished less than you would like

- Yes
- No
- select one

Didn't do work or other activities as **carefully** as usual

- Yes
- No

select one

During the past **four weeks**, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?

- Not at all
- Slightly
- Moderately
- Quite a bit
- Extremely

select one

How much **bodily** pain have you had during the past **four weeks**?

- None
- Very mild
- Mild
- Moderate
- Severe
- Very severe

select one

During the **past four weeks**, how much did **pain** interfere with your normal work (including both work outside the home and housework)?

- Not at all
- A little bit
- Moderately
- Quite a bit
- Extremely

select one

These questions are about how you feel and how things have been with you during the past 4 weeks. For each item, please give the one that answer that comes closest to the way you have been feeling.

How much of the time during the **past four weeks** -

Did you feel full of life?

- All the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time
- select one

Have you been a very nervous person?

- All the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time
- select one

Have you felt so down in the dumps that nothing could cheer you up?

- All the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time
- select one

Have you felt calm and peaceful?

- ☐ All the time
- ☐ Most of the time
- ☐ A good bit of the time
- ☐ Some of the time
- ☐ A little of the time
- ☐ None of the time
- ☉ select one

Did you have a lot of energy?

- ☐ All the time
- ☐ Most of the time
- ☐ A good bit of the time
- ☐ Some of the time
- ☐ A little of the time
- ☐ None of the time
- ☉ select one

Have you felt downhearted and low?

- ☐ All the time
- ☐ Most of the time
- ☐ A good bit of the time
- ☐ Some of the time
- ☐ A little of the time
- ☐ None of the time
- ☉ select one

Did you feel worn out?

- ☐ All the time
- ☐ Most of the time
- ☐ A good bit of the time

- Some of the time
- A little of the time
- None of the time
- select one

Have you been a happy person?

- All the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time
- select one

Did you feel tired?

- All the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time
- select one

During the **past 4 weeks**, how much of the time has your **physical health or emotional problems** interfered with your social activities (like visiting with friends, relatives, etc.)?

- All the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time

- None of the time
- select one

How TRUE or FALSE is each of the following statements for you?

I seem to get ill more easily than other people.

- Definitely true
- Mostly true
- Don't know
- Mostly false
- Definitely false
- select one

I am as healthy as anybody I know.

- Definitely true
- Mostly true
- Don't know
- Mostly false
- Definitely false
- select one

I expect my health to get worse.

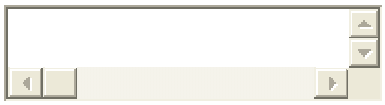
- Definitely true
- Mostly true
- Don't know
- Mostly false
- Definitely false
- select one

My health is excellent.

- Definitely true

- Mostly true
- Don't know
- Mostly false
- Definitely false
- select one

If you would like to say more about your responses to the questions above, please use this space:



next page >

Please note. We are unable to prevent occasional malfunctions in the questionnaire. If you experience any difficulties, please e-mail us because we want to know about them. It is likely that your information has not been lost.

This part of the survey asks about symptoms, and should take around five minutes.

The following questions are about any symptoms you have had in the past week.

Tiredness

- ☐ None
- ☐ Very mild
- ☐ Mild
- ☐ Moderate
- ☐ Severe
- ☐ Very severe
- ☉ select one

Daytime sleepiness

- ☐ None
- ☐ Very mild
- ☐ Mild
- ☐ Moderate
- ☐ Severe
- ☐ Very severe
- ☉ select one

Bodily tension

- ☐ None
- ☐ Very mild
- ☐ Mild
- ☐ Moderate
- ☐ Severe
- ☐ Very severe
- ☉ select one

Weakness/dizziness

- [ ] None
- [ ] Very mild
- [ ] Mild
- [ ] Moderate
- [ ] Severe
- [ ] Very severe
- [x] select one

Loss of appetite

- [ ] None
- [ ] Very mild
- [ ] Mild
- [ ] Moderate
- [ ] Severe
- [ ] Very severe
- [x] select one

Back pain

- [ ] None
- [ ] Very mild
- [ ] Mild
- [ ] Moderate
- [ ] Severe
- [ ] Very severe
- [x] select one

Muscle pain

- [ ] None
- [ ] Very mild
- [ ] Mild

- [ ] Moderate
- [ ] Severe
- [ ] Very severe
- [x] select one

Stomach ache

- [ ] None
- [ ] Very mild
- [ ] Mild
- [ ] Moderate
- [ ] Severe
- [ ] Very severe
- [x] select one

Fever

- [ ] None
- [ ] Very mild
- [ ] Mild
- [ ] Moderate
- [ ] Severe
- [ ] Very severe
- [x] select one

Nausea

- [ ] None
- [ ] Very mild
- [ ] Mild
- [ ] Moderate
- [ ] Severe
- [ ] Very severe

select one

Cold

- None
- Very mild
- Mild
- Moderate
- Severe
- Very severe

select one

Allergic complaints (allergy)

- None
- Very mild
- Mild
- Moderate
- Severe
- Very severe

select one

Itching

- None
- Very mild
- Mild
- Moderate
- Severe
- Very severe

select one

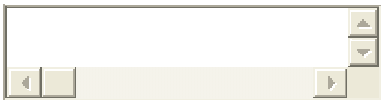
Heart palpitations

- None
- Very mild
- Mild
- Moderate
- Severe
- Very severe
- select one

Constipation / Diarrhea

- None
- Very mild
- Mild
- Moderate
- Severe
- Very severe
- select one

If you would like to say more about your responses to the questions above, please use this space:



next page >

Please note. We are unable to prevent occasional malfunctions in the questionnaire. If you experience any difficulties, please e-mail us because we want to know about them. It is likely that your information has not been lost.

This part of the survey asks about you, and should take around five minutes.

Please enter your postcode

What is your date of birth?

What is your relationship status?

- Single (never married)
- In a relationship
- Married (first marriage)
- Separated (but still legally married)
- Divorced
- Widowed
- select one

Your gender

- Male
- Female

I consider myself to be transgender:

Male to Female

Female to Male

- select one

To which of these ethnic groups do you consider you belong?

White

- select one
- British
- Irish

☐ Any other white background, please write in:

Mixed

☐ White and Black Caribbean

☐ White and Black African

☐ White and Asian

☐ Any other Mixed background, please write in:

Asian or Asian British

☐ Indian

☐ Pakistani

☐ Bangladeshi

☐ Any other Asian background, please write in:

Black or Black British

☐ Caribbean

☐ African

☐ Any other Black background
please write in:

Chinese or other ethnic group

☐ Chinese

☐ Any other,
please write in:

What is your sexual orientation? Please mark on the scale

☐ Heterosexual / Straight

☐

- ☐
- ☐ Bisexual
- ☐
- ☐
- ☐ Gay
- ☉ select one

Are you a student?

- ☐ No
- ☐ In full time education
- ☐ In part time education
- ☉ select one

Last week, were you doing any work: (1) as an employee, or on a Government sponsored training scheme (2) as self-employed/freelance, or in your own/family business?

Answer "yes" if away from work ill, on maternity leave, on holiday or temporarily laid off
Answer "yes" for any paid work, including temporary or casual work, even if only for one hour
Answer "yes" if you worked, paid or unpaid, in your own/family business

- ☐ Yes - CLICK HERE to go to next question
- ☐ No
- ☉ select one

Were you actively looking for any kind of paid work during the last 4 weeks?

- ☐ Yes
- ☐ No
- ☉ select one

If a job had been available last week, could you have started it within 2 weeks?

- ☐ Yes
- ☐ No

select one

Last week, were you waiting to start a job already obtained?

Yes

No

select one

Have you ever worked?

Yes

No - to go to next question CLICK HERE

select one

Answer the questions below for the *main* job you were doing last week, or if not working last week, your last *main* job.
Your *main* job is the job in which you usually work the most hours.
Do (did) you work as an employee or are (were) you self-employed?

Employee

Self-employed

Self-employed/freelance without employees

select one

How many people work (worked) for your employer at the place where you work (worked)?

1-9

10-24

25-499

500 or more

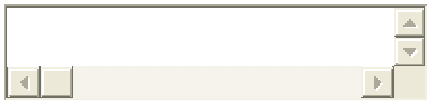
select one

What is (was) the full title of your main job?

For example, PRIMARY SCHOOL TEACHER, STATE REGISTERED NURSE, CAR MECHANIC, TELEVISION SERVICE ENGINEER, BENEFITS ASSISTANT.

Civil Servants, Local Government Officers - give job title not grade or pay band.

Describe what you do (did) in your *main* job



What is (was) the business of your employer at the place where you work (worked)?

For example, MAKING SHOES, REPAIRING CARS, SECONDARY EDUCATION, FOOD WHOLESALE, CLOTHING RETAIL, DOCTOR'S SURGERY. If you are (were) self-employed/freelance or have (had) your own business, what is (was) the nature of your business?

Civil Servants, Local Government Officers - please specify your Department.

How do you usually travel to work?

- Work mainly at or from home
- Underground, metro, light rail, tram
- Train
- Bus, minibus or coach
- Motor cycle, scooter or moped
- Driving a car or van
- Passenger in a car or van
- Taxi
- Bicycle
- On foot
- select one

How many hours a week do you usually work in your *main* job? Answer to nearest whole hour. Give average for last four weeks.

Number of hours worked a week

Which of these qualifications do you have?

Tick all the qualifications that apply or, if not specified, the nearest equivalent.

- ☐ 1 or more O levels/CSEs/GCSEs (any grades)
- ☐ 5 or more O levels, 5 or more CSEs (grade 1), 5 or moreGCSEs (grades A-C), School Certificate
- ☐ 1 or more A levels / AS levels, Higher School Certificate
- ☐ First Degree (eg BA, BSc)
- ☐ Higher Degree (eg MA, PhD, PGCE, post-graduate certificate diplomas)
- ☐ NVQ Level 1, Foundation GNVQ
- ☐ NVQ Level 2, Intermediate GMVQ
- ☐ NVQ Level 3, Advanced GNVQ
- ☐ NVQ Levels 4-5, HNC, HND
- ☐ Other qualifications (eg City and Guilds, RSA/OCR, BTEC/Edexcel)
- ☐ None of these

Scotland

- ☐ 'O' Grade, Standard Grade, Intermediate 1, Intermediate 2, GCSE, CSE, Senior Certificate *or equivalent*
- ☐ Higher Grade, CSYS, Scottish Group Award at Higher, 'A' Level, AS Level, Advanced Senior Certificate *or equivalent*
- ☐ GSVQ/SVQ Level 1 or 2, SCOTVEC/National Certificate Module, BTEC First Diploma, City and Guilds Craft, RSA Diploma *or equivalent*

- [ ] GSVQ/SVQ Level 3, ONC, OND, SCOTVEC National Diploma, City and Guilds Advanced Craft, RSA Advanced Diploma *or equivalent*
- [ ] First Degree, Higher Degree
- [ ] None of these

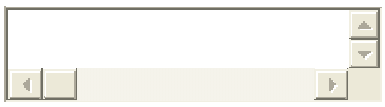
Do you have any of the following professional qualifications?

- [ ] No Professional Qualifications
- [ ] Qualified Dentist
- [ ] Qualified Teacher Status (for schools)
- [ ] Qualified Medical Doctor
- [ ] Qualified Nurse, Midwife, Health Visitor
- [ ] Other professional qualifications

Last week, were you any of the following? Tick all the boxes that apply.

- [ ] Retired
- [ ] Student
- [ ] Looking after home/family
- [ ] Permanently sick/disabled
- [ ] None of the above

If you would like to say more about your responses to the questions above, please use this space:



CLICK HERE to submit the final page >

Thank You!

Please note. We are unable to prevent occasional malfunctions in the questionnaire. If you experience any difficulties, please e-mail us because we want to know about them. It is likely that your information has not been lost.

Anonymous ID Number: 68

Participant Information Sheet

Version 3, 17th November 2005

For the research study

Personality and Stress

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this information.

What is the purpose of the study?

We are interested in how personality traits and health might be related. Stress hormones and the immune system might be involved in their relationship.

Why have I been chosen?

We are looking for healthy volunteers in the U.K. to take part.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?

If you decide to take part, we will ask you to complete a questionnaire. The questionnaire takes approximately one hour to complete. Then, over two days, we would like you to provide eight saliva samples over two consecutive weekdays:

First day

- 3 hours after you wake up
- 6 hours after you wake up
- 9 hours after you wake up
- 12 hours after you wake up

Second day

- 3 hours after you wake up
- 6 hours after you wake up
- 9 hours after you wake up
- 12 hours after you wake up

We will also ask you to write down the time you woke up. The samples will be tested for a stress hormone called cortisol and a measure of how well your immune system is working, called IgA. They will be stored with a security code that prevents anyone from identifying them as yours. The anonymous samples may be used for future research if you agree to this.

What would I have to do? (KEEP THIS SECTION WITH YOU)

Here are the instructions showing how to provide an accurate sample:

- Remember to provide the samples at the correct times
- Over two consecutive weekdays, provide a sample 3,6,9 and 12 hours after waking.
- Do not eat or drink (except water) 30 MINUTES prior to sample collection
- Do not drink ALCOHOL before and between giving samples
- Provide the sample before brushing your teeth
- Wash your mouth out with water 10 minutes prior to giving a sample
- Swallow the saliva that is already in your mouth
- Place the straw in your mouth under the tongue. Allow saliva to collect PASSIVELY into the tube for a timed 2-minute period, then spit it into the tube.
- Throw the straw away.
- Seal the tube with the screw cap and label with the stickers provided
- After sampling keep all tubes together in a plastic bag and place in a freezer as soon as possible (your domestic freezer is suitable)
- Return all eight samples to the 'Personality and Stress Study' box in the psychology concourse. You can use a carrier bag with a knot tied in it.
- Spare tubes are available on request.

What are the possible disadvantages of taking part?

Some people find providing a saliva sample uncomfortable. You may find that remembering to provide the samples at the correct time is inconvenient.

What are the possible advantages of taking part?

We will provide a free report on the personality questionnaire, which you may find useful.

What if something goes wrong?

If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, please contact Dr. Martha Whiteman.

Will my taking part in this study be kept confidential?

All information collected about you during the course of the research will be kept strictly confidential. Any information shared outside the research team will have names and addresses removed so that you cannot be recognised from it.

What will happen to the results of the research study?

The results of the study will be written up into reports and publications. We will provide a simplified version of the results written in lay terms, in 2007.

Who is organising and funding the research?

The research is funded by the Economic and Social Research Council (ESRC) and organized by the Department of Psychology, The University of Edinburgh.

Who has reviewed the study?

This study reviewed and granted approval by the Multi Region Ethics Committee for Scotland (A).

Contact for Further Information

Mr Gareth Hagger-Johnson
Psychology
The School of Philosophy Psychology and Language Sciences
The University of Edinburgh
7 George Square
Edinburgh EH8 9JZ

G.E.Hagger-Johnson@ed.ac.uk
0131 650 3272 (office)
07967 157 241 (mobile)

Dr. Martha Whiteman
Psychology
The School of Philosophy Psychology and Language Sciences
The University of Edinburgh
7 George Square
Edinburgh EH8 9JZ

M.Whiteman@ed.ac.uk
0131 650 3317

Anonymous ID Number: 68

CONSENT FORM (DETACH)

Version 3
17th November 2005
Title of Project: Personality and Stress

Name of Researchers: Gareth Hagger-Johnson and Dr. Martha Whiteman

Please initial box

1. I confirm that I have read and understand the information sheet dated ..... (version .....) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐

3. I agree to take part in the above study. ☐

4. I agree to allow my anonymous saliva samples to be used in future research ☐

_____ _____
Name of participant Date Signature

_____ _____
E-mail address of participant

_____ _____
Name of researcher Date Signature

Please return this consent form to Gareth Hagger-Johnson's mail folder in the filing cabinet in the mail room of the psychology concourse

Anonymous ID Number: 68

SF-36 Health Questionnaire

1. In general, would you say your health is:

(Circle One Number)

Excellent	1
Very good	2
Good	3
Fair	4
Poor	5

2. **Compared to one year ago**, how would you rate your health in general **now**?

(Circle One Number)

Much better now than one year ago	1
Somewhat better now than one year ago	2
About the same	3
Somewhat worse now than one year ago	4
Much worse now than one year ago	5

The following items are about activities you might do during a typical day. Does **your health now limit you** in these activities? If so, how much?

(Circle One Number on Each Line)

		Yes, Limited a Lot	Yes, Limited a Little	No, Not Limited at All
3.	**Vigorous activities,** such as running, lifting heavy objects, participating in strenuous sports	1	2	3
4.	**Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	1	2	3
5.	Lifting or carrying groceries	1	2	3
6.	Climbing **several** flights of stairs	1	2	3
7.	Climbing **one** flight of stairs	1	2	3
8.	Bending, kneeling, or stooping	1	2	3
9.	Walking **more than a mile**	1	2	3
10.	Walking **several blocks**	1	2	3
11.	Walking **one block**	1	2	3
12.	Bathing or dressing yourself	1	2	3

During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of your physical health**?

(Circle One Number on Each Line)

		Yes	No
13.	Cut down the **amount of time** you spent on work or other activities	1	2
14.	**Accomplished less** than you would like	1	2
15.	Were limited in the **kind** of work or other activities	1	2
16.	Had **difficulty** performing the work or other activities (for example, it took extra effort)	1	2

During the **past 4 weeks,** have you had any of the following problems with your work or other regular daily activities **as a result of any emotional problems** (such as feeling depressed or anxious)?

(Circle One Number on Each Line)

		Yes	No
17.	Cut down the **amount of time** you spent on work or other activities	1	2
18.	**Accomplished less** than you would like	1	2
19.	Didn't do work or other activities as **carefully** as usual	1	2

20. During the **past 4 weeks,** to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

(Circle One Number)

Not at all	1
Slightly	2
Moderately	3
Quite a bit	4
Extremely	5

21. How much **bodily** pain have you had during the **past 4 weeks**?

(Circle One Number)

None	1
Very mild	2
Mild	3
Moderate	4
Severe	5
Very severe	6

22. During the **past 4 weeks,** how much did **pain** interfere with your normal work (including both work outside the home and housework)?

(Circle One Number)

Not at all	1
A little bit	2
Moderately	3
Quite a bit	4
Extremely	5

These questions are about how you feel and how things have been with you **during the past 4 weeks**. For each question, please give the one answer that comes closest to the way you have been feeling.

How much of the time during the **past 4 weeks** . . .

(Circle One Number on Each Line)

		All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time
23.	Did you feel full of pep?	1	2	3	4	5	6
24.	Have you been a very nervous person?	1	2	3	4	5	6
25.	Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
26.	Have you felt calm and peaceful?	1	2	3	4	5	6
27.	Did you have a lot of energy?	1	2	3	4	5	6
28.	Have you felt downhearted and blue?	1	2	3	4	5	6
29.	Did you feel worn out?	1	2	3	4	5	6
30.	Have you been a happy person?	1	2	3	4	5	6
31.	Did you feel tired?	1	2	3	4	5	6

32. During the **past 4 weeks,** how much of the time has your **physical health or emotional problems** interfered with your social activities (like visiting with friends, relatives, etc.)?

(Circle One Number)

All of the time	1
Most of the time	2
Some of the time	3
A little of the time	4
None of the time	5

How TRUE or FALSE is each of the following statements for you.

(Circle One Number on Each Line)

		Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
33.	I seem to get sick a little easier than other people.....	1	2	3	4	5
34.	I am as healthy as anybody I know.	1	2	3	4	5
35.	I expect my health to get worse.....	1	2	3	4	5
36.	My health is excellent.	1	2	3	4	5

Health Behaviours Questionnaire

Do you smoke?	Yes / No
If 'YES' How many cigarettes do you smoke per day? How long have you been a smoker? YEARS: MONTHS:	If 'NO': have you ever smoked cigarettes for more than one month? Yes / No If 'YES': how many cigarettes per day did you used to smoke? How long were you a smoker? YEARS: MONTHS:
Would you describe yourself as	Circle one: • A non-drinker • A very occasional drinker (special occasions only) • An occasional drinker • A regular drinker
On how many days over the past two weeks did you have a drink?	_____days
On the days that you did drink, how many units did you have, on average?	_____units (approx.) What is a unit? • A half pint of standard strength beer/lager(3.5% ABV) • A small glass of table wine 125ml (8% ABV) • A small pub measure of spirits (25ml)
Would you like to reduce the amount that you drink?	Yes / No

Demographic and Equal Opportunities Questionnaire

- Please circle the appropriate answer
- Please do not write your name on this questionnaire
- You do not have to answer any questions that you would prefer to leave blank

U.K. Postcode	
What is your date of birth?	
What is your relationship status?	Single (never married) In a relationship Married (first marriage) Married (second or later marriage) Separated (but still legally married) Divorced Widowed
Your gender	Male Female I consider myself to be transgender: Male to Female Female to Male
Are you a student?	No In full time education In part time education
Sexual orientation	0. Heterosexual / Straight 1. 2. 3. Bisexual 4. 5. 6. Gay
To which of these ethnic groups do you consider you belong?	White Mixed Asian or Asian British Black or Black British Chinese or Other Ethnic Group
Please list any current medications that you are taking:	
Last week, were you doing any work: (1) as an employee, or on a Government sponsored training scheme (2) as self-employed/freelance, or in your own/family business?	Yes – **go to Section B** No Answer "yes" if away from work ill, on maternity leave, on holiday or temporarily laid off Answer "yes" for any paid work, including temporary or casual work, even if only for one hour Answer "yes" if you worked, paid or unpaid, in your own/family business
Were you actively looking for any kind of paid work during the last 4 weeks?	
If a job had been available last week, could you have started it within 2 weeks?	

Last week, were you waiting to start a job already obtained?	
Last week, were you any of the following?	Retired Student Looking after home/family Permanently sick/disabled None of the above
Have you ever worked?	Yes No – **go to Section C**
Section B. Answer the questions below for the *main* job you were doing last week, or if not working last week, your last *main* job. Your *main* job is the job in which you usually work the most hours.	
Do (did) you work as an employee or are (were) you self-employed?	Employee Self-employed Self-employed/freelance without employees
How many people work (worked) for your employer at the place where you work (worked)?	1-9 10-24 25-499 500 or more
What is (was) the full title of your main job? For example, PRIMARY SCHOOL TEACHER, STATE REGISTERED NURSE, CAR MECHANIC, TELEVISION SERVICE ENGINEER, BENEFITS ASSISTANT. Civil Servants, Local Government Officers - give job title not grade or pay band.	
Describe what you do (did) in your *main* job. For example, MAKING SHOES, REPAIRING CARS, SECONDARY EDUCATION, FOOD WHOLESALE, CLOTHING RETAIL, DOCTOR'S SURGERY. If you are (were) self-employed/freelance or have (had) your own business, what is (was) the nature of your business? Civil Servants, Local Government Officers - please specify your Department.	
What is (was) the business of your employer at the place where you work (worked)?	
How do you usually travel to work?	Work mainly at or from home Underground, metro, light rail, tram Train Bus, minibus or coach Motor cycle, scooter or moped Driving a car or van Passenger in a car or van Taxi Bicycle On foot
How many hours a week do you usually work in your *main* job? Answer to nearest whole hour. Give average for last four weeks.	Number of hours worked a week:

Section C. Which of these qualifications do you have?	Scotland:
1 or more O levels/CSEs/GCSEs (any grades) 5 or more O levels, 5 or more CSEs (grade 1), 5 or more GCSEs (grades A-C), School Certificate 1 or more A levels / AS levels, Higher School Certificate First Degree (eg BA, BSc) Higher Degree (eg MA, PhD, PGCE, post-graduate certificate diplomas) NVQ Level 1, Foundation GNVQ NVQ Level 2, Intermediate GMVQ NVQ Level 3, Advanced GNVQ NVQ Levels 4-5, HNC, HND Other qualifications (eg City and Guilds, RSA/OCR, BTEC/Edexcel) None of these	'O' Grade, Standard Grade, Intermediate 1, Intermediate 2, GCSE, CSE, Senior Certificate or equivalent Higher Grade, CSYS, Scottish Group Award at Higher, 'A' Level, AS Level, Advanced Senior Certificate or equivalent GSVQ/SVQ Level 1 or 2, SCOTVEC/National Certificate Module, BTEC First Diploma, City and Guilds Craft, RSA Diploma or equivalent GSVQ/SVQ Level 3, ONC, OND, SCOTVEC National Diploma, City and Guilds Advanced Craft, RSA Advanced Diploma or equivalent First Degree, Higher Degree None of these
Do you have any of the following professional qualifications?	No Professional Qualifications Qualified Dentist Qualified Teacher Status (for schools) Qualified Medical Doctor Qualified Nurse, Midwife, Health Visitor Other professional qualifications
Last week, were you any of the following? Tick all the boxes that apply.	Retired Student Looking after home/family Permanently sick/disabled None of the above
If you would like to say more about your responses to the questions above, please use this space:	

Anonymous ID Number: 68

Perceived Stress Questionnaire

The questions in this scale ask you about your feelings and thoughts during the last month. In each case, you will be asked to indicate how often you felt or thought a certain way. Although some of the questions are similar, there are differences between them and you should treat each one as a separate question. The best approach is to answer each question fairly quickly. That is, don't try to count up the number of times you felt a particular way, but rather indicate the alternative that seems like a reasonable estimate.

For each question choose from the following alternatives:

0 = never
1 = almost never
2 = sometimes
3 = fairly often
4 = very often

In the last month, how often have you been upset because of something that happened unexpectedly?	0	1	2	3	4
In the last month, how often have you felt that you were unable to control the important things in your life?	0	1	2	3	4
In the last month, how often have you felt nervous and stressed?	0	1	2	3	4
In the last month, how often have you dealt with irritating life hassles?	0	1	2	3	4
In the last month, how often have you felt that you were effectively coping with important changes that were occurring in your life?	0	1	2	3	4
In the last month, how often have you felt confident about your ability to handle your personal problems?	0	1	2	3	4
In the last month, how often have you felt that things were going your way?	0	1	2	3	4
In the last month, how often have you found that you could not cope with all the things you had to do?	0	1	2	3	4
In the last month, how often have you been able to control irritations in your life?	0	1	2	3	4
In the last month, how often have you felt that you were on top of things?	0	1	2	3	4
In the last month, how often have you been angered because of things that happened that were outside of your control?	0	1	2	3	4
In the last month, how often have you found yourself thinking about things that you have to accomplish?	0	1	2	3	4
In the last month, how often have you been able to control the way you spend your time?	0	1	2	3	4
In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?	0	1	2	3	4

Hassles in the Past Month

Hassles are irritants that can range from minor annoyances to fairly major pressures, problems, or difficulties. They can occur few or many times. Listed on the following pages are a number of ways in which a person can feel hassled. First, circle the hassles that have happened to you in the past month. Then look at the numbers to the right of the items you circled. Indicate by circling 1, 2, or 3 how SEVERE each of the circled hassles has been to you in the past month. If a hassle did not occur in the past month do NOT circle it.

HOW SEVERE? 1 = Somewhat severe, 2 = Moderately severe, 3 = Extremely severe

1.	Displacing or losing things	1	2	3
2.	Troublesome neighbours	1	2	3
3.	Social obligations	1	2	3
4.	Inconsiderate smokers	1	2	3
5.	Troubling thoughts about your future	1	2	3
6.	Thoughts about death	1	2	3
7.	Health of a family member	1	2	3
8.	Not enough money for clothing	1	2	3
9.	Not enough money for housing	1	2	3
10.	Concerns about owing money	1	2	3
11.	Concerns about getting credit	1	2	3
12.	Concerns about money for emergencies	1	2	3
13.	Someone owes you money	1	2	3
14.	Financial responsibility for someone who doesn't live with you	1	2	3
15.	Cutting down on electricity, gas, etc.	1	2	3
16.	Smoking too much	1	2	3
17.	Use of alcohol	1	2	3
18.	Personal use of drugs	1	2	3
19.	Too many responsibilities	1	2	3
20.	Decisions about having children	1	2	3
21.	Non-family members living in your house	1	2	3
22.	Care for pet	1	2	3
23.	Planning meals	1	2	3
24.	Concerned about the meaning of life	1	2	3
25.	Trouble relaxing	1	2	3
26.	Trouble making decisions	1	2	3
27.	Problems getting along with fellow workers	1	2	3
28.	Customer or clients giving you a hard time	1	2	3
29.	Home maintenance (inside)	1	2	3
30.	Concerns about job security	1	2	3
31.	Concerns about retirement	1	2	3
32.	Laid-off or out of work	1	2	3
33.	Don't like current work duties	1	2	3
34.	Don't like fellow workers	1	2	3
35.	Not enough money for basic necessities	1	2	3
36.	Not enough money for food	1	2	3
37.	Too many interruptions	1	2	3
38.	Unexpected company	1	2	3
39.	Too much time on hands	1	2	3
40.	Having to wait	1	2	3

41.	Concerns about accidents	1	2	3
42.	Being lonely	1	2	3
43.	Not enough money for health care	1	2	3
44.	Fear of confrontation	1	2	3
45.	Financial security	1	2	3
46.	Silly practical mistakes	1	2	3
47.	Inability to express yourself	1	2	3
48.	Physical illness	1	2	3
49.	Side effects of medication	1	2	3
50.	Concerns about medical treatment	1	2	3
51.	Physical appearance	1	2	3
52.	Fear of rejection	1	2	3
53.	Difficulties with getting pregnant	1	2	3
54.	Sexual problems that result from physical problems	1	2	3
55.	Sexual problems other than those resulting from physical problems	1	2	3
56.	Concerns about health in general	1	2	3
57.	Not seeing enough people	1	2	3
58.	Friends or relatives too far away	1	2	3
59.	Preparing meals	1	2	3
60.	Wasting time	1	2	3
61.	Car maintenance	1	2	3
62.	Filling-in forms	1	2	3
63.	Neighbourhood deterioration	1	2	3
64.	Financing children's education	1	2	3
65.	Problems with employees	1	2	3
66.	Problems in job due to being a woman or man	1	2	3
67.	Declining physical abilities	1	2	3
68.	Being exploited	1	2	3
69.	Concerns about bodily functions	1	2	3
70.	Rising prices of basic necessities	1	2	3
71.	Not getting enough rest	1	2	3
72.	Not getting enough sleep	1	2	3
73.	Problems with ageing parents	1	2	3
74.	Problems with your children	1	2	3
75.	Problems with persons younger than yourself	1	2	3
76.	Problems with your partner	1	2	3
77.	Difficulties seeing or hearing	1	2	3
78.	Overloaded with family responsibilities	1	2	3
79.	Too many things to do	1	2	3
80.	Unchallenging work	1	2	3
81.	Concerns about meeting high standards	1	2	3
82.	Financial dealings with friends or acquaintances	1	2	3
83.	Job dissatisfactions	1	2	3
84.	Worries about decisions to change jobs	1	2	3
85.	Trouble with reading, writing, or spelling abilities	1	2	3
86.	Too many meetings	1	2	3
87.	Problems with divorce or separation	1	2	3
88.	Trouble with arithmetic skills	1	2	3
89.	Gossip	1	2	3
90.	Legal problems	1	2	3

91. Concerns about weight	1	2	3
92. Not enough time to do things you need to do	1	2	3
93. Television	1	2	3
94. Not enough personal energy	1	2	3
95. Concerns about inner conflicts	1	2	3
96. Feel conflict over what to do	1	2	3
97. Regrets over past decisions	1	2	3
98. Menstrual (period) problems	1	2	3
99. The weather	1	2	3
100. Nightmares	1	2	3
101. Concerns about getting ahead	1	2	3
102. Hassles from boss or supervisor	1	2	3
103. Difficulties with friends	1	2	3
104. Not enough time for family	1	2	3
105. Transport problems	1	2	3
106. Not enough money for transport	1	2	3
107. Not enough money for entertainment and recreation	1	2	3
108. Shopping	1	2	3
109. Prejudice and discrimination from others	1	2	3
110. Property, investments or taxes	1	2	3
111. Not enough time for entertainment and recreation	1	2	3
112. Garden or outside home maintenance	1	2	3
113. Concerns about news events	1	2	3
114. Noise	1	2	3
115. Crime	1	2	3
116. Traffic	1	2	3
117. Pollution	1	2	3
Have we missed any of your hassles? If so, write them below:	1	2	3

One more thing: has there been a change in your life that affected how you answered this scale? If so tell us what it was:

Uplifts in the Past Month

Uplifts are events that make you feel good. They can be sources of peace, satisfaction, or joy. Some occur often, others are relatively rare. On the following pages, circle the events that have made you feel good **in the past month**. Then look at the numbers to the right of the items you circled. Indicate by circling a 1, 2, or 3 how OFTEN each of the circled uplifts has occurred in the last month. If an uplift did not occur in the last month do NOT circle it.

HOW OFTEN? 1 = Somewhat often, 2 = Moderately often, 3 = Extremely often

1.	Getting enough sleep	1	2	3
2.	Practising your hobby	1	2	3
3.	Being lucky	1	2	3
4.	Saving money	1	2	3
5.	Nature	1	2	3
6.	Liking fellow workers	1	2	3
7.	Not working (on holiday, laid off, etc.)	1	2	3
8.	Gossiping	1	2	3
9.	Successful financial dealings	1	2	3
10.	Being rested	1	2	3
11.	Feeling healthy	1	2	3
12.	Finding something presumed lost	1	2	3
13.	Recovering from illness	1	2	3
14.	Staying, or getting into, good physical shape	1	2	3
15.	Being with children	1	2	3
16.	'Pulling something off': getting away with something	1	2	3
17.	Visiting, phoning, or writing to someone	1	2	3
18.	Relating well to your partner	1	2	3
19.	Completing a task	1	2	3
20.	Giving a compliment	1	2	3
21.	Meeting family responsibilities	1	2	3
22.	Relating well to friends	1	2	3
23.	Being efficient	1	2	3
24.	Meeting your responsibilities	1	2	3
25.	Stopping, or cutting down on, alcohol	1	2	3
26.	Stopping, or cutting down on, smoking	1	2	3
27.	Solving an on-going practical problem	1	2	3
28.	Daydreaming	1	2	3
29.	Weight	1	2	3
30.	Financially supporting someone who doesn't live with you	1	2	3
31.	Sex	1	2	3
32.	Friendly neighbours	1	2	3
33.	Having enough time to do what you want	1	2	3
34.	Divorce or separation	1	2	3
35.	Eating out	1	2	3
36.	Having enough (personal energy)	1	2	3
37.	Resolving inner conflicts	1	2	3
38.	Being with older people	1	2	3
39.	Finding no prejudice or discrimination when you expect it	1	2	3
40.	Cooking	1	2	3

41. Capitalizing on an unexpected opportunity	1	2	3
42. Using drugs or alcohol	1	2	3
43. Life being meaningful	1	2	3
44. Being well-prepared	1	2	3
45. Eating	1	2	3
46. Relaxing	1	2	3
47. Having the 'right' amount of things to do	1	2	3
48. Being visited or phoned, or receiving a letter	1	2	3
49. The weather	1	2	3
50. Thinking about the future	1	2	3
51. Spending time with family	1	2	3
52. Home (inside) pleasing to you	1	2	3
53. Being with younger people	1	2	3
54. Buying things for the house	1	2	3
55. Reading	1	2	3
56. Shopping	1	2	3
57. Smoking	1	2	3
58. Buying clothes	1	2	3
59. Giving a present	1	2	3
60. Becoming pregnant, your spouse becoming pregnant	1	2	3
61. Getting a present	1	2	3
62. Having enough money for health care	1	2	3
63. Travelling or commuting	1	2	3
64. Doing gardening or outside housework	1	2	3
65. Having enough money for transport	1	2	3
66. Health of a family member improving	1	2	3
67. Resolving conflicts over what to do	1	2	3
68. Thinking about health	1	2	3
69. Being a 'good' listener	1	2	3
70. Socializing (parties, being with friends, etc.)	1	2	3
71. Making a friend	1	2	3
72. Sharing something	1	2	3
73. Having someone listen to you	1	2	3
74. Your garden or the outside of your house is pleasing	1	2	3
75. Looking forward to retirement	1	2	3
76. Having enough money for entertainment and recreation	1	2	3
77. Entertainment (movies, concerts, TV, etc.)	1	2	3
78. Good news on local or world level	1	2	3
79. Getting good advice	1	2	3
80. Recreation (sports, games, hiking, etc.)	1	2	3
81. Paying off debts	1	2	3
82. Using skills well at work	1	2	3
83. Past decisions 'working out'	1	2	3
84. Growing as a person	1	2	3
85. Being complimented	1	2	3
86. Having good ideas at work	1	2	3
87. Improving or gaining new skills	1	2	3
88. Job satisfying despite discrimination due to your sex	1	2	3
89. Free time	1	2	3
90. Expressing yourself well	1	2	3
91. Laughing	1	2	3

92.	Holidaying without spouse or children	1	2	3
93.	Liking work duties	1	2	3
94.	Having good credit	1	2	3
95.	Music	1	2	3
96.	Getting unexpected money	1	2	3
97.	Changing jobs	1	2	3
98.	Dreaming	1	2	3
99.	Having fun	1	2	3
100.	Going somewhere different	1	2	3
101.	Deciding to have children	1	2	3
102.	Enjoying non-family members living in your house	1	2	3
103.	Pets	1	2	3
104.	Car running well	1	2	3
105.	Neighbourhood improving	1	2	3
106.	Children's accomplishments	1	2	3
107.	Things going well with employee(s)	1	2	3
108.	Pleasant smells	1	2	3
109.	Receiving love	1	2	3
110.	Successfully avoiding or dealing with bureaucracy or institutions	1	2	3
111.	Making decisions	1	2	3
112.	Thinking about the past	1	2	3
113.	Giving good advice	1	2	3
114.	Praying	1	2	3
115.	Meditating	1	2	3
116.	Fresh air	1	2	3
117.	Confronting someone or something	1	2	3
118.	Being accepted	1	2	3
119.	Giving love	1	2	3
120.	Boss pleased with your work	1	2	3
121.	Being alone	1	2	3
122.	Feeling safe	1	2	3
123.	Working well with fellow workers	1	2	3
124.	Knowing your job is secure	1	2	3
125.	Feeling safe in your neighbourhood	1	2	3
126.	Doing volunteer work	1	2	3
127.	Contributing to a charity	1	2	3
128.	Learning something	1	2	3
129.	Being 'at one' with the world	1	2	3
130.	Fixing/repairing something (besides at your job)	1	2	3
131.	Making something (besides at your job)	1	2	3
132.	Exercising	1	2	3
133.	Meeting a challenge	1	2	3
134.	Hugging and/or kissing	1	2	3
135.	Flirting	1	2	3
136.	Have we missed any of your uplifts? If so, write them in below:	1	2	3

One more thing: has there been a change in your life that affected how you answered this scale? If so, tell us what it was:

Please return this questionnaire to the 'Personality and Stress Study' box in the psychology concourse

Mplus syntax

Conscientiousness and health behaviours with age and sex as covariates (chapter 7)

```
MODEL: c BY c1-c6;
y BY y1-y3 c6;
y ON c c5;
c y ON age sex;
c3 WITH c6;
```

Conscientiousness, health behaviours and health outcomes with age and educational level as covariates(chapter 7)

```
MODEL: c BY c1-c6;
h BY h1-h5 fagday units;
c5 WITH c2;
c3 WITH c2;
h ON c age  ;
fagday WITH h2;
h5 WITH c6;
mcs ON c ;
mcs on h@0;
c ON age ;
edlevel ON c;
mcs WITH edlevel@0;
```

Lifestyle and CVD risk factors influencing all-cause mortality in males, with age as covariate (chapter 8)

```
USEOBSERVATIONS =
age > 39 AND bpcat < 7 AND male EQ 1
AND hpb1 EQ 0
AND hpb2 EQ 0
AND discon5-discon6 EQ 0
AND discon15 EQ 0;
SURVIVAL = agedeath;
TIMECENSORED = cenall (0 = NOT 1 = RIGHT);
MODEL: f1 BY z1-z4;
f2 BY Zwhr Zbpsys Zbpdias;
f1 WITH f2@0;
f1 f2 ON age;
agedeath ON age f1;
agedeath ON f2@0;
ANALYSIS: ALGORITHM = INTEGRATION;
INTEGRATION = 10;
BASEHAZARD = OFF;
```

Latent growth curve model of diurnal cortisol profile with Neuroticism facets as covariates (chapter 10)

```
MODEL: i s | x9@3 x10@2 x11@1 x12@0;
s@0;
i ON male n1-n6 pss pcs mcs waking;
```

References

Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, N.J. ; Hove: Lawrence Erlbaum.

Abercrombie, H. C., Giese-Davis, J., Sephton, S., Epel, E. S., Turner-Cobb, J. M., & Spiegel, D. (2004). Flattened cortisol rhythms in metastatic breast cancer patients. *Psychoneuroendocrinology*, *29*(8), 1082–1092.

Acton, G. S., & Revelle, W. (2004). Evaluation of ten psychometric criteria for circumplex structure. *MPR-online*, *9*, 1–27.

Adler, N., & Matthews, K. (1994). Health psychology: Why do some people get sick and some stay well? *Annual Review of Psychology*, *45*(1), 229–259.

Allen, P. I. M., Batty, K. A., & Dodd, C. A. S. (1985). Dissociation between emotional and endocrine responses preceding an academic examination in male medical students. *Journal of Endocrinology*, *107*(2), 163–170.

Almada, S. J., Zonderman, A. B., Shekelle, R. B., Dyer, A. R., Daviglus, M. L., Costa, P. T., et al. (1991). Neuroticism and cynicism and risk of death in middle-aged men: the western electric study. *Psychosomatic Medicine*, *53*(2), 165–175.

Ambrose, J. A., & Barua, R. S. (2004). The pathophysiology of cigarette smoking and cardiovascular disease: An update. *Journal of the American College of Cardiology*, *43*(10), 1731–1737.

Amelang, M. (1997). Using personality variables to predict cancer and heart disease. *European Journal of Personality*, *11*(5), 319–342.

Amelang, M., Schmidt-Rathjens, C., & Matthews, G. (1996). Personality, cancer and coronary heart disease: Further evidence on a controversial issue. *British Journal of Health Psychology*, *1*(3), 191–205.

American-Psychiatric-Association. (1994). *Diagnostic and statistical manual of mental disorders, DSM-IV-TR (text revision)* (Fourth ed.). Washington, DC: American Psychiatric Publishing.

Andrews, W., Parker, G., & Barrett, E. (1998). The SSRI antidepressants: Exploring their "other" possible properties. *Journal of Affective Disorders*, *49*(2), 141–144.

Armenian, H. K., & Shapiro, S. (1998). *Epidemiology and health services*. Oxford: Oxford University Press.

Ashton, M. C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior*, *19*(3), 289–303.

Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality*, *29*(4), 432–442.

Ashton, M. C., & Lee, K. (2005a). A defence of the lexical approach to the study of personality structure. *European Journal of Personality*, *19*(1), 5–24.

Ashton, M. C., & Lee, K. (2005b). Honesty-humility, the big five, and the five-factor model. *Journal of Personality*, *73*(5), 1321–1354.

Asparouhov, T. (2006). Continuous - time survival analysis in mplus. *Mplus technical appendices.*.

Atkins, D., & Clancy, C. (2004). Multiple risk factors interventions: Are we up to the challenge? *American Journal of Preventive Medicine*, *27*(2, Suppl. 1), 102–103.

Austin, E. J., & Deary, I. J. (2002). Why smart people can be so stupid. In R. J. Sternberg (Ed.), (pp. 187–211). Cambridge: Cambridge University Press.

Aviv, A. L., Zelenski, J. M., Rallo, L., & Larsen, R. J. (2002). Who comes when: Personality differences in early and later participation in a university subject pool. *Personality and Individual Differences*, *33*(3), 487–496.

Babcock, Q., & Byrne, T. (2000). Student perceptions of methylphenidate abuse at a public liberal arts college. *Journal of American college health*, *49*(3), 143–145.

Bagby, R. M., Levitan, R. D., Kennedy, S. H., Levitt, A. J., & Joffe, R. T. (1999). Selective alteration of personality in response to noradrenergic and serotonergic antidepressant medication in depressed sample: Evidence of nonspecificity. *Psychiatry Research*, *86*(3), 211–216.

Bailey, R. D., Foote, W. E., & Throckmorton, B. (2000). Human sexual behaviour: A comparison of college and internet surveys. , 141–168.

Bamforth, C. W. (2002). Nutritional aspects of beer - a review. *Nutrition Research*, *22*(1-2), 227–237.

Bandelow, B., Wedekind, D., Sandvoss, V., Broocks, A., Hajak, G., Pauls, J., et al. (2000). Diurnal variation of cortisol in panic disorder. *Psychiatry Research*, *95*(3), 245–250.

Bartley, M., Fitzpatrick, R., Firth, D., & Marmot, M. (2000). Social distribution of cardiovascular disease risk factors: Change among men in England 1984-1993. *Journal of Epidemiology and Community Health*, *54*(11), 806–814.

Batty, G. D., Shipley, M. J., Marmot, M., & Smith, G. D. (2001). Physical activity and cause-specific mortality in men: Further evidence from the Whitehall study. *European Journal of Epidemiology*, *17*(9), 863–869.

Batty, G. D., Shipley, M. J., Marmot, M. G., & Smith, G. D. (2003). Leisure time physical activity and disease-specific mortality among men with chronic bronchitis: evidence from the Whitehall study. *American Journal of Public Health*, *93*(5), 817–821.

Bekaert, S., De Meyer, T., & Van Oostveldt, P. (2005). Telomere attrition as ageing biomarker. *Anticancer Research*, *25*(4), 3011–3022.

Belloc, N. B. (1973). Relationship of health practices and mortality. *Preventive Medicine*, *2*(1), 67–81.

Belloc, N. B., & Breslow, L. (1972). Relationship of physical health status and health practices. *Preventive Medicine*, *1*(3), 409–421.

Berrigan, D., Dodd, K., Troiano, R. P., Krebs-Smith, S. M., & Ballard. (2003). Patterns of health behavior in U.S. adults. *Preventive Medicine*, *36*(5), 615–623.

Best, S. J., Krueger, B., Hubbard, C., & Smith, A. (2001). An assessment of the generalizability of internet surveys. *Social Science Computer Review*, *19*(2), 131–145.

Bhopal, R. (2002). *Concepts of epidemiology: An integrated introduction to the ideas, theories, principles and methods of epidemiology*. Oxford: Oxford University Press, USA.

Bigaard, J., Frederiksen, K., Tjnneland, A., Thomsen, B. L., Overvad, K., Heitmann, B. L., et al. (2004). Waist and hip circumferences and all-cause mortality: Usefulness of the waist-to-hip ratio? *International Journal of Obesity*, *28*(6), 741–747.

Birks, Y., Roebuck, A., & Thompson, D. R. (2004). A validation study of the cardiac depression scale (CDS) in a UK population. *British Journal of Health Psychology*, *9*(1), 15–24.

Blane, B. D. (2006). The life course, the social gradient, and health. In M. Marmot & R. G. Wilkinson (Eds.), *Social determinants of health* (Second ed., pp. 54–77). Oxford: Oxford University Press.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*(2), 187–215.

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin*, *130*(6), 887–919.

Bolger, N., & Schilling, E. A. (1991). Personality and the problems of everyday life: The role of neuroticism in exposure and reactivity to daily stressors. *Journal of Personality*, *59*(3), 355–386.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314.

Bollen, K. A., & Curran, P. J. (2005). *Latent curve models : A structural equation perspective*. Hoboken, N.J. ; Chichester: Wiley Interscience.

Boniface, D. R., Cottee, M. J., Neal, D., & Skinner, A. (2001). Social and demographic factors predictive of change over seven years in CHD-related behaviours in men aged 18-49 years. *Public Health*, *115*(4), 246–252.

Boniface, D. R., & Tefft, M. E. (1997). The application of structural equation modelling to the construction of an index for the measurement of health-related behaviours. *The Statistician*, *46*(4), 505–514.

Bonow, R. O. (2002). Primary prevention of cardiovascular disease: A call to action. *Circulation*, *106*(25), 3140–3141.

Booth-Kewley, S. B., & Vickers, R. R. (1994). Associations between major domains of personality and health behavior. *Journal of Personality*, *62*(3), 281–298.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press.

Borsboom, D., Mellenbergh, G. J., & Heerden, J. van. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219.

Bower, J. E., Ganz, P. A., Dickerson, S. S., Petersen, L., Aziz, N., & Fahey, J. L. (2005). Diurnal cortisol rhythm and fatigue in breast cancer survivors. *Psychoneuroendocrinology*, *30*(1), 92–100.

Bowers, K. S. (1987). Toward a multidimensional view of personality and health. *Journal of Personality*, *55*(2), 343–349.

Bowling, A. (1997). *Measuring health: A review of quality of life measurement scales* (Second ed.). Buckingham: Open University Press.

Braveman, P. A., Cubbin, C., Egerter, S., Chideya, S., Marchi, K. S., Metzler, M., et al. (2005). socio-economic status in health research: one size does not fit all. *JAMA*, *294*(22), 2879–2888.

Breslow, L., & Enstrom, J. E. (1980a). Persistence of health habits and their relationship to mortality. *Preventive Medicine*, *9*(4), 469–483.

Breslow, L., & Enstrom, J. E. (1980b). Persistence of health habits and their relationship to mortality. *Preventive Medicine*, *9*(4), 469–483.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, *54*(1), 106–148.

Brock, B. M., Haefner, D. P., & Noble, D. S. (1988). Alameda county redux: replication in michigan. *Preventive Medicine*, *17*(4), 483–495.

Brock, D. W. (1998). Enhancements of human function: Some distinctions for policymakers. In E. Parens (Ed.), *Enhancing human traits* (pp. 48–69). Washington: Georgetown University Press.

Brunner, E., & Marmot, M. (2006). Social organization, stress, and health. In M. Marmot & R. G. Wilkinson (Eds.), *Social determinants of health* (Second ed., pp. 6–30). Oxford: Oxford University Press.

Buchanan, T. (2000). Potential of the internet for personality research. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. 121–140). San Diego, CA: Academic Press.

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the internet. *European Journal of Psychological Assessment*, *21*(2), 115–127.

Burke, H. M., Davis, M. C., Otte, C., & Mohr, D. C. (2005). Depression and cortisol responses to psychological stress: A meta-analysis. *Psychoneuroendocrinology*, *30*(9), 846–856.

Butt, T. (2004). *Understanding people*. Basingstoke: Palgrave Macmillan.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Capitanio, J. P., Mendoza, S. P., & Bentson, K. L. (2004). Personality characteristics and basal cortisol concentrations in adult male rhesus macaques (Macaca mulatta). *Psychoneuroendocrinology*, *29*(10), 1300–1308.

Caraballo, R. S., Giovino, G. A., & Pechacek, T. F. (2004). Self-reported cigarette smoking vs. serum cotinine among u.s. adolescents. *Nicotine and Tobacco Research*, *6*(1), 19–25.

Caraballo, R. S., Giovino, G. A., Pechacek, T. F., & Mowery, P. D. (2001). Factors associated with discrepancies between self-reports on cigarette smoking and measured serum cotinine levels among persons aged 17 years or older : Third National Health and Nutrition Examination Survey, 1988-1994. *American Journal of Epidemiology*, *153*(8), 807–814.

Caspi, A., Harrington, H., Moffitt, T. E., Begg, D., Dickson, N., Langley, J., et al. (1997). Personality differences predict health-risk behaviors in young adulthood: Evidence from a longitudinal study. *Journal of Personality and Social Psychology*, *73*(5), 1052–1063.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, *56*, 453–484.

Catley, D., Kaell, A. T., Kirschbaum, C., & Stone, A. A. (2000). A naturalistic evaluation of cortisol secretion in persons with fibromyalgia and rheumatoid arthritis. *Arthritis Care Research*, *13*(1), 51–61.

Cervone, D., & Pervin, L. (2007). *Personality: Theory and research* (Tenth ed.). Hoboken, N.J.: Wiley.

Chamberlin, T. C. (1965). The method of multiple working hypotheses. *Science, New Series*, *148*, 754–759.

Chapman, B. P., Duberstein, P. R., & Lyness, J. M. (2007). The distressed personality type: replicability and general health associations. *European Journal of Personality*, *21*(7), 911–929.

Chapman, M. (1988). *Constructive evolution: Origins and development of Piaget's thought*. Cambridge: Cambridge University Press.

Chard. (1995). *An introduction to radioimmunoassay and related techniques*. Amsterdam: Elsevier Science.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods Research*, *29*(4), 468–508.

Chiolero, A., Wietlisbach, V., Ruffieux, C., Paccaud, F., & Cornuz, J. (2006). Clustering of risk behaviors with cigarette consumption: A population-based survey. *Preventive Medicine*, *42*(5), 348–353.

Christensen, A. J., Ehlers, S. L., Wiebe, J. S., Moran, P. J., Raichle, K., Ferneyhough, K., et al. (2002). Patient personality and mortality: a 4-year prospective examination of chronic renal insufficiency. *Health Psychology*, *21*(4), 315–320.

Christo, G., Jones, S. L., Haylett, S., Stephenson, G. M., Lefever, R. M., & Lefever, R. (2003). The shorter PROMIS questionnaire: Further validation of a tool for simultaneous assessment of multiple addictive behaviours. *Addictive Behaviors*, *28*(2), 225–248.

Clara, I. P., Cox, B. J., & Enns, M. W. (2003). Hierarchical models of personality and psychopathology: the case of self-criticism, neuroticism, and depression. *Personality and Individual Differences*, *35*(1), 91–99.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, *18*(1), 115–126.

Cloninger, C. R., Przybeck, T. R., & Svrakic, D. M. (1991). The Tridimensional Personality Questionnaire: U.S. normative data. *Psychological Reports*, *69*(3 Pt 1), 1047–1057.

Cloninger, C. R., Svrakic, D. M., & Przybeck, T. R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry*, *50*(12), 975–990.

Clow, A., Thorn, L., Evans, P., & Hucklebridge, F. (2004). The awakening cortisol response: Methodological issues and significance. *Stress*, *7*(1), 29–37.

Cohen, R. K., & Swerdlik, M. K. (1998). *Psychological testing and assessment: An introduction to tests and measurement* (Fourth ed.). Mountain View, California: Mayfield Publishing Company.

Cohen, S., Doyle, W. J., & Baum, A. (2006). socio-economic status is associated with stress hormones. *Psychosomatic Medicine*, *68*(3), 414–420.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385–396.

Costa, P. T., Herbst, J. H., McCrae, R. R., Samuels, J., & Ozer, D. J. (2002). The replicability and utility of three personality types. *European Journal of Personality*, *16*(S1), S73–S87.

Costa, P. T., & McCrae, R. R. (1987). Neuroticism, somatic complaints, and disease: Is the bark worse than the bite? *Journal of Personality*, *55*(2), 299–316.

Costa, P. T., & McCrae, R. R. (1992a). NEO PI-R professional manual [Computer software manual]. Odessa, FL: Psychological Assessment Resources, Inc.

Costa, P. T., & McCrae, R. R. (1992b). Reply to Ben-Porath and Waller. *Journal of Personality Assessment*, *4*, 20–22.

Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO personality inventory. *Personality and Individual Differences*, *12*(9), 887–898.

Cox, B. D. (1988). *Health and lifestyle survey, 1984-1985 (HALS1).* Colchester, Essex: UK Data Archive [distributor].

Cox, B. D. (1995). *Health and lifestyle survey : Seven year follow-up, 1991-1992 (HALS2).* Colchester, Essex: UK Data Archive [distributor].

Cox, B. D. (2005). *Health and lifestyle survey (HALS) deaths data : Sixth listing (fifth update), June 2005 [computer file]* (Sixth ed.). Colchester, Essex: UK Data Archive [distributor].

Crawley, M. J. (2007). *The R book*. Chichester: Wiley.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *55*, 281–302.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*(3), 391–418.

Csikszentmihalyi, M. (2002). *Flow : the classic work on how to achieve happiness*. London: Rider & Co.

Cummings, J. H., & Bingham, S. A. (1998). Fortnightly review. diet and the prevention of cancer. *British Medical Journal*, *317*(7173), 1636–1640.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*(4), 529–569.

Dagenais, G. R., Yi, Q., Mann, J. F., Bosch, J., Pogue, J., & Yusuf, S. (2005). Prognostic impact of body weight and abdominal obesity in women and men with cardiovascular disease. *American Heart Journal*, *149*(1), 54–60.

Dahlgren, A., Kecklund, G., & Akerstedt, T. (2005). Different levels of work-related stress and the effects on sleep, fatigue and cortisol. *Scandinavian Journal of Work, Environment & Health*, *31*(4), 277–285.

Dani, J. A., Ji, D., & Zhou, F.-M. (2001). Synaptic plasticity and nicotine addiction. *Neuron*, *31*(3), 349–352.

Daruna, J. H. (1996). Neuroticism predicts normal variability in the number of circulating leucocytes. *Personality and Individual Differences*, *20*(1), 103–108.

de Raad, B., & Goldberg, L. R. (2002). Big five factor assessment: Introduction. In B. de Raad & M. Perugini (Eds.), *Big five assessment* (pp. 1–26). Gottingen: Hogrefe & Huber Publishing.

Dean, K., & Salem, N. (1998). Detecting measurement confounding in epidemiological research: construct validity in scaling risk behaviours: based on a population sample in Minnesota, USA. *Journal of Epidemiology and Community Health*, *52*(3), 195–199.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. Oxford: Oxford University Press, USA.

Deary, I. J., Blenkin, H., Agius, R. M., Endler, N. S., Zealley, H., & Wood, R. (1996). Models of job-related stress and personal achievement among consultant doctors. *British Journal of Psychology*, *86*(1), 3–29.

Deary, I. J., Clyde, Z., & Frier, B. M. (1997). Constructs and models in health psychology: The case of personality and illness reporting in diabetes mellitus. *British Journal of Health Psychology*, *2*(1), 35–54.

Deary, I. J., & Hettema, P. J. (1993). *Foundations of personality* (I. J. Deary & P. J. Hettema, Eds.). Dordrecht: Kluwer.

De Fruyt, F., Van Leeuwen, K., Bagby, R. M., Rolland, J. P., & Rouillon, F. (2006). Assessing and interpreting personality change and continuity in patients treated for major depression. *Psychological Assessment*, *18*(1), 71–80.

Del Boca, F. K., & Darkes, J. (2003). The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction*, *98*(Suppl. 2), 1–12.

Demitrack, M. A., Dale, J. K., Straus, S. E., Laue, L., Listwak, S. J., Kruesi, M. J. P., et al. (1991). Evidence for impaired activation of the hypothalamic-pituitary-adrenal axis in patients with chronic fatigue syndrome. *Journal of Clinical Endocrinology and Metabolism*, *73*(6), 1224–1234.

Denollet, J., Sys, S. U., & Brutsaert, D. L. (1995). Personality and mortality after myocardial infarction. *Psychosomatic Medicine*, *57*(6), 582–591.

De Weerth, C., Zijl, R. H., & Buitelaar, J. K. (2003). Development of cortisol circadian rhythm in infancy. *Early Human Development*, *73*(1-2), 39–52.

Diamantopoulos, Adamantios, Siguaw, & Judy, A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, *17*(4), 263–282.

Digman, J. M. (1989). Five robust trait dimensions: Development, stability, and utility. *Journal of Personality*, *57*(2), 195–214.

Dillard, A. (1990). *The writing life*. Harper Perennial.

Doll, R., Peto, R., Boreham, J., & Sutherland, I. (2004). Mortality in relation to smoking: 50 years' observations on male british doctors. *British Medical Journal*, *328*(7455), 1519–1519.

Donohew, L., Zimmerman, R., Cupp, P. S., Novak, S., Colon, S., & Abell, R. (2000). Sensation seeking, impulsive decision-making, and risky sex: implications for risk-taking and design of interventions. *Personality and Individual Differences*, *28*(6), 1079–1091.

Donovan, J. E., & Jessor, R. (1985). Structure of problem behavior in adolescence and young adulthood. *Journal of Consulting and Clinical Psychology*, *53*(6), 890–904.

Douglas, R. M., Hemila, H., Chalker, E., & Treacy, B. (2007). Vitamin C for preventing and treating the common cold. *Cochrane Database of Systematic Reviews*(3).

Dowd, J. B., & Goldman, N. (2006). Do biomarkers of stress mediate the relation between socio-economic status and health? *Journal of Epidemiology and Community Health*, *60*(7), 633–639.

Drossman, D. A., Leserman, J., Li, Z., Keefe, F., Hu, Y. J. B., & Toomey, T. C. (2000). Effects of coping on health outcome among women with gastrointestinal disorders. *Psychosomatic Medicine*, *62*(3), 309–317.

Du, L., Bakish, D., Ravindran, A. V., & Hrdina, P. D. (2002). Does fluoxetine influence major depression by modifying five-factor personality traits? *Journal of Affective Disorders*, *71*(1-3), 235–241.

Duffy, J. C. (1995). Alcohol consumption and all-cause mortality. *International Journal of Epidemiology*, *24*(1), 100–105.

Ebstein, R. P. (2006). The molecular genetic architecture of human personality: beyond self-report questionnaires. *Molecular Psychiatry*, *11*, 427–445.

Edwards, R. (2004). The problem of tobacco smoking. *British Medical Journal*, *328*(7433), 217–219.

Edwards, S., Clow, A., Evans, P., & Hucklebridge, F. (2001). Exploration of the awakening cortisol response in relation to diurnal cortisol secretory activity. *Life Sciences*, *68*(18), 2093–2103.

Edwards, S., Hucklebridge, F., Clow, A., & Evans, P. (2003). Components of the diurnal cortisol cycle in relation to upper respiratory symptoms and perceived stress. *Psychosomatic Medicine*, *65*(2), 320–327.

Egan, V., Charlesworth, P., Richardson, C., Blair, M., & McMurran, M. (2001). Sensational interests and sensation seeking in mentally disordered offenders. *Personality and Individual Differences*, *30*(6), 995–1007.

Egan, V., Deary, I., & Austin, E. (2000). The NEO-FFI: emerging British norms and an item-level analysis suggest N, A and C are more reliable than O and E. *Personality and Individual Differences*, *29*(5), 907–920.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates.

Endler, N. S., Denisoff, E., & Rutherford, A. (1998). Anxiety and depression: Evidence for the differentiation of commonly cooccurring constructs. *Journal of Psychopathology and Behavioral Assessment*, *20*(2), 149–171.

Evans, P., Hucklebridge, F., Clow, A., Hucklebridge, & Angela, C. (2000). *Mind, immunity and health: The science of psychoneuroimmunology*. Free Association Books.

Eysenck, H. (2006). *The biological basis of personality*. London: Transaction Publishers.

Eysenck, H., & Eysenck, S. (1964). *The Eysenck Personality Inventory*. London: University of London Press.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences*, *13*(6), 667–673.

Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology*, *73*(6), 1224–1237.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. New York: Springer.

Eysenck, H. J., & Eysenck, S. B. G. (1975). Manual of the Eysenck Personality Questionnaire (Junior and Adult).

Farah, M. J. (2005). Neuroethics: the practical and the philosophical. *Trends in Cognitive Sciences*, *9*(1), 34–40.

Farah, M. J., Illes, J., Cook-Deegan, R., Gardner, H., Kandel, E., King, P., et al. (2004). Neurocognitive enhancement: what can we do and what should we do? *Nature Reviews Neuroscience*, *5*(5), 421–425.

Fenech, M., Shyong, & Gillies, P. J. (2007). Nutrition and genome health. *Forum Nutrition*, *60*, 49–65.

Fernander, A. F., Patten, C. A., Schroeder, D. R., Stevens, S. R., Eberman, K. M., & Hurt, R. D. (2005). Exploring the association of John Henry active coping and education on smoking behavior and nicotine dependence among Blacks in the USA. *Social Science & Medicine*, *60*(3), 491–500.

Field, A. (2005). *Discovering statistics using spss : (and sex, drugs and rock'n'roll)* (Second ed.). London: Sage Publications Ltd.

Fine, L. J., Philogene, G. S., Gramling, R., Coups, E. J., & Sinha, S. (2004). Prevalence of multiple chronic disease risk factors: 2001 national health interview survey. *American Journal of Preventive Medicine*, *27*(Suppl.), 18–24.

Fisher, E. B. (1997). Two approaches to social support in smoking cessation: commodity model and nondirective support. *Addictive Behaviors*, *26*(2), 819–833.

Fitzpatrick, R. (1994). Global questions and scores. In C. Jenkinson (Ed.), *Measuring health and medical outcomes* (pp. 27–41). Oxford: UCL Press.

Freres, D. R., Gillham, J. E., Reivich, K., & Shatte, A. J. (2002). Preventing depressive symptoms in middle school students: The Penn Resiliency Program. *International Journal of Emergency Mental Health*, *4*(1), 31–40.

Friedman, H. S. (2000). Long-term relations of personality and health: Dynamisms, mechanisms, tropisms. *Journal of Personality*, *68*(6), 1089–1107.

Friedman, H. S., & Booth-Kewley, S. (1987). The "disease-prone personality". a meta-analytic view of the construct. *American Psychologist*, *42*(6), 539–555.

Friedman, H. S., Tucker, J. S., Schwartz, J. E., Martin, L. R., Tomlinson-Keasey, C., Wingard, D. L., et al. (1995). Childhood conscientiousness and longevity: health behaviors and cause of death. *Journal of Personality and Social Psychology*, *68*(4), 696–703.

Friedman, H. S., Tucker, J. S., Tomlinson-Keasey, C., Schwartz, J. E., Wingard, D. L., & Criqui, M. H. (1993). Does childhood personality predict longevity? *Journal of Personality and Social Psychology*, *65*(1), 176–185.

Fujitaka, M., Nomura, S., Sakura, N., Ueda, K., Matuura, R., & Yumiba, C. (2000). Morning and afternoon serum levels of cortisone and cortisol in asthmatic patients. *Clinica Chimica Acta*, *299*(1-2), 101–108.

Funder, D. C., & Funder, D. (2001). *The personality puzzle* (Second ed.). London: W. W. Norton & Company.

Gaab, J., Huster, D., Peisen, R., Engert, V., Heitz, V., Schad, T., et al. (2002). Hypothalamic-pituitary-adrenal axis reactivity in chronic fatigue syndrome and health under psychological, physiological, and pharmacological stimulation. *Psychosomatic Medicine*, *64*(6), 951–962.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gholson, B., & Barker, P. (1985). Kuhn, Lakatos, and Laudan: Applications in the history of Physics and Psychology. *American Psychologist*, *40*(7), 755–769.

Gillham, Jane, Hamilton, John, Freres, Derek, et al. (2006). Preventing depression among early adolescents in the primary care setting: A randomized controlled study of the penn resiliency program. *Journal of Abnormal Child Psychology*, *34*(2), 195–211.

Gillham, J., & Reivich, K. (2004). Cultivating optimism in childhood and adolescence. *Annals of the American Academy of Political and Social Science*, *591*, 146–163.

Glasgow, R. E., Goldstein, M. G., Ockene, J. K., & Pronk, N. P. (2004). Translating what we have learned into practice: Principles and hypotheses for interventions addressing multiple behaviors in primary care. *American Journal of Preventive Medicine*, *27*(2), 88–101.

Glendinning, A. (2002). Self-esteem and smoking in youth — muddying the waters? *Journal of Adolescence*, *25*(4), 415–425.

Goffin, R. D. (2007). Assessing the adequacy of structural equation models: Golden rules and editorial policies. *Personality and Individual Differences*, *42*(5), 831–839.

Goldberg, L. R. (1993). The structure of personality traits: Vertical and horizontal aspects. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 169–188). Washington, DC: American Psychological Association.

Goldberg, L. R. (1998). What is beyond the big five? *Journal of Personality*, *66*(4), 495–524.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In Mervielde, Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in europe, vol. 7* (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, R. C., et al. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96.

Goldberg, L. R., & Strycker, L. A. (2002). Personality traits and eating habits: the assessment of food preferences in a large community sample. *Personality and Individual Differences*, *32*(1), 49–65.

Goodyer, I. M., Park, R. J., Netherton, C. M., & Herbert, J. (2001). Possible role of cortisol and dehydroepiandrosterone in human development and psychopathology. *British Journal of Psychiatry*, *179*(SEPT.), 243–249.

Gordon, L. (1997). *Measuring stress: A guide for health and social scientists*. Oxford: Oxford University Press.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, *59*(2), 93–104.

Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in Psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry*, *160*(4), 636–645.

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

Gould, S. J. (1996). *The mismeasure of man* (Second ed.). London: W. W. Norton & Company.

Gow, A., Whiteman, M., Pattie, A., & Deary, I. (2005). Goldberg's "IPIP" big-five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, *39*(2), 317–329.

Gruenewald, T. L., Kemeny, M. E., Aziz, N., & Fahey, J. L. (2004). Acute threat to the social self: Shame, social self-esteem, and cortisol activity. *Psychosomatic Medicine*, *66*(6), 915–924.

Hagger-Johnson, G. E. (2004). *Personality traits and health behaviours*. Unpublished master's thesis, The University of Edinburgh.

Hagtvet, K. A., & Nasser, F. M. (2004). How well do item parcels represent conceptually defined latent constructs? a two-facet approach. *Structural Equation Modeling*, *11*(2), 168–193.

Hales, C. N., & Barker, D. J. (2001). The thrifty phenotype hypothesis. *British Medical Bulletin*, *60*, 5–20.

Hamer, M. (2006). Exercise and psychobiological processes: implications for the primary prevention of coronary heart disease. *Sports Medicine*, *36*(10), 829–838.

Hampson, S. E., Andrews, J. A., & Barckley, M. (2007). Predictors of the development of elementary-school children's intentions to smoke cigarettes: Hostility, prototypes, and subjective norms. *Nicotine & Tobacco Research*, *9*(7), 751–760.

Hampson, S. E., Andrews, J. A., Barckley, M., Lichtenstein, E., & Lee, M. E. (2006). Personality traits, perceived risk, and risk-reduction behaviors: A further study of smoking and radon. *Health Psychology*, *25*(4), 530–536.

Hampson, S. E., Andrews, J. A., Barckley, M., & Severson, H. H. (2006). Personality predictors of the development of elementary school children's intentions to drink alcohol: The mediating effects of attitudes and subjective norms. *Psychology of Addictive Behaviors*, *20*(3), 288–297.

Hampson, S. E., Dubanoski, J. P., Hamada, W., Marsella, A. J., Matsukawa, J., Suarez, E., et al. (2001). Where are they now? locating former elementary-school students after nearly 40 years for a longitudinal study of personality and health. *Journal of Research in Personality*, *35*(3), 375–387.

Hampson, S. E., & Goldberg, L. R. (2006). A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. *Journal of Personality and Social Psychology*, *91*(4), 763–779.

Hampson, S. E., Goldberg, L. R., Vogt, T. M., & Dubanoski, J. P. (2006). Forty years on: Teachers' assessments of children's personality traits predict self-reported health behaviors and outcomes at midlife. *Health Psychology*, *25*(1).

Hampson, S. E., Goldberg, L. R., Vogt, T. M., & Dubanoski, J. P. (2007). Mechanisms by which childhood personality traits influence adult health status: Educational attainment and healthy behaviors. *Health Psychology*, *26*(1), 121–125.

Hanson, B. S., Isacsson, S.-O., Janzon, L., & Lindell, S.-E. (1990). Social support and quitting smoking for good. Is there an association? Results from the population study, "men born in 1914," Malmo, Sweden. *Addictive Behaviors*, *15*(3), 221–233.

Hatfield, C. F., Herbert, J., Van Someren, E. J. W., Hodges, J. R., & Hastings, M. H. (2004). Disrupted daily activity/rest cycles in relation to daily cortisol rhythms of home-dwelling patients with early alzheimer's dementia. *Brain*, *127*(5), 1061–1074.

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three — testing the theory in structural equation models! *Personality and Individual Differences*, *42*(5), 841–850.

Hayes, V., Morris, J., Wolfe, C., & Morgan, M. (1995). The SF-36 Health Survey Questionnaire: Is it suitable for use with older adults? *Age & Ageing*, *24*(2), 120–125.

Haylett, S. A., Stephenson, G. M., & Lefever, R. M. H. (2004). Covariation in addictive behaviours: A study of addictive orientations using the shorter PROMIS questionnaire. *Addictive Behaviors*, *29*(1), 61–71.

Hemingway, H., Stafford, M., Stansfeld, S., Shipley, M., & Marmot, M. (1997). Is the SF-36 a valid measure of change in population health? results from the Whitehall II study. *British Medical Journal*, *315*(7118), 1273–1279.

Herek, G. M., Garnets, L. D., & Nolen-Hoeksema, S. (2007). Sexual orientation and mental health. *Annual Review of Clinical Psychology*, *3*, 353–375.

Hewson, C. M., Laurent, D., & Vogel, C. M. (1996). Proper methodologies for psychological and sociological studies conducted via the internet. *Behavior Research Methods, Instruments, and Computers*, *28*(2), 186–191.

Higgins, J., & Green, S. (2006). *Cochrane handbook for systematic reviews of interventions (version 4.2.6)* (J. Higgins & S. Green, Eds.). Chichester, UK: John Wiley & Sons, Ltd.

Hogan, J., & Roberts, B. W. (1996a). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*, *17*(6), 627–637.

Hogan, J., & Roberts, B. W. (1996b). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*, *17*(6), 627–637.

Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance*, *18*(4), 331–341.

Holbrook, A. L., Krosnick, J. A., Moore, D., & Tourangeau, R. (2007). Response order effects in dichotomous categorical questions presented orally: The impact of question and respondent attributes. *Public Opinion Quarterly*, *71*(3), 325–348.

Holmes, S. (2006). Nutrition and the prevention of cancer. *Journal of Family Planning and Reproductive Health Care*, *16*(2), 43–46.

Home Office. (1999). *The health of minority ethnic groups, health survey for england.* London: DoH.

Home Office. (2004). *Choosing health.* London: HMO.

Horney, K. (1993). *Our inner conflicts: A constructive theory of neurosis*. London: W. W. Norton & Company.

Hough, L. M., & Oswald, F. L. (2005). They're right, well... mostly right: Research evidence and an agenda to rescue personality testing from 1960s insights. *Human Performance*, *18*(4), 373–387.

Houtman, I. L. D., & Bakker, F. C. (1991). Individual differences in reactivity to and coping with the stress of lecturing. *Journal of Psychosomatic Research*, *35*(1), 11–24.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods*, *12*(2), 205–218.

Hoyle, R. H. (1999). *Statistical strategies for small sample research*. Thousand Oaks, CA: Sage Publications, Inc.

Hoyle, R. H., Fejfar, M. C., & Miller, J. D. (2000). Personality and sexual risk taking: a quantitative review. *Journal of Personality*, *68*(6), 1203–1231.

Hruschka, D. J., Kohrt, B. A., & Worthman, C. M. (2005). Estimating between- and within-individual variation in cortisol levels using multilevel models. *Psychoneuroendocrinology*, *30*(7), 698–714.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453.

Huovinen, E., Kaprio, J., & Koskenvuo, M. (2001). Asthma in relation to personality traits, life satisfaction, and stress: A prospective study among 11 000 adults. *Allergy*, *56*(10), 971–977.

Huppert, F. A., & Whittington, J. E. (1995). Symptoms of psychological distress predict 7-year mortality. *Psychological Medicine*, *25*(5), 1073–1086.

Iglesias, S. L., Azzara, S., Squillace, M., Jeifetz, M., Lores, M. R., Desimone, M. F., et al. (2005). A study on the effectiveness of a stress management programme for college students. *Pharmacy Education*, *5*(1), 27–31.

Ignarro, L. J., Balestrieri, M. L., & Napoli, C. (2007). Nutrition, physical activity, and cardiovascular disease: An update. *Cardiovascular Research*, *73*(2), 326–340.

Iwasa, H., Masui, Y., Gondo, Y., Inagaki, H., Kawaai, C., & Suzuki, T. (2008). Personality and all-cause mortality among older adults dwelling in a japanese community: A five-year population-based prospective cohort study. *American Journal of Geriatric Psychiatry*, *16*(5), 399–405.

Jarvis, M. J., & Wardle, J. (2006). Social patterning of individual health behaviours: the case of cigarette smoking. In M. Marmot & R. G. Wilkinson (Eds.), *Social determinants of health* (Second ed., pp. 224–237). Oxford: Oxford University Press.

Jenkinson, C., Bardsley, M., & Lawrence, K. (1994). Measurement in subjective health assessment: Themes and prospects. In C. Jenkinson (Ed.), *Measuring health and medical outcomes* (pp. 176–185). Oxford: UCL Press.

Jenkinson, C., Coulter, A., & Wright, L. (1993). Short form 36 (SF 36) health survey questionnaire: Normative data for adults of working age. *British Medical Journal*, *306*(6890), 1437–1440.

Jenkinson, C., Layte, R., & Lawrence, K. (1997). Development and testing of the Medical Outcomes Study 36-Item Short Form Health Survey Summary Scale scores in the United Kingdom: Results from a large-scale survey and a clinical trial. *Medical Care*, *35*(4), 410–416.

Jenkinson, C., Layte, R., Wright, L., & Coulter, A. (1996). *The U.K. SF-36: An analysis and interpretation manual. A guide to health status measurement with particular reference to the Short Form 36 Health Survey*. Oxford: Health Services Research Unit, Department of Public Health and Primary Care, University of Oxford.

Jenkinson, C., Stewart-Brown, S., Petersen, S., & Paice, C. (1999). Assessment of the SF-36 version 2 in the United Kingdom. *Journal of Epidemiology and Community Health*, *53*(1), 46–50.

Johnson, E. O., Kamilaris, T. C., Chrousos, G. P., & Gold, P. W. (1992). Mechanisms of stress: A dynamic overview of hormonal and behavioral homeostasis. *Neuroscience & Biobehavioral Reviews*, *16*(2), 115–130.

Johnson, W., & Krueger, R. F. (2005). Higher perceived life control decreases genetic variance in physical health: Evidence from a national twin study. *Journal of Personality and Social Psychology*, *88*(1), 165–173.

Jonassaint, C. R., Boyle, S. H., Williams, R. B., Mark, D. B., Siegler, I. C., & Barefoot, J. C. (2007). Facets of openness predict mortality in patients with cardiac disease. *Psychosomatic Medicine*, *69*(4), 319–322.

Kalman, D. (2002). The subjective effects of nicotine: methodological issues, a review of experimental studies, and recommendations for future research. *Nicotine & Tobacco Research*, *4*(1), 25–70.

Kendler, K. S., Kuhn, J., & Prescott, C. A. (2004). The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *American Journal of Psychiatry*, *161*(4), 631–636.

Kessler, D., Bennewith, O., Lewis, G., & Sharp, D. (2002). Detection of depression and anxiety in primary care: follow up study. *British Medical Journal*, *325*(7371), 1016–1017.

Kimmel, A. J. (1996). *Ethical issues in behavioral research: A survey*. Cambridge: Wiley-Blackwell.

Kirkwood, B., & Sterne, J. (2001). *Essentials of medical statistics*. Oxford: Blackwell Publishing Limited.

Kirschbaum, C., & Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. *Psychoneuroendocrinology*, *19*(4), 313–333.

Kirschbaum, C., Prussner, J. C., Stone, A. A., Federenko, I., Gaab, J., Lintz, D., et al. (1995). Persistent high cortisol responses to repeated psychological stress in a subpopulation of healthy men. *Psychosomatic Medicine*, *57*(5), 468–474.

Kirschbaum, C., Steyer, R., Eid, M., Patalla, U., Schwenkmezger, P., & Hellhammer, D. H. (1990). Cortisol and behavior: 2. Application of a latent state-trait model to salivary cortisol. *Psychoneuroendocrinology*, *15*(4), 297–307.

Kline, P. (2000). *New psychometrics: Science, psychology and measurement*. London: Routledge.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (Second ed.). New York; London: The Guilford Press.

Kloner, R. A., & Rezkalla, S. H. (2007). To drink or not to drink? that is the question. *Circulation*, *116*(11), 1306–1317.

Knight, R. B., Atkins, A., & Eagle, C. J. (1979). Psychological stress, ego defenses, and cortisol production in children hospitalized for elective surgery. *Psychosomatic Medicine*, *41*(1), 40–49.

Knutson, B., Wolkowitz, O. M., Cole, S. W., Chan, T., Moore, E. A., Johnson, R. C., et al. (1998). Selective alteration of personality and social behavior by serotonergic intervention. *American Journal of Psychiatry*, *155*(3), 373–379.

Korotkov, D., & Hannah, T. E. (2004). The five-factor model of personality: Strengths and limitations in predicting health status, sick-role and illness behaviour. *Personality and Individual Differences*, *36*(1), 187–199.

Korte, S. M., Koolhaas, J. M., Wingfield, J. C., & McEwen, B. S. (2005). The Darwinian concept of stress: Benefits of allostasis and costs of allostatic load and the trade-offs in health and disease. *Neuroscience & Biobehavioral Reviews*, *29*(1), 3–38.

Korten, A. E., Jorm, A. F., Jiao, Z., Letenneur, L., Jacomb, P. A., Henderson, A. S., et al. (1999). Health, cognitive, and psychosocial factors as predictors of mortality in an elderly community sample. *Journal of Epidemiology and Community Health*, *53*(2), 83–88.

Kreuter, M. W., & Skinner, C. S. (2000). Tailoring: What's in a name? *Health Education Research*, *15*(1), 1–4.

Kristenson, M., Eriksen, H. R., Sluiter, J. K., Starke, D., & Ursin, H. (2004). Psychobiological mechanisms of socio-economic differences in health. *Social Science & Medicine*, *58*(8), 1511–1522.

Krueger, R. F., Caspi, A., & Moffitt, T. E. (2000). Epidemiological personology: The unifying role of personality in population-based research on problem behaviors. *Journal of Personality*, *68*(6), 967–998.

Kudielka, B. M., Buske-Kirschbaum, A., Hellhammer, D. H., & Kirschbaum, C. (2004). HPA axis responses to laboratory psychosocial stress in healthy elderly adults, younger adults, and children: impact of age and gender. *Psychoneuroendocrinology*, *29*(1), 83–98.

Kudielka, B. M., & Kirschbaum, C. (2003). Awakening cortisol responses are influenced by health status and awakening time but not by menstrual cycle phase. *Psychoneuroendocrinology*, *28*(1), 35–47.

Kuhn, T. S. (1996). *The structure of scientific revolutions*. Chicago: University Of Chicago Press.

Kuller, L. H. (2006). Nutrition, lipids, and cardiovascular disease. *Nutrition Reviews*, *64*(2 Pt 2).

Kunz-Ebrecht, S. R., Kirschbaum, C., & Steptoe, A. (2004). Work stress, socioeconomic status and neuroendocrine activation over the working day. *Social Science & Medicine*, *58*(8), 1523–1530.

Kuo, P.-H., Neale, M. C., Riley, B. P., Patterson, D. G., Walsh, D., Prescott, C. A., et al. (2007). A genome-wide linkage analysis for the personality trait neuroticism in the irish affected sib-pair study of alcohol dependence. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *144B*(4), 463–468.

Kypri, K., & McAnally, H. M. (2005). Randomized controlled trial of a web-based primary care intervention for multiple health risk behaviors. *Preventive Medicine*, *41*(3-4), 761–766.

Laaksonen, M., Lahelma, E., & Prttl, R. (2002). Associations among health-related behaviours: Sociodemographic variation in finland. *Sozial- und Prventivmedizin/Social and Preventive Medicine*, *47*(4), 225–232.

Laaksonen, M., Prattala, R., Helasoja, V., Uutela, A., & Lahelma, E. (2003). Income and health behaviours. evidence from monitoring surveys among finnish adults. *Journal of Epidemiology and Community Health*, *57*(9), 711–717.

Langdridge, D. (2003). *Introduction to research methods & data analysis in psychology*. Prentice Hall (UK).

Larsen, K. (2005). The cox proportional hazards model with a continuous latent variable measured by multiple binary indicators. *Biometrics*, *61*(4), 1049–1055.

LeBlanc, J., & Ducharme, M. B. (2005). Influence of personality traits on plasma levels of cortisol and cholesterol. *Physiology & Behavior*, *84*(5), 677–680.

Lee, K., Ogunfowora, B., & Ashton, M. C. (2005). Personality traits beyond the big five: Are they within the hexaco space? *Journal of Personality*, *73*(5), 1437–1463.

Leproult, R., Copinschi, G., Buxton, O., & Van Cauter, E. (1997). Sleep loss results in an elevation of cortisol levels the next evening. *Sleep*, *20*(10), 865–870.

Leventhal, H., Weinman, J., Leventhal, E. A. A., & Alison. (2008). Health psychology: The search for pathways between behavior and health. *Annual Review of Psychology*, *59*(1), 477–505.

Li, I., Chiou, H.-H., & Shen, P.-S. (2007). Correlations between cortisol level and internalizing disposition of young children are increased by selecting optimal sampling times and aggregating data. *Developmental Psychobiology*, *49*(6), 633–639.

Lilienfeld, D. E., & Stolley, P. D. (1994). *Foundations of epidemiology*. Oxford: Oxford University Press.

Lipkus, I. M., Barefoot, J. C., Williams, R. B., & Siegler, I. C. (1994). Personality measures as predictors of smoking initiation and cessation in the UNC Alumni Heart Study. *Health Psychology*, *13*(2), 149–155.

Lloyd, A., Jenkinson, C., & Stewart-Brown, S. (1999). Assessment of the SF-36 version 2 in the United Kingdom (multiple letters). *Journal of Epidemiology and Community Health*, *53*(10), 651–652.

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis*. London: Lawrence Erlbaum Associates.

Londen, L. van, Goekoop, J. G., Kempen, G. M. van, Frankhuijzen-Sierevogel, A. C., Wiegant, V. M., Velde, E. A. van der, et al. (1997). Plasma levels of arginine vasopressin elevated in patients with major depression. *Neuropsychopharmacology*, *17*(4), 284–292.

Luoto, R., Prättälä, R., Uutela, A., & Puska, P. (1998). Impact of unhealthy behaviors on cardiovascular mortality in Finland, 1978-1993. *Preventive Medicine*, *27*(1), 93–100.

Machale, S. M., Cavanagh, J. T. O., Bennie, J., Carroll, S., Goodwin, G. M., & Lawrie, S. M. (1998). Diurnal variation of adrenocortical activity in chronic fatigue syndrome. *Neuropsychobiology*, *38*(4), 213–217.

Mackinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000, December). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181.

Maddox, G. L., & Douglass, E. B. (1973). Self-assessment of health: a longitudinal study of elderly subjects. *Journal of Health and Social Behavior*, *14*(1), 87–93.

Maier, H., & Smith, J. (1999). Psychological predictors of mortality in old age. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *54*(1), P44–54.

Markland, D. (2007). The golden rule is that there are no golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences*, *42*(5), 851–858.

Marklund, N., Peltonen, M., Nilsson, T. K., & Olsson, T. (2004). Low and high circulating cortisol levels predict mortality and cognitive dysfunction early after stroke. *Journal of Internal Medicine (GBR)*, *256*(1), 15–21.

Marsh, H. W., & Hau, K. T. (1999). Confirmatory factor analysis: Strategies for small sample sizes. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 252–287). Thousand Oaks, CA: Sage Publications Ltd.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing hu and bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320–341.

Martin, J. L. (2001). The authoritarian personality, 50 years later: What questions are there for political psychology? *Political Psychology*, *22*(1), 1–26.

Martin, L. R., Friedman, H. S., & Schwartz, J. E. (2007). Personality and mortality risk across the life span: The importance of conscientiousness as a biopsychosocial attribute. *Health Psychology*, *26*(4), 428–436.

Martin, L. R., Friedman, H. S., Tucker, J. S., Tomlinson-Keasey, C., Criqui, M. H., & Schwartz, J. E. (2002). A life course perspective on childhood cheerfulness and its relation to mortality risk. *Personality and Social Psychology Bulletin*, *28*(9), 1155–1165.

Martin, P., & Ashcroft, R. (2005). *Neuroscience, ethics and society: a review of the field. Background paper prepared for the 2005 Wellcome Trust Summer School on "neuroethics"*. The Wellcome Trust.

Martin, P. R. (1998). *The sickening mind*. London: Flamingo.

Masson, L. F., McNeill, G., Tomany, J. O., Simpson, J. A., Peace, H. S., Wei, L., et al. (2003). Statistical approaches for assessing the relative validity of a food-frequency questionnaire: Use of correlation coefficients and the kappa statistic. *Public Health Nutrition*, *6*(3), 313–321.

Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits* (Second ed.). Cambridge: Cambridge University Press.

Matthews, G., Yousfi, S., Schmidt-Rathjens, C., & Amelang, M. (2003). Personality variable differences between disease clusters. *European Journal of Personality*, *17*(2), 157–177.

McCain, G. A., & Tilbe, K. S. (1989). Diurnal hormone variation in fibromyalgia syndrome: A comparison with rheumatoid arthritis. *Journal of Rheumatology*, *16*(Suppl. 19), 154–157.

McCrae, R. R., & Costa, P. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). New York: Guildford.

McCrae, R. R., & Costa, P. T. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences*, *6*(5), 587–597.

McCrae, R. R., Costa, P. T., Hrebickova, M., Ostendorf, F., Angleitner, A., Avia, M. D., et al. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, *78*(1), 173–186.

McCrae, R. R., Costa, P. T., Lima, M. P. de, Simoes, A., Ostendorf, F., Angleitner, A., et al. (1999). Age differences in personality across the adult life span: Parallels in five cultures,. *Developmental Psychology*, *35*(2), 466–477.

McCrae, R. R., Costa, P. T., Ostendorf, F., Angleitner, A., Hrebickova, M., Avia, M. D., et al. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, *78*(1), 173–186.

McCrae, R. R., Costa, P. T., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., et al. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional, and cross-cultural analyses,. *Journal of Personality and Social Psychology*, *83*(6), 1456–1468.

McCrae, R. R., Costa, P. T., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., et al. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional, and cross-cultural analyses. *Journal of Personality and Social Psychology*, *83*(6), 1456–1468.

McCue, P., Buchanan, T., & Martin, C. (2006). Screening for psychological distress using internet administration of the Hospital Anxiety and Depression Scale (HADS) in individuals with chronic fatigue syndrome. *British Journal of Clinical Psychology*, *45*(4), 483–498.

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum Associates.

McEwan, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, *840*(1), 33–44.

McEwen, B. S., Biron, C. A., Brunson, K. W., Bulloch, K., Chambers, W. H., Dhabhar, F. S., et al. (1997). The role of adrenocorticoids as modulators of immune function in health and disease: neural, endocrine and immune interactions. *Brain Research Reviews*, *23*(1-2), 79–133.

McIntosh, C. N. (2007). Rethinking fit assessment in structural equation modelling: A commentary and elaboration on barrett (2007). *Personality and Individual Differences*, *42*(5), 859–867.

Mechanic, D., & Cleary, P. D. (1980). Factors associated with the maintenance of positive health behavior. *Preventive Medicine*, *9*(6), 805–814.

Meng, L., Maskarinec, G., Lee, J., & Kolonel, L. N. (1999). Lifestyle factors and chronic diseases: Application of a composite risk index. *Preventive Medicine*, *29*, 296–304.

Mershon, B., & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology*, *55*(4), 675–680.

Metzner, H. L., Carman, W. J., & House, J. (1983). Health practices, risk factors, and chronic disease in tecumseh. *Preventive Medicine*, *12*(4), 491–507.

Miles, J. (2003). A framework for power analysis using a structural equation modelling procedure. *BMC Medical Research Methodology*, *3*(1).

Miles, J., & Shevlin, M. (2000). *Applying regression and correlation: A guide for students and researchers*. London: Sage Publications Ltd.

Miles, J., & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and Individual Differences*, *42*(5), 869–874.

Miller, T. Q., Smith, T. W., Turner, C. W., Guijarro, M. L., & Hallet, A. J. (1996). A meta-analytic review of research on hostility and physical health. *Psychological Bulletin*, *119*(2), 322–348.

Mink, P. J., Goodman, M., Barraj, L. M., Imrey, H., Kelsh, M. A., & Yager, J. (2004). Evaluation of uncontrolled confounding in studies of environmental exposures and neurobehavioral testing in children. *Epidemiology*, *15*(4), 385–393.

Mischel, W. (1968). *Personality and assessment*. New York: John Wiley & Sons, Inc.

Mokdad, A. H., Marks, J. S., Stroup, D. F., & Gerberding, J. L. (2004). Actual causes of death in the united states, 2000. *JAMA*, *291*(10), 1238–1245.

Monden, C., & Kraaykamp, G. (2006). Neuroticism, education and self-assessed health in the general population of the united states. can smoking behaviour explain the associations? *Social Indicators Research*, *78*(2), 271–285.

Morris, R., & Carstairs, V. (1991). Which deprivation? a comparison of selected deprivation indexes. *Journal of Public Health Medicine*, *13*(4), 318–326.

Munafo, M. R., Clark, T. G., Moore, L. R., Payne, E., Walton, R., & Flint, J. (2003). Genetic polymorphisms and personality in healthy adults: A systematic review and meta-analysis. *Molecular Psychiatry*, *8*(5), 471–484.

Murberg, T. A., Bru, E., & Aarsland, T. (2001). Personality as predictor of mortality among patients with congestive heart failure: A two-year follow-up study. *Personality and Individual Differences*, *30*(5), 749–757.

Murray, C. J. L., & Lopez, A. D. (2006). *Global burden of disease and risk factors*. New York, N.Y.; Washington, DC: Oxford University Press; World Bank Publications.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*(1), 81–117.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*(3), 431–462.

Muthen, B. O. (1992). Latent variable modeling in epidemiology. *Alcohol Health and Research World*, *16*(4), 286–292.

Muthén, B. O., & Muthén, L. K. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620.

Muthén, L., & Muthén, B. (1998–2007). *Mplus user's guide* (Fifth ed.). Los Angeles, CA: Muthén and Muthén.

Myrtek, M. (2001). Meta-analyses of prospective studies on coronary heart disease, type a personality, and hostility. *International Journal of Cardiology*, *79*(2-3), 245–251.

Nakaya, N., Tsubono, Y., Hosokawa, T., Nishino, Y., Ohkubo, T., Hozawa, A., et al. (2003). Personality and the risk of cancer. *Journal of the National Cancer Institute*, *95*(11), 799–805.

Nelson, L. D., Drebing, C., Satz, P., & Uchiyama, C. (1998). Personality change in head trauma: A validity study of the neuropsychology behavior and affect profile. *Archives of Clinical Neuropsychology*, *13*(6), 549–560.

Ness, A. R., & Powles, J. W. (1997). Fruit and vegetables, and cardiovascular disease: A review. *International Journal of Epidemiology*, *26*(1), 1–13.

Nicolson, N. A., & Van Diest, R. (2000). Salivary cortisol patterns in vital exhaustion. *Journal of Psychosomatic Research*, *49*(5), 335–342.

Noar, S. M., Benac, C. N., & Harris, M. S. (2007). Does tailoring matter? meta-analytic review of tailored print health behavior change interventions. *Psychological Bulletin*, *133*(4), 673–693.

O'Cleirigh, C., Ironson, G., Weiss, A., & Costa, P. T. (2007). Conscientiousness predicts disease progression (CD4 number and viral load) in people living with HIV. *Health Psychology*, *26*(4), 473–480.

O'Connor, M. C., & Paunonen, S. V. (2007). Big five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*(5), 971–990.

Office of National Statistics. (2001). *Census 2001.* London: Office of National Statistics.

Ogden, J. (2007). *Health psychology*. Open University Press.

O'Keefe, J. H., Bybee, K. A., & Lavie, C. J. (2007). Alcohol and cardiovascular health: the razor-sharp double-edged sword. *Journal of the American College of Cardiology*, *50*(11), 1009–1014.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, *17*(6), 609–626.

Ormel, J., Rosmalen, J., & Farmer, A. (2004). Neuroticism: a non-informative marker of vulnerability to psychopathology. *Social Psychiatry and Psychiatric Epidemiology*, *39*(11), 906–912.

Oswald, L. M., Zandi, P., Nestadt, G., Potash, J. B., Kalaydjian, A. E., & Wand, G. S. (2006). Relationship between cortisol responses to stress and personality. *Neuropsychopharmacology*, *31*(7), 1583–1591.

O'Toole, B. I., & Stankov, L. (1992). Ultimate validity of psychological tests. *Personality and Individual Differences*, *13*(6), 699–716.

Paaunonen, S. V., & Jackson, D. N. (1986). Nomothetic and idiothetic measurement in personality. *Journal of Personality*, *54*(2), 447–459.

Pace-Schott, E. F., & Hobson, J. A. (2002). The neurobiology of sleep: genetics, cellular physiology and subcortical networks. *Nature Reviews Neuroscience*, *3*(8), 591–605.

Palmgreen, P., Donohew, L., Lorch, E. P., Hoyle, R. H., & Stephenson, M. T. (2001). Television campaigns and adolescent marijuana use: Tests of sensation seeking targeting. *American Journal of Public Health*, *91*(2), 292–296.

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., et al. (1995). Physical activity and public health. a recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA*, *273*(5), 402–407.

Patton, D., Barnes, G. E., & Murray, R. P. (1993). Personality characteristics of smokers and ex-smokers. *Personality and Individual Differences*, *15*(6), 653–664.

Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology*, *74*(2), 538–556.

Paunonen, S. V. (2003). Big five factors of personality and replicated predictions of behavior. *Journal of Personality and Social Psychology*, *84*(2), 411–424.

Paunonen, S. V., & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, *81*(3), 524–539.

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the big five? plenty! *Journal of Personality*, *68*(5), 821–835.

Paz-Elizur, T., Krupsky, M., Blumenstein, S., Elinger, D., Schechtman, E., & Livneh, Z. (2003). DNA repair activity for oxidative damage and risk of lung cancer. *Journal of the National Cancer Institute*, *95*(17), 1312–1319.

Peele, S. (1999). *Alcohol and pleasure: a health perspective*. London: Brunner/Mazel.

Pervin, L., Cervone, D., & John, O. (2004). *Personality: Theory and research* (Ninth ed.). Hoboken, N.J.: Wiley.

Phillips, Douglas, C., Burns, & Mark, D. (2005). Neuroticism, cortisol reactivity, and antibody response to vaccination. *Psychophysiology*, *42*(2), 232–238.

Piedmont, R. L. (2001). Cracking the plaster cast: Big five personality change during intensive outpatient counseling. *Journal of Research in Personality*, *35*(4), 500–520.

Piedmont, R. L., & Aycock, W. (2007). An historical analysis of the lexical emergence of the big five personality adjective descriptors. *Personality and Individual Differences*, *42*(6), 1059–1068.

Piedmont, R. L., McCrae, R. R., Rieman, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, *78*(3), 582–593.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*(10), 421–425.

Plat, L., Byrne, M. M., Sturis, J., Polonsky, K. S., Mockel, J., Fery, F., et al. (1996). Effects of morning cortisol elevation on insulin secretion and glucose regulation in humans. *American Journal of Physiology — Endocrinology and Metabolism*, *270*(1), E36–42.

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.

Posthuma, D., Cherny, S. S., & Boomsma, D. I. (2006). Introduction to the special issue: Human linkage studies for behavioral traits. *Behavior Genetics*, *36*(1), 1–3.

Powell, K. E., & Pratt, M. (1996). Physical activity and health. avoiding the short and miserable life. *British Medical Journal*, *313*(7050), 126–127.

Praag, H. M. van, Kloet, E. R. de, & Os, J. van. (2004). *Stress, the brain and depression*. Cambridge: Cambridge University Press.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swifts electric factor analysis machine. *Understanding Statistics*, *2*(1), 13–43.

Preventing depression among early adolescents in the primary care setting: A randomized controlled study of the Penn Resiliency Program. (2006). *Journal of Abnormal Child Psychology*, *34*(2), 195–211.

Pronk, N. P., Peek, C. J., & Goldstein, M. G. (2004). Addressing multiple behavioral risk factors in primary care: A synthesis of current knowledge and stakeholder dialogue sessions. *American Journal of Preventive Medicine*, *27*(2, Suppl. 1), 4–17.

Pruessner, J. C., Gaab, J., Hellhammer, D. H., Lintz, D., Schommer, N., & Kirschbaum, C. (1997). Increasing correlations between personality traits and cortisol stress responses obtained by data aggregation. *Psychoneuroendocrinology*, *22*(8), 615–625.

Renaud, S., & Lorgeril, M. de. (1992). Wine, alcohol, platelets, and the French paradox for coronary heart disease. *The Lancet*, *339*(8808), 1523–1526.

Rhodes, S. D., Bowie, D. A., & Hergenrather, K. C. (2003). Collecting behavioural data using the world wide web: Considerations for researchers. *Journal of Epidemiology and Community Health*, *57*(1), 68–73.

Rimm, E. B., Klatsky, A., Grobbee, D., & Stampfer, M. J. (1996). Review of moderate alcohol consumption and reduced risk of coronary heart disease: is the effect due to beer, wine, or spirits? *British Medical Journal*, *312*(7033), 731–736.

Roberts, B. W., & Bogg, T. (2004). A longitudinal study of the relationships between conscientiousness and the social-environmental factors and substance-use behaviors that influence health. *Journal of Personality*, *72*(2), 325–353.

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: an empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, *58*(1), 103–139.

Robertson, A., Brunner, E., & Sheiham, A. (2006). Food is a political issue. In M. Marmot & R. G. Wilkinson (Eds.), *Social determinants of health* (Second ed., pp. 172–195). Oxford: Oxford University Press.

Rosmond, R., Dallman, M. F., & Bjorntorp, P. (1998). Stress-related cortisol secretion in men: Relationships with abdominal obesity and endocrine, metabolic and hemodynamic abnormalities. *Journal of Clinical Endocrinology & Metabolism*, *83*(6), 1853–1859.

Rowan, K. (1994). Global questions and scores. In C. Jenkinson (Ed.), *Measuring health and medical outcomes* (pp. 54–76). Oxford: UCL Press.

Rust, J., & Lord, W. (2006). *NEO PI-R Manual UK edition (adapted from the NEO PI-R Professional Manual by Paul T. Costa and Robert R. McCrae)*. Oxford: Hogrefe Ltd.

Ryan, H., Wortley, P. M., Easton, A., Pederson, L., & Greenwood, G. (2001). Smoking among lesbians, gays, and bisexuals: a review of the literature. *American Journal of Preventive Medicine*, *21*(2), 142–149.

Ryan, J. M., Corry, J. R., Attewell, R., & Smithson, M. J. (2002). A comparison of an electronic version of the sf-36 general health questionnaire to the standard paper version. *Quality of Life Research*, *11*(1), 19–26.

Saboonchi, F., & Lundh, L. G. (2003). Perfectionism, anger, somatic health, and positive affect. *Personality and Individual Differences*, *35*(7), 1585–1599.

Saucier, G., & Goldberg, L. R. (2001). Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of Personality*, *69*(6), 847–879.

Saucier, G., & Goldberg, L. R. (2002). Assessing the big five: Applications of 10 psychometric criteria to the development of marker scales. In B. de Raad & M. Perugini (Eds.), *Big five assessment* (pp. 29–58). Gottingen: Hogrefe & Huber Publishing.

Schapiro, I. R., Ross-Petersen, L., Salan, H., Garde, K., Olsen, J. H., & Johansen, C. (2003). Extroversion and neuroticism and the associated risk of cancer: A danish cohort study. *American Journal of Epidemiology*, *153*(8), 757–763.

Schmidt, W. C. (1997). World-wide web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers*, *29*, 274–279.

Schmidt-Reinwald, A., Pruessner, J. C., Hellhammer, D. H., Federenko, I., Rohleder, N., Schurmeyer, T. H., et al. (1999). The cortisol response to awakening in relation to different challenge tests and a 12-hour cortisol rhythm. *Life Sciences*, *64*(18), 1653–1660.

Schommer, N. C., Kudielka, B. M., Hellhammer, D. H., & Kirschbaum, C. (1999). No evidence for a close relationship between personality traits and circadian cortisol rhythm or a single cortisol stress response. *Psychological Reports*, *84*(3), 840–842.

Schreiber, J. E., Van Hulle, C. A., Clark, P. L. B., Lemery, K., & Goldsmith, H. H. (2003). Genetic and environmental influence on afternoon and evening cortisol levels. *Behavior Genetics*, *33*(6), 718–718.

Schuit, J. A., Loon van, Tijhuis, M., & Ocke, M. C. (2002). Clustering of lifestyle risk factors in a general adult population. *Preventive Medicine*, *35*(3), 219–224.

Schwartz, P. H. (2005). Defending the distinction between treatment and enhancement. *American Journal of Bioethics*, *5*(3), 17–19.

Scott Long, J. (1983). *Covariance structure models: An introduction to LISREL*. London: Sage.

Segerstrom, S. C., & Miller, G. E. (2004). Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological bulletin*, *130*(4), 601–630.

Segovia, J., Bartlett, R. F., & Edwards, A. C. (1991). Health status and health practices: Alameda and beyond. *International Journal of Epidemiology*, *20*(1), 259–263.

Sephton, S., & Spiegel, D. (2003). Circadian disruption in cancer: a neuroendocrine-immune pathway from stress to disease? *Brain, Behavior, and Immunity*, *17*(5), 321–328.

Sephton, S. E., Sapolsky, R. M., Kraemer, H. C., & Spiegel, D. (2000). Diurnal cortisol rhythm as a predictor of breast cancer survival. *Journal of the National Cancer Institute*, *92*(12), 994–1000.

Shenkin, S. D., Starr, J. M., & Deary, I. J. (2004). Birth weight and cognitive ability in childhood: a systematic review. *Psychological Bulletin*, *130*(6), 989–1013.

Shiner, R. L., Masten, A. S., & Roberts, J. M. (2003). Childhood personality foreshadows adult personality and life outcomes two decades later. *Journal of Personality*, *71*(6), 1145–1170.

Shipley, B., Weiss, A., Der, G., Taylor, M. D. D., & Deary, I. (2007). Neuroticism, extraversion, and mortality in the uk health and lifestyle survey: A 21-year prospective cohort study. *Psychosomatic Medicine*, *69*, 923–931.

Sieber, J. E., & Saks, M. J. (1989). A census of subject pool characteristics and policies. *American Psychologist*, *44*(7), 1053–1061.

Silveira, E., Taft, C., Sundh, V., Waern, M., Palsson, S., & Steen, B. (2005). Performance of the SF-36 Health Survey in screening for depressive and anxiety disorders in an elderly female swedish population. *Quality of Life Research*, *14*(5), 1263–1274.

Silverman, D. (2001). *Interpreting qualitative data: Methods for analysing talk, text and interaction*. Sage Publications Ltd.

Singh-Manoux, A. (2005). Commentary: Modelling multiple pathways to explain social inequalities in health and mortality. *International Journal of Epidemiology*, *34*(3), 638–639.

Skinner, C. S., Strecher, V. J., & Hospers, H. (1994). Physicians' recommendations for mammography: Do tailored messages make a difference? *American Journal of Public Health*, *84*(1), 43–49.

Sluiter, J. K., Frings-Dresen, M. H. W., & Van Der Beek, A. J. (2000). The relation between neuroendocrine reactivity and recovery from different natures of work, subjective need for recovery, and health status. In (pp. 220–223). Coronel Inst. Occup./Environ. Hlth., Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands.

Smith, D. G., Shipley, M. J., Batty, G. D., Morris, J. N., & Marmot, M. (2000). Physical activity and cause-specific mortality in the Whitehall study. *Public Health*, *114*(5), 308–315.

Smith, G. D., Whitley, E., Dorling, D., & Gunnell, D. (2001). Area based measures of social and economic circumstances: cause specific mortality patterns depend on the choice of index. *Journal of Epidemiology and Community Health*, *55*(2), 149–150.

Smith, T. (2006). Personality as risk and resilience in physical health. *Current Directions in Psychological Science*, *15*(5), 227–231.

Smith, T. W., & Williams, P. G. (1992). Personality and health: advantages and limitations of the five-factor model. *Journal of Personality*, *60*(2), 395–423.

Soanes, C., & Stevenson, A. (2005). *Oxford dictionary of English* (Second ed.). Oxford: Oxford University Press.

Solari, A. (2005). Role of health-related quality of life measures in the routine care of people with multiple sclerosis. *Health and Quality of Life Outcomes*, *3*(1).

Spiegel, D., & Sephton, S. E. (2001). Psychoneuroimmune and endocrine pathways in cancer: effects of stress and support. *Seminars in Clinical Neuropsychiatry*, *6*(4), 252–265.

Spirduso, W. W., & Cronin, D. L. (2001). Exercise dose-response effects on quality of life and independent living in older adults. *Medicine & Science in Sports & Exercise*, *33*(6, Suppl.).

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, *84*(5), 1041–1053.

Stafford, M., & Mccarthy, M. (2006). Neighbourhoods, housing and health. In M. Marmot & R. G. Wilkinson (Eds.), *Social determinants of health* (Second ed., pp. 297–317). Oxford: Oxford University Press.

Stansfeld, S. A. (2002a). Commentary: The problem with stress: minds, hearts and disease. *International Journal of Epidemiology*, *31*, 1113–1116.

Stansfeld, S. A. (2002b). Commentary: The problem with stress: minds, hearts and disease. *International Journal of Epidemiology*, *31*, 1113–1116.

Steer, R. A. (1999). Use of the beck depression inventory for primary care to screen for major depression disorders - the info-roc technique. *General Hospital Psychiatry*, *21*(2), 106–111.

Stefansdottir, I. K., & Vilhjalmsson, R. (2007). Dimensions of health-related lifestyle in young adulthood: results from a national population survey. *Scandinavian Journal of Caring Sciences*, *21*(3), 321–328.

Stelmack, R. M. (2004). *On the psychobiology of personality: Essays in honor of marvin zuckerman*. Amsterdam ; Oxford: Elsevier.

Stevens, C. D., & Ash, R. A. (2001). The conscientiousness of students in subject pools: Implications for "laboratory" research. *Journal of Research in Personality*, *35*(1), 91–97.

Stone, A. A., Schwartz, J. E., Smyth, J., Kirschbaum, C., Cohen, S., Hellhammer, D., et al. (2001). Individual differences in the diurnal cycle of salivary free cortisol: a replication of flattened cycles for some individuals. *Psychoneuroendocrinology*, *26*(3), 295–306.

Stone, J., Townend, E., Kwan, J., Haga, K., Dennis, M. S., & Sharpe, M. (2004). Personality change after stroke: Some preliminary observations. *Journal of Neurology, Neurosurgery and Psychiatry*, *75*(12), 1708–1713.

Straten, Annemieke, Cuijpers, Pim, Zuuren, Florence, et al. (2007). Personality traits and health-related quality of life in patients with mood and anxiety disorders. *Quality of Life Research*, *16*(1), 1–8.

Streiner, D. L. (2002). Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Canadian Journal of Psychiatry*, *47*(3), 262–266.

Strickland, P., Morriss, R., Wearden, A., & Deakin, B. (1998). A comparison of salivary cortisol in chronic fatigue syndrome, community depression and healthy controls. *Journal of Affective Disorders*, *47*(1-3), 191–194.

Swann, W. B., & Seyle, C. (2005). Personality psychology's comeback and its emerging symbiosis with social psychology. *Personality and Social Psychology Bulletin*, *31*(2), 155–165.

Tabachnick, B. G., & Fidell, L. S. (2000). *Using multivariate statistics* (Fourth ed.). Boston, MA: Allyn & Bacon.

Takano, A., Arakawa, R., Hayashi, M., Takahashi, H., Ito, H., & Suhara, T. (2007). Relationship between neuroticism personality trait and serotonin transporter binding. *Biological Psychiatry*, *62*(6), 588–592.

Taylor, R. S., Unal, B., Critchley, J. A., & Capewell, S. (2006). Mortality reductions in patients receiving exercise-based cardiac rehabilitation: How much can be attributed to cardiovascular risk factor improvements? *European Journal of Cardiovascular and Preventive Rehabilitation*, *13*(3), 369–374.

Tefft, M. E. E., & Boniface, D. R. R. (2000). Estimating food and nutrient intake from food frequency questionnaire data by reference to a standard weighed diet survey. *Journal of Human Nutrition and Dietetics*, *13*(3), 219–224.

Terracciano, A., Costa, P. T., & McCrae, R. R. (2006). Personality plasticity after age 30. *Personality and Social Psychology Bulletin*, *32*(8), 999–1009.

Thorn, L., Hucklebridge, F., Evans, P., & Clow, A. (2006). Suspected nonadherence and weekend versus week day differences in the awakening cortisol response. *Psychoneuroendocrinology*, *31*(8), 1009–1018.

Tochigi, M., Kato, C., Otowa, T., Hibino, H., Marui, T., Ohtani, T., et al. (2006). Association between corticotropin-releasing hormone receptor 2 (crhr2) gene polymorphism and personality traits. *Psychiatry and Clinical Neurosciences*, *60*(4), 524–526.

Torgerson, C. (2003). *Systematic reviews*. London: Continuum International Publishing Group.

Traustadóttir, T., Bosch, P. R., & Matt, K. S. (2003). Gender differences in cardiovascular and hypothalamic-pituitary-adrenal axis responses to psychological stress in healthy older adult men and women. *Stress*, *6*(2), 133–140.

Turner, D. C., & Sahakian, B. J. (2006). Neuroethics of cognitive enhancement. *Biosocieties*, *1*(1), 113–123.

Tyrka, A. R., Mello, A. F., Mello, M. F., Gagne, G. G., Grover, K. E., Anderson, G. M., et al. (2006). Temperament and hypothalamic-pituitary-adrenal axis function in healthy adults. *Psychoneuroendocrinology*, *31*(9), 1036–1045.

Tyrka, A. R., Wier, L. M., Price, L. H., Rikhye, K., Ross, N. S., Anderson, G. M., et al. (2008). Cortisol and ACTH responses to the Dex/CRH Test: Influence of temperament. *Hormones & Behavior*, *53*(4), 518–525.

Ussher, M., West, R., Evans, P., Steptoe, A., McEwen, A., Clow, A., et al. (2004). Acute reduction in secretory immunoglobulin a following smoking cessation. *Psychoneuroendocrinology*, *29*(10), 1335–1340.

Van Gestel, S., & Van Broeckhoven, C. (2003). Genetics of personality: are we making progress? *Molecular Psychiatry*, *8*(10), 840–852.

Van Itallie, T. B. (2002). Stress: A risk factor for serious illness. *Metabolism — Clinical and Experimental*, *51*(6, Suppl. 1), 40–45.

Van Os, J., & Jones, P. B. (1999). Early risk factors and adult person–environment relationships in affective disorder. *Psychological Medicine*, *29*(5), 1055–1067.

Vedhara, K., Miles, J., Bennett, P., Plummer, S., Tallon, D., Brooks, E., et al. (2003). An investigation into the relationship between salivary cortisol, stress, anxiety and depression. *Biological Psychology*, *62*(2), 89–96.

Vedhara, K., Stra, J. T., Miles, J. N., Sanderman, R., & Ranchor, A. V. (2006). Psychosocial factors associated with indices of cortisol production in women with breast cancer and controls. *Psychoneuroendocrinology*, *31*(3), 299–311.

Vickers, & Hervig, L. K. (1984). *Health behaviors: Empirical consistency and theoretical significance of subdomains.* San Diego: Naval Health Research Center.

Vickers, R. R., Conway, T. L., & Hervig, L. K. (1990). Demonstration of replicable dimensions of health behaviors. *Preventive Medicine*, *19*(4), 377–401.

Vollrath, M., & Torgersen, S. (2002). Who takes health risks? a probe into eight personality types. *Personality and Individual Differences*, *32*(7), 1185–1197.

Walker, B. R., Phillips, D. I., Noon, J. P., Panarelli, M., Andrew, R., Edwards, H. V., et al. (1998). Increased glucocorticoid activity in men with cardiovascular risk factors. *Hypertension*, *31*(4), 891–895.

Walsh, J. J., Wilding, J. M., Eysenck, M. W., & Valentine, J. D. (1997). Neuroticism, locus of control, type a behaviour pattern and occupational stress. *Work & Stress*, *11*(2), 148–159.

Warburton, D. E., Nicol, C. W., & Bredin, S. S. (2006). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal*, *174*(6), 801–809.

Wardle, J., & Steptoe, A. (1991). The european health and behaviour survey: Rationale, methods and initial results from the united kingdom. *Social Science & Medicine*, *33*(8), 925–936.

Ware, J. E., & Gandek, B. (1998). Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *Journal of Clinical Epidemiology*, *51*(11), 903–912.

Ware, J. E., Kosinski, M., Bayliss, M. S., Mchorney, C. A., Rogers, W. H., & Raczek, A. (1994). Comparison of methods for the scoring and statistical analysis of sf-36 health profile and summary measures: Summary of results from the medical outcomes study. *Medical Care*, *33*(4), AS264–AS279.

Ware, J. E., Kosinski, M., & Keller, S. K. (1994). *SF-36 physical and mental health summary scales: A user's manual.* Boston, MA: The Health Institute.

Ware, J. E., Snow, K. K., Kosisnki, M., & Gandek, B. (1993). *SF-36 health survey manual and interpretation guide*. Boston, MA: The Health Institute.

Wasserman, , Geijer, , Sokolowski, , et al. (2007). Genetic variation in the hypothalamicpituitaryadrenocortical axis regulatory factor, t-box 19, and the angry/hostility personality trait. *Genes, Brain & Behavior*, *6*(4), 321–328.

Wasylkiw, L., & Fekken, C. G. (2002). Personality and self-reported health: matching predictors and criteria. *Personality and Individual Differences*, *33*(4), 607–620.

Wasylkiw, L., & Fekken, G. C. (1999). The dimensionality of health behaviors. *Journal of Social Behavior and Personality*, *14*(4), 585–596.

Watson, D., & Pennebaker, J. W. (1989). Health complaints, stress, and distress: exploring the central role of negative affectivity. *Psychological Review*, *96*(2), 234–254.

Weiss, A., & Costa. (2005). Domain and facet personality predictors of all-cause mortality among medicare patients aged 65 to 100. *Psychosomatic Medicine*, *67*(5), 724–733.

Welling, L., & Thomson, L. (2003). *PHP and MySQL web development* (Second ed.). London: Pearson Education.

Wellings, K., Field, J., Johnson, A. M., Wadsworth, J., & Bradshaw, S. (1994). *Sexual behaviour in Britain*. London: Penguin Books Ltd.

Westrin, A., Engstom, G., Ekman, R., & Traskman-Bendz, L. (1998). Correlations between plasma-neuropeptides and temperament dimensions differ between suicidal patients and healthy controls. *Journal of Affective Disorders*, *49*(1), 45–54.

White, I. R. (1999). The level of alcohol consumption at which all-cause mortality is least. *Journal of Clinical Epidemiology*, *52*(10), 967–975.

Whiteman, M., Deary, I., Lee, A., & Fowkes, F. (1997). Submissiveness and protection from coronary heart disease in the general population: Edinburgh Artery Study. *The Lancet*, *350*(9077), 541–545.

Whiteman, M., Fowkes, F., Deary, I., & Lee, A. (1997). Hostility, cigarette smoking and alcohol consumption in the general population. *Social Science & Medicine*, *44*(8), 1089–1096.

Whiteman, M. C., Bedford, A., Grant, E., Fowkes, F. G. R., & Deary, I. J. (2001). The five-factor model (NEO-FFI) and the Personality Deviance Scales-Revised (PDS-R): Going around in interpersonal circles. *Personality and Individual Differences*, *31*(2), 259–267.

Whiteman, M. C., Fowkes, & Deary, I. J. (1997). Hostility and the heart. *British Medical Journal*, *315*(7105), 379–380.

Wiebe, D. J., & Fortenberry, K. (2007). Mechanisms relating personality and health. In M. E. Vollrath (Ed.), *Handbook of personality and health* (pp. 137–156). Chichester: John Wiley & Sons Inc.

Wilhelm, I., Born, J., Kudielka, B. M., Schlotz, W., & Wust, S. (2007). Is the cortisol awakening rise a response to awakening? *Psychoneuroendocrinology*, *32*(4), 358–366.

Williams, A. F., & Wechsler, H. (1972). Interrelationship of preventive actions in health and other areas. *Health Services Report*, *87*(10), 969–976.

Willoughby, M., Vandergrift, N., Blair, C., & Granger, D. A. (2007). A structural equation modeling approach for the analysis of cortisol data collected using prepostpost designs. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(1), 125–145.

Wilson, R. S., Krueger, K. R., Gu, L., Bienias, J. L., Mendes, & Evans, D. A. (2005). Neuroticism, extraversion, and mortality in a defined population of older persons. *Psychosomatic Medicine*, *67*(6), 841–845.

Wingard, D. L., Berkman, L. F., & Brand, R. J. (1982). A multivariate analysis of health-related practices: a nine-year mortality follow-up of the Alameda County Study. *American Journal of Epidemiology*, *116*(5), 765–775.

Winkelspecht, C., Lewis, P., & Thomas, A. (2006). Potential effects of faking on the NEO-PI-R: Willingness and ability to fake changes who gets hired in simulated selection decisions. *Journal of Business and Psychology*, *21*(2), 243–259.

Winter, D. G. (1997). Allport's life and allport's psychology. *Journal of Personality*, *65*(3), 723–731.

Wolf, O. T., Convit, A., Thorn, E., & De Leon, M. J. (2002). Salivary cortisol day profiles in elderly with mild cognitive impairment. *Psychoneuroendocrinology*, *27*(7), 777–789.

Wolpe, P. R. (2002). Treatment, enhancement, and the ethics of neurotherapeutics. *Brain and Cognition*, *50*(3), 387–395.

World Health Organization. (2001). *The world health report 2001 - mental health: New understanding, new hope*. Geneva: World Health Organization.

World Health Organization. (2002). *The world health report 2002 - reducing risks, promoting healthy life.* Geneva: WHO.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256–293). Newbury Park, CA: SAGE.

Wright, C. E., & Steptoe, A. (2005). Subjective socio-economic position, gender and cortisol responses to waking in an elderly population. *Psychoneuroendocrinology*, *30*(6), 582–590.

Wright, S., Johnston, M., & Weinman, J. (1995). *Measures in health psychology portfolio*. Windsor: NFER-Nelson.

Wust, S., Entringer, S., Federenko, I. S., Schlotz, W., & Hellhammer, D. H. (2005). Birth weight is associated with salivary cortisol responses to psychosocial stress in adult life. *Psychoneuroendocrinology*, *30*(6), 591–598.

Wust, S., Federenko, I., Hellhammer, D. H., & Kirschbaum, C. (2000). Genetic factors, perceived chronic stress, and the free cortisol response to awakening. *Psychoneuroendocrinology*, *25*(7), 707–720.

Wust, S., Federenko, I. S., Van Rossum, E. F. C., Koper, J. W., & Hellhammer, D. H. (2005). Habituation of cortisol responses to repeated psychosocial stress — further characterization and impact of genetic factors. *Psychoneuroendocrinology*, *30*(2), 199–211.

Yehuda, R., Teicher, M. H., Trestman, R. L., Levengood, R. A., & Siever, L. J. (1996). Cortisol regulation in posttraumatic stress disorder and major depression: A chronobiological analysis. *Biological Psychiatry*, *40*(2), 79–88.

Young-Bruehl, E. (2003). *Where do we fall when we fall in love?* New York: Other Press.

Zobel, A., Barkow, K., Schulze-Rauschenbach, S., Widdern, O. von, Metten, M., Pfeiffer, U., et al. (2004). High neuroticism and depressive temperament are associated with dysfunctional regulation of the hypothalamic-pituitary-adrenocortical system in healthy volunteers. *Acta Psychiatrica Scandinavica*, *109*(5), 392–399.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores.* Ottawa, ON: Directorate of Human Resources, Research and Evaluation, Department of National Defense.