

https://dspace.org/

Examining the relationships between socio-economic status and hypertension: an application of structural equation modelling

Morgan, Raquel

2017

http://hdl.handle.net/10539/23585

Morgan, Raquel (2017) Examining the relationships between socio-economic status and hypertension: an application of structural equation modelling, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/23585>

Downloaded from a DSpace repository. https://dspace.org/



Examining the relationships between socio-economic status and hypertension: An application of structural equation modelling

by:

Raquel Morgan

Supervisor: Renette Krommenhoek

A Research Report submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfillment of the requirements for the degree of Master of Science

May 31, 2017

Declaration

I, Raquel Gelinda Morgan, declare that this research report is my own, unaided work. It is being submitted for the Degree of Master of Science by coursework and research report to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

Signed: _____

Date: _____

Abstract

Over the years, epidemiological research has seen differing levels of the prevalence of hypertension across socio-economic strata. However in Sub-Saharan Africa, the patterns of association and underlying risk factors have often been poorly understood. In this study, we examined the extent to which socio-economic factors affect systolic and diastolic blood pressure across gender. Furthermore we explored whether certain risk factors associated with hypertension mediate this relationship. We used data from the third phase of the National Income Dynamic Study conducted in South Africa in 2012 on more than 18,000 adult individuals.

Structural equation modelling and multiple linear regression were used to estimate the relationship between blood pressure and various behavioural, demographic and socio-economic variables. These results were then compared to determine which technique provides more meaningful results. A higher socio-economic status was associated with a higher systolic and diastolic blood pressure in both males and females. Furthermore, body mass index was a mediator of the indirect effect of socio-economic status on blood pressure. Smoker status, alcohol consumption, physical exercise, emotional well-being and resting heart rate were also mediators; however their role was modest in comparison to BMI. One of the findings of this study is that a reduction in the BMI of an individual will have an impact on lowering hypertension. Furthermore, the promotion of healthy behaviours that target higher income groups need to be established so that these groups can make rational decisions in choosing their behaviours.

Acknowledgements

First and foremost, I would like to thank my Lord and Saviour for gracing me with the strength to pursue this dream and for helping me to make it a reality. I would then like to thank my husband and family for their constant support, encouragement and prayers that have carried me throughout this journey. I express my sincere appreciation and gratitude to Mrs. Renette Krommenhoek for her guidance, time and insights throughout this study. It was an honour having her as a supervisor and I thank her for her assistance and patience. Finally, thanks go to the reviewers of the proposal from the School of Statistics and Actuarial Science for their valuable suggestions and comments.

Contents

Declaration i

Abstract ii

Acknowledgements iii

List of Figures vii

List of Tables ix

List of Abbreviations xiii

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Problem Statement . . . 3
- 1.3 Structure of report . . . 4

2 Literature Review 5
- 2.1 History of Structural Equation Modelling . . . 5
  - 2.1.1 Basics of Structural Equation Modelling . . . 11
  - 2.1.2 Model fit indices . . . 16
  - 2.1.3 Mathematical Notation . . . 22
  - 2.1.4 Mediation Analysis . . . 25
- 2.2 Comparison of SEM and multiple linear regression . . . 30
- 2.3 Hypertension . . . 32
  - 2.3.1 Definition . . . 33
  - 2.3.2 Risk factors associated with hypertension . . . 33
- 2.4 Association between socio-economic status and health outcomes . . . 34
- 2.5 Summary . . . 37

3 Methodology 38
- 3.1 Data source and sampling design . . . 38
- 3.2 Measures used in the analysis . . . 39
  - 3.2.1 Demographic variables . . . 39
  - 3.2.2 Socio-economic variables . . . 39
  - 3.2.3 Blood pressure and resting heart rate . . . 41
  - 3.2.4 Bio-behavioural risk factors . . . 41
  - 3.2.5 Hypothesised model for structural equation model . . . 42
- 3.3 Techniques applied in analysis . . . 45

3.4 Summary 48

4 Data analysis and results **49**
4.1 Sample characteristics 49
4.2 Data screening 52
4.2.1 Missing data 52
4.2.2 Univariate and Multivariate normality 53
4.2.3 Outliers 55
4.2.4 Linearity and Homoscedasticity 56
4.2.5 Multicollinearity 57
4.3 Structural equation modelling 57
4.3.1 Cronbach's alpha and Exploratory factor analysis 57
4.3.2 Discussion of SEM results 62
4.3.2.1 Confirmatory factor analysis for emotional well-being 63
4.3.2.2 Confirmatory factor analysis for SES 66
4.3.2.3 SEM Model 1: Investigating the direct relationship between SES and blood pressure 68
4.3.2.4 SEM Model 2: Investigating whether bio-behavioural risk factors mediate the relationship between SES and blood pressure 71
4.4 Multiple linear regression 90
4.4.1 Discussion of multiple linear regression results 93
4.4.1.1 SBP multiple linear regression results for males . . 96
4.4.1.2 DBP multiple linear regression results for males . . 101
4.4.1.3 SBP multiple linear regression results for females . 106
4.4.1.4 DBP multiple linear regression results for females . 111
4.5 Comparison of SEM and multiple linear regression results 116
4.6 Summary 125

5 Conclusion **127**
5.1 Discussion of results 127
5.2 Strengths and limitations of the study 131
5.3 Recommendations for further work 132

References **133**

Appendix A Description of variables **146**

Appendix B Sample characteristics **148**

Appendix C Normality **154**

Appendix D Outliers **157**

Appendix E Cronbach's alpha **159**

Appendix F Exploratory Factor Analysis **161**

Appendix G Confirmatory Factor Analysis **164**

Appendix H SEM model 1 results **171**

Appendix I SEM model 2 results **178**

Appendix J Multiple linear regression results **217**

Appendix K Fit diagnostics **221**

Appendix L SAS code for multiple linear regression models **225**

List of Figures

2.1 A structural equation model depicting measurement and structural components (Source: Byrne 2010) . . . 7
2.2 Panel A: Illustration of direct effect where X affects Y. Panel B: Illustration of mediation design where X indirectly affects Y through M. . . . 26
2.3 Panel A: Illustration of direct effect where X affects Y. Panel B: Illustration of multiple mediation design where X indirectly affects Y through $M_1, M_2, ..., M_j$. . . . 27

3.1 Hypothesised pathways between SES and blood pressure . . . 43

4.1 Modified model: Emotional well-being CFA model for males and females . . . 64
4.2 Modified model: SES CFA model for males . . . 67
4.3 Modified model: SES CFA model for females . . . 67
4.4 Modified model: Model investigating direct relationship of SES and blood pressure for males . . . 70
4.5 Modified model: Model investigating direct relationship of SES and blood pressure for females . . . 70
4.6 Model 2.1 investigating mediated relationship of SES and blood pressure for males excluding controls . . . 73
4.7 Model 2.1 investigating mediated relationship of SES and blood pressure for females excluding controls . . . 74
4.8 Model investigating mediated relationship of SES and blood pressure for males including controls . . . 81
4.9 Model investigating mediated relationship of SES and blood pressure for females including controls . . . 82

D.1 Age boxplots for both genders . . . 157
D.2 Income boxplots for both genders . . . 157

G.1 Baseline model: Emotional well-being CFA model for males and females . . . 164
G.2 Baseline model: SES CFA model for males and females . . . 167

H.1 Baseline model: Model investigating direct relationship of SES and blood pressure for males . . . 171
H.2 Baseline model: Model investigating direct relationship of SES and blood pressure for females . . . 174

K.1 Test dataset: Fit diagnostics for model 3 . . . 221
K.2 Test dataset: Fit diagnostics for model 6 . . . 221
K.3 Test dataset: Fit diagnostics for model 9 . . . 222

K.4 Validation dataset: Fit diagnostics for model 11 222
K.5 Test dataset: Fit diagnostics for model 15 223
K.6 Test dataset: Fit diagnostics for model 18 223
K.7 Test dataset: Fit diagnostics for model 21 224
K.8 Test dataset: Fit diagnostics for model 24 224

List of Tables

2.1 Two-Index Presentation Strategy 20

4.1 Reliability Statistics for 7 variables loading onto emotional well-being for males 58
4.2 Inter-Item Correlation Matrix for 7 variables loading onto emotional well-being for males 58
4.3 Item-Total Statistics for 7 variables loading onto emotional well-being for males 59
4.4 KMO and Bartlett's Test for males iteration 2 60
4.5 Variance explained when eigenvalues exceed 1 for males iteration 2 . 60
4.6 Rotated Component matrix for males iteration 2 61
4.7 Description of mediation models for males and females 72
4.8 Model fit indices for SEM model 2.1 for males 75
4.9 Model fit indices for SEM model 2.1 for females 75
4.10 Model fit indices for SEM model 2.2 for males 83
4.11 Model fit indices for SEM model 2.2 for females 83
4.12 Description of each model built 91
4.13 Fit of each multiple linear regression model 93
4.14 Test dataset: Stepwise regression results for model 3 96
4.15 Test dataset: Overall summary of results for model 3 97
4.16 Test dataset: Parameter estimates for model 3 97
4.17 Test dataset: Stepwise regression results for model 6 99
4.18 Test dataset: Overall summary of results for model 6 99
4.19 Test dataset: Parameter estimates for model 6 100
4.20 Test dataset: Stepwise regression results for model 9 101
4.21 Test dataset: Overall summary of results for model 9 101
4.22 Test dataset: Parameter estimates for model 9 102
4.23 Validation dataset: Stepwise regression results for model 11 103
4.24 Validation dataset: Overall summary of results for model 11 104
4.25 Validation dataset: Parameter estimates for model 11 105
4.26 Test dataset: Stepwise regression results for model 15 106
4.27 Test dataset: Overall summary of results for model 15 106
4.28 Test dataset: Parameter estimates for model 15 107
4.29 Test dataset: Stepwise regression results for model 18 109
4.30 Test dataset: Overall summary of results for model 18 109
4.31 Test dataset: Parameter estimates for model 18 110
4.32 Test dataset: Stepwise regression results for model 21 111
4.33 Test dataset: Overall summary of results for model 21 112
4.34 Test dataset: Parameter estimates for model 21 112
4.35 Test dataset: Stepwise regression results for model 24 114
4.36 Test dataset: Overall summary of results for model 24 114

4.37 Test dataset: Parameter estimates for model 24 115

A.1 Description of variables used in analysis 146

B.1 Demographic and SES statistics of population 148
B.2 Bio-behavioural factors statistics for population 150
B.3 Blood pressure, resting heart rate and medication use 153

C.1 Skewness and kurtosis for variables exceeding the limits for males dataset 154
C.2 Skewness and kurtosis for variables exceeding the limits for females dataset 154
C.3 Univariate and Multivariate normality for males 154
C.4 Univariate and Multivariate normality for females 155

D.1 First 10 Mahalanobis d-squared values for males 158
D.2 First 10 Mahalanobis d-squared values for females 158

E.1 Reliability Statistics for 7 variables loading onto emotional well-being for females 159
E.2 Inter-Item Correlation Matrix for 7 variables loading onto emotional well-being for females 159
E.3 Item-Total Statistics for 7 variables loading onto emotional well-being for females 159

F.1 Rotated Component matrix for males iteration 1 161
F.2 KMO and Bartlett's Test for females 162
F.3 Variance explained when eigenvalues exceed 1 for females 162
F.4 Rotated Component matrix for females 162

G.1 Baseline model: Covariance modification indices for emotional well-being CFA model for males 164
G.2 Baseline model: Model fit indices for emotional well-being CFA model for males 165
G.3 Modified model: Covariance modification indices for emotional well-being CFA model for males 165
G.4 Modified model: Model fit indices for emotional well-being CFA model for males 165
G.5 Modified model: Unstandardised regression coefficients for emotional well-being CFA model for males 165
G.6 Baseline model: Covariance modification indices for emotional well-being CFA model for females 166
G.7 Baseline model: Model fit indices for emotional well-being CFA model for females 166
G.8 Modified model: Covariance modification indices for emotional well-being CFA model for females 166

G.9 Modified model: Model fit indices for emotional well-being CFA model for females . . . 167
G.10 Modified model: Unstandardised regression coefficients for emotional well-being CFA model for females . . . 167
G.11 Baseline model: Covariance modification indices for SES CFA model for males . . . 167
G.12 Baseline model: Model fit indices for SES CFA model for males . . . 168
G.13 Modified model: Covariance modification indices for SES CFA model for males . . . 168
G.14 Modified model: Model fit indices for SES CFA model for males . . . 168
G.15 Modified model: Unstandardised regression coefficients for SES CFA model for males . . . 169
G.16 Baseline model: Covariance modification indices for SES CFA model for females . . . 169
G.17 Baseline model: Model fit indices for SES CFA model for females . . . 169
G.18 Modified model: Covariance modification indices for SES CFA model for females . . . 170
G.19 Modified model: Model fit indices for SES CFA model for females . . . 170
G.20 Modified model: Unstandardised regression coefficients for SES CFA model for females . . . 170

H.1 Baseline model: Model fit indices for SEM model 1 for males . . . 171
H.2 Baseline model: Unstandardised regression coefficients for SEM model 1 for males . . . 171
H.3 Baseline model: Covariances for SEM model 1 for males . . . 172
H.4 Modified model: Model fit indices for SEM model 1 for males . . . 172
H.5 Modified model: Unstandardised regression coefficients for SEM model 1 for males . . . 172
H.6 Modified model: Covariances for SEM model 1 for males . . . 173
H.7 Modified model: Variances for SEM model 1 for males . . . 173
H.8 Modified model: Squared multiple correlations for SEM model 1 for males . . . 174
H.9 Baseline model: Model fit indices for SEM model 1 for females . . . 174
H.10 Baseline model: Unstandardised regression coefficients for SEM model 1 for females . . . 175
H.11 Baseline model: Covariances for SEM model 1 for females . . . 175
H.12 Modified model: Model fit indices for SEM model 1 for females . . . 175
H.13 Modified model: Unstandardised regression coefficients for SEM model 1 for females . . . 176
H.14 Modified model: Covariances for SEM model 1 for females . . . 176
H.15 Modified model: Variances for SEM model 1 for females . . . 177
H.16 Modified model: Squared multiple correlations for SEM model 1 for females . . . 177

I.1 Unstandardised regression coefficients for SEM model 2.1 for males 178
I.2 Standardised regression coefficients for SEM model 2.1 for males . . . 179

I.3 Squared multiple correlations for SEM model 2.1 for males 181
I.4 Standardised indirect effects for SEM model 2.1 for males 183
I.5 Significance of standardised indirect effects for SEM model 2.1 for males 184
I.6 Specific indirect effects for SEM model 2.1 for males 185
I.7 Standardised total effects for SEM model 2.1 for males 186
I.8 Unstandardised regression coefficients for SEM model 2.1 for females 188
I.9 Standardised regression coefficients for SEM model 2.1 for females . 189
I.10 Squared multiple correlations for SEM model 2.1 for females 191
I.11 Standardised indirect effects for SEM model 2.1 for females 192
I.12 Significance of standardised indirect effects for SEM model 2.1 for females 193
I.13 Specific indirect effects for SEM model 2.1 for females 194
I.14 Standardised total effects for SEM model 2.1 for females 195
I.15 Unstandardised regression coefficients for SEM model 2.2 for males 197
I.16 Standardised regression coefficients for SEM model 2.2 for males . . 198
I.17 Squared multiple correlations for SEM model 2.2 for males 200
I.18 Standardised indirect effects for SEM model 2.2 for males 202
I.19 Significance of standardised indirect effects for SEM model 2.2 for males 203
I.20 Specific indirect effects for SEM model 2.2 for males 204
I.21 Standardised total effects for SEM model 2.2 for males 205
I.22 Unstandardised regression coefficients for SEM model 2.2 for females 207
I.23 Standardised regression coefficients for SEM model 2.2 for females . 208
I.24 Squared multiple correlations for SEM model 2.2 for females 210
I.25 Standardised indirect effects for SEM model 2.2 for females 212
I.26 Significance of standardised indirect effects for SEM model 2.2 for females 213
I.27 Specific indirect effects for SEM model 2.2 for females 214
I.28 Standardised total effects for SEM model 2.2 for females 215

J.1 Dummy variables used in regression analysis 217
J.2 Independent variables used after stepwise regression 219

List of Abbreviations

ACS	Acute coronary syndrome
ADF	Asymptotic Distribution-Free
AGFI	Adjusted goodness of fit index
AIC	Akaike information criterion
ANOVA	Analysis of variance
BMI	Body mass index
CAIC	Consistent version of akaike information criterion
CFA	Confirmatory factor analysis
CFI	Comparative fit index
DBP	Diastolic blood pressure
EFA	Exploratory factor analysis
GFI	Goodness of fit index
IQR	Interquartile range
KMO	Kaiser-Meyer-Olkin
LISREL	Linear structural relations model
MAR	Missing at random
MCAR	Missing completely at random
MCMC	Markov Chain Monte Carlo
MSE	Mean square error
NFI	Normed fit index
NIDS	National Income Dynamic Study
PGFI	Parsimony goodness of fit index
PNFI	Parsimonious normed fit index
RMR	Root mean square residual
RMSEA	Root mean square error of approximation
RMSE	Root mean square error
SBP	Systolic blood pressure
SEM	Structural equation modelling

SES	Socio-economic status
SRMR	Standardised root mean square residual
TLI	Tucker-Lewis index
VIF	Variance inflation factor
WHO	World Health Organisation

Chapter 1

Introduction

1.1 Background

Hypertension, often referred to as high blood pressure, is becoming a central medical issue globally. The World Health Organisation (WHO) has reported that one of the leading risk factors for global mortality is increased blood pressure (WHO, 2014). In 2010, an estimated 9.4 million deaths were caused by increased blood pressure (WHO, 2014). An elevated blood pressure, if left uncontrolled, can cause many complications resulting in various diseases such as: renal disease, dementia, stroke and cardiac failure to name but a few. Furthermore as a result of these complications, heavy financial burdens and service burdens are imposed on the healthcare system.

Globally the overall prevalence of high blood pressure was approximately 22% in 2014 for adults 18 years and older. In addition, the highest prevalence of hypertension in adults was recorded in Africa at approximately 30% while the lowest prevalence of high blood pressure was found in America at 18% (WHO, 2014). The global health report states that factors such as: alcohol abuse, physical inactivity, unhealthy diet, tobacco use, ageing, psychological stress, genetic factors, determinants of socio-economic status and inadequate access to health care all contribute to the high prevalence of hypertension. Some of these underlying risk factors are targeted through a change in lifestyle which may result in a reduction in high blood pressure (NICE, 2011).

Extensive research has shown that lower socio-economic status (SES) is often associated with poorer health and poorer bio-behavioural risk profiles which in

turn results in a higher systolic blood pressure (SBP) (Brummett, Babyak, Siegler, Shanahan, Harris, Elder and Williams, 2011). Disparities in SES in relation to the prevalence of hypertension have been observed over the years particularly in countries with a higher income where extensive epidemiological research relates a higher SES to a lower prevalence of hypertension and cardiovascular disease (Cois and Ehrlich, 2014). In contrast however, various studies in Sub-Saharan Africa have depicted mixed associations between SES and blood pressure (Lang, Pariente, Salem and Tap, 1988; Norman, Bradshaw and Steyn, 2001; Ploubidis, Mathenge, De Stavola, Grundy, Foster and Kuper, 2013). In some cases in Sub-Saharan Africa, the control and awareness of hypertension has been associated with higher SES (Norman *et al.*, 2001) while in other cases education which is often a determinant of SES was not related to the prevalence of hypertension (Ploubidis *et al.*, 2013). Small non-representative samples, inconsistent measures of SES or health outcomes and design flaws have often been argued as some of the potential reasons for the conflicting results (Fernald and Adler, 2008). Lam (2011) stated that SES does not directly impact hypertension but exerts its effects via a complex interaction of bio-behavioural factors such as diet and exercise. These mediating factors are potentially modifiable and a better understanding of these risk factors may provide opportunities for preventative interventions which in turn may help to address the health inequalities among different social groups.

Classical statistical analysis such as multiple linear regression or analysis of variance (ANOVA) are often not suitable to explore the complexity of the preceding inter-relationships since the effects of SES on the health outcome would ignore the indirect effects of the mediating factors and would simply be adjusted for each other (De Stavalo, Nitsch, dos Santos Silva and Leon, 2006). Structural equation modelling (SEM) is often used to deduce meaningful insight about the associations between these variables. Kastorini, Milionis, Georgousopoulouy, Kalantzi, Nikolaouz, Vemmosx, Goudevenos and Panagiotakos (2015) highlighted the importance of using SEM when testing hypotheses consisting of complex pathways with interacting variables as it overcomes some of the shortcomings of common statistical methodologies such as ANOVA or multiple linear regression by illuminating

relationships between underlying concepts of a theory.

1.2 Problem Statement

The objective of this study is to evaluate how socio-economic factors affect hypertension lifestyle related risk factors and to clarify the potential pathways explaining the association between SES and blood pressure using SEM and regression analysis. Various researchers have assessed the mediating effects of modifiable bio-behavioural risk factors on the association between SES and health outcomes (Chaix, Bean, Leal, Thomas, Havard, Evans, Jégo and Pannier, 2010; Brummett *et al.*, 2011; Ploubidis *et al.*, 2013; Miyaki, Song, Taneichi, Tsutsumi, Hashimoto, Kawakami, Takahashi, Shimazu, Inoue, Kurioka and Shimbo, 2013; Bardenheier, Bullard, Caspersen, Cheng, Gregg and Geiss, 2013; Cois and Ehrlich, 2014; Kastorini *et al.*, 2015). The studies were predominantly conducted in western countries with a higher level of income and a different morbidity profile to South Africa and equivalent studies in lower income countries are lacking. Furthermore, the burden of hypertension has been observed to be disproportionate across socio-economic levels and its risk factors often vary across genders (Wong, Wang, Leung, Tsang, Lo and Griffiths, 2015).

Recently, the National Income Dynamics study (NIDS) conducted in South Africa has provided quality behavioural, socio-demographic and anthropometric data on a large sample of individuals for research purposes (De Villiers, Brown, Woolard, Daniels and Leibbrandt, 2013). In this study, this data will be analysed in order to provide further insights on the relationships and associations between SES and hypertension in South Africa. The objective of this study is thus to assess the patterns of the diastolic and systolic blood pressure by gender and to examine its determinants with respect to behavioural and socio-economic factors.

The specific research questions addressed in this study are:

1. What are the relationships between demographic, behavioural, socio-economic variables and blood pressure?

2. Is there an independent association of education, income, occupation status and ownership of durable goods (all regarded as measures of SES) with blood pressure?

3. How useful are the differences in health behaviours such as physical activity, emotional well-being, smoking, alcohol consumption, resting heart rate and body mass index in explaining these relationships?

The results from this analysis could have important implications as it will enable an identification of population groups at risk for hypertension and could inform interventions aimed at reducing social inequalities in healthcare in South Africa. The SEM and multiple linear regression will be compared with regards to answering the research questions above to determine which one of the two approaches provides better results.

1.3 Structure of report

The report is structured in the following manner: Chapter 2 presents a review of the literature concerning SEM, multiple linear regression and the applications of these techniques in understanding the patterns and determinants of hypertension. Chapter 3 outlines the methodology for applying SEM and multiple linear regression as well as a description of the data used in the analysis. The results of these analyses are presented and discussed in Chapter 4, with conclusions and recommendations in Chapter 5.

Chapter 2

Literature Review

This chapter reviews some of the published literature on the concept of SEM, multiple linear regression and the applications of these techniques in understanding the determinants of hypertension. Section 2.1 outlines the history of SEM, basics of SEM and the mathematical notation required for understanding SEM. Section 2.2 highlights the similarities and differences between structural equation modelling and multiple linear regression. Section 2.3 provides a description of hypertension and the risk factors associated with hypertension while section 2.4 describes previous studies that have aimed to understand the relationship between socio-economic status and health outcomes.

2.1 History of Structural Equation Modelling

Early development of structural equation models began in the 1900's however it is only in the last three decades that the growing interest in applying structural equation models particularly in econometrics and social sciences has been on the rise. SEM has been described in literature as a "comprehensive statistical approach to testing hypothesis about relations among observed and latent variables" (Hoyle, 1995). In addition, Rigdon (1998) described SEM as a methodology to represent, estimate, and test a theoretical framework of (mostly) linear relations between variables. Kline (2011) emphasised that the two main purposes of SEM are firstly to understand the covariance or correlation patterns amongst observed variables and secondly to interpret as much of the variance as possible in the specified model. Byrne (2010) highlighted that the term SEM describes two essential features of the procedure: (1) the causal processes that are studied are represented by a series of

structural equations and (2) these structural relations can be depicted pictorially to enable a clearer conceptualisation of the theory being studied. Thereafter the hypothesised model is examined statistically in a simultaneous analysis of the entire system of variables to determine the degree to which it is consistent with the data.

SEM is often described as a collection of related techniques and as a result does not have a single inventor. It originated from the work of Charles Spearman in 1904 who developed exploratory factor analysis (Spearman, 1904) . This was then followed in 1918 by a young genetist called Sewall Wright who performed the first application of path analysis (Wright, 1918). Breakthroughs in the development of SEM took place in the 1970's and 1980's. Jöreskog (1973), Wiley (1973) and Keesling (1972) were three authors who were involved in the early development of structural equation models as they combined both the measurement (factor analysis) and structural (path analysis) approaches. Their approach was initially known as the Jöreskog, Keesling and Wiley (JKW) model but through the development of the first software program by Jöreskog and Sörbom (Jöreskog and Sörbom, 1989), it then became known as the linear structural relations model (LISREL). The term 'structural relation' pertains to the ability to handle the relationships amongst latent variables which are often described as unobserved variables in a model that can be approximated by observed variables. These relations are often described by linear regression equations which are represented graphically by path diagrams using arrows (Nachtigal, Kroehne, Funke, Steyer and Schiller, 2003) as depicted in the following example.

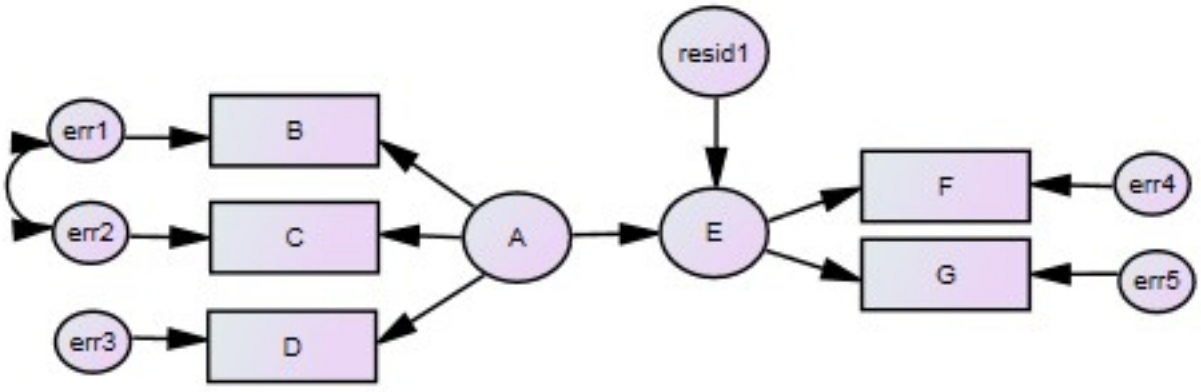


Figure 2.1: A structural equation model depicting measurement and structural components (Source: Byrne 2010)

In any SEM analysis it is standard convention to use squares or rectangles to signify the observed or manifest variables and circles or ellipses to depict the latent or unobserved variables (Hoyle, 1995). Latent variables are usually hypothetical constructs of interest in a study. These variables are often referred to as common factors in factor analysis literature (Ho, Stark and Chernyshenko, 2012). Path analysis usually describes three types of effects: indirect, direct and total effects. Direct effects depict the influence of one variable on another unmediated by other variables in the model while indirect effects are mediated by at least one intervening variable (Bollen, 1989). The total effect is the sum of the indirect and direct effects. Lines which are directed from one observed variable to another denote the direct effects (Schumacker and Lomax, 2004).

In figure 2.1 there are two latent factors depicted by **A** and **E** and five observed variables. Three of the observed variables are hypothesised to measure **A** (**B**; **C**; **D**), and two are hypothesised to measure **E** (**F**; **G**). There is an error term (denoted by err1 to err5) which is enclosed in a circle and is associated with each of the observed variables and this represents the measurement error. In this example, some of variable **B** will be explained by variable **A** and some will not. This unexplained portion depicts the error term which indicates other influences on variable **B** that are not contained in this model. The residual term which is depicted by resid1 denotes the error in the prediction of the endogenous factor

(**E**), also known as the dependent variable, from the exogenous factor (**A**) which is the independent variable. In this example it is thus hypothesised that variable **A** has a direct influence on variable **E** which implies that variable **A** may result in a decrease or increase in variable **E**. The covariance of a variable is denoted by a curved double-headed line between two independent observed variables. In the example, the double-headed line indicates that the measurement error associated with variable **B** and variable **C** are specified to be correlated. This implies that there are influences on both of these independent variables that are outside of the path model thus one would expect the same unmeasured variables to influence both independent variables.

This general structural equation model thus comprises of two submodels: a measurement model and structural model. The measurement model describes the relationships between the unobserved and the observed variables and represents the confirmatory factor analysis (CFA) model which will be described later. Conversely, the structural model depicts the relationships among the unobserved variables. Thus it emphasises the way in which specific latent variables indirectly or directly influence changes in values of other latent variables in the model (Byrne, 2010). From this example, it is evident that SEM is a very flexible technique because it deals with simple or multiple linear regression and also a system of simultaneous regression equations.

One of the major disadvantages of using models only with observed variables is that the measurement error is not taken into account. If one only used observed variables then one assumes that all measured variables are perfectly valid and reliable which is highly unlikely (Schumacker and Lomax, 2004). For example, education level is not a perfect measure of socio-economic status nor is the amount of exercise per week a perfect measure of fitness. As a result, validity and reliability of scores are first examined and if they are at an acceptable level then these scores are used in the analysis. Factor analysis was thus developed to take into account the measurement error of the variables. The aim of factor analysis is to discover which sets of observed variables share common variance-covariance characteristics that define theoretical constructs or factors called latent variables (Schumacker

and Lomax, 2004). The main assumption is that some factors which are smaller than the number of observed variables are responsible for the shared-covariance among the observed variables thus factor analysis is often referred to as a data reduction technique (Bollen, 1989). As a result researchers often collect data on observed variables and use factor analysis to either confirm that a set of variables define the constructs or explore which variables relate to factors.

There are two types of factor analysis namely exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA entails finding a model that has theoretical support and that fits the data while CFA assesses the significance of a hypothesised factor model. Schumacker and Lomax (2004) indicate that in CFA one specifies the number of factors which are correlated and which observed variable measures each factor while in EFA the individual determines how many factors there are, whether these factors are correlated and which observed variables best measure each factor. Thus the main difference between CFA and EFA is that in CFA one has a priori-specified theoretical model while in EFA there is no such model. In EFA, the number of underlying factors and the factor loadings are estimated using statistical techniques.

Transformations such as the varimax rotation can be used to improve the interpretation of the results when EFA is applied (Hox and Bechger, 2007) however factor rotation and extraction do not form part of CFA. Often some factor loadings are constrained to be zero thus CFA is particularly useful since it can estimate such a structure exactly while in EFA approximating a simple structure requires rotation. Furthermore for each factor, one loading needs to be fixed to one for interpretability of the latent factor (Hox and Bechger, 2007). For each latent factor it is vital that one either estimates the loading given a fixed variance for a latent factor or estimates the factor variances given at least one fixed loading. Moreover in SEM, the CFA model is imposed upon the data as one first obtains parameter estimates and thereafter assesses the model fit.

Today, most statistical programs like AMOS, LISREL or EQS provide the user with a graphical interface which enables one to construct path diagrams, determine

the model fit and estimate the parameters with ease. Previously SEM programs required the covariance matrix of observed variables as input data however due to advancements over the years software can now handle raw input data files which makes the application much easier. When applying SEM an adequate sample size is required in order to maintain power and ensure stable standard errors and parameter estimates (Schumacker and Lomax, 2004). In any analysis of a complex model, more cases are required than that of a simpler model because complex models have more parameters than simpler models. Models with more parameters require more estimates, so larger samples are necessary in order for the results to be stable. The type of estimation algorithm that is used in the analysis will also greatly affect the sample size requirements (Kline, 2011).

Jackson (2003) describes a useful rule of thumb called the $N : q$ rule which relates the sample size to the model complexity. This rule is particularly useful when one applies the maximum likelihood estimation method which will be described in more detail later on. In maximum likelihood estimation Jackson (2003) indicates that the sample size should be thought of in terms of the ratio of cases (N) to the number of model parameters that require statistical estimates (q). It is desirable to have a 20:1 ratio of the number of cases to model parameters however a 10:1 ratio is often seen as a realistic minimum target. Any ratio below this, leads to estimates that are unstable and one should not trust the results. In addition, Kline (2011) suggests that the sample size should at least be 200 observations when applying maximum likelihood estimation. The weighted least squares estimation method however provides an asymptotically distribution free approach and thus requires an exceptionally large sample size. A sample size of at least 2000 was proposed by Yung and Bentler (1994) in their simulation study in order to provide satisfactory results. Many researchers have shown that the stability and accuracy of SEM declines with a decreasing sample size and an increasing number of variables.

Over the years, epidemiological studies have focused their attention on the cause of the disease however in recent years a view has developed that a disease is often a mix of economic, environmental, political and social factors. Questions raised by public health researchers have thus become more focused on social, economic and

political factors. One of the tools that can assist healthcare researchers in dealing with these complexities is SEM and this will described in our application of SEM.

2.1.1 Basics of Structural Equation Modelling

The basic building blocks of any SEM analysis follow a logical sequence of six basic steps as described by Kline (2011). Any SEM analysis will begin with model specification and will be followed by model identification. This will then be followed by model estimation. Model respecification will occur in cases where there is a poor initial model fit. Lastly reporting of the results will take place in order to ensure that model results are accurately described in written reports.

Model specification is a process that entails using all of the relevant theoretical research and information in order to develop a theoretical model. A researcher thus needs to determine which variables to include in the theoretical model and how these variables are related, prior to any data collection (Schumacker and Lomax, 2004). This process often involves constructing a model diagram using graphical symbols that can alternatively be described as a series of equations. These equations explain the model's parameters which describe the relations among the observed or latent variables that are estimated from the sample data. Specification is the most important step (Kline, 2011) and is described as the hardest part of structural equation modelling (Cooley, 1978).

Model identification pertains to the correspondence between the free parameters which are to be estimated and the information from which they are to be estimated that is, the observed covariances and variances (Hoyle, 1995). Identification occurs if it is theoretically possible to derive a single unique value for each model parameter (Kline, 2011). If models are not identified then they should instead be respecified before continuing with any further analysis. If a value can be acquired through one manipulation of the observed data for each parameter then the model is just identified and has zero degrees of freedom. A just identified model thus has an equal amount of known and unknown information. If instead a value for the parameters

can be obtained in many different ways then the model is overidentified and has degrees of freedom equal to the number of observed variances and covariances minus the number of free parameters (Hoyle, 1995). An overidentified model thus has more known information than unknown information. A model is underidentified if it is not possible to uniquely estimate all of its parameters. This occurs when there is more unknown than known information. Therefore in any SEM analysis, the model specification ensures that the model to be estimated is either just identified or overidentified. Overidentified models however are the only useful models as they enable one to examine the fit of the overall model.

Once the model has been identified after assessing the free parameters to be estimated and the information from which these need to be estimated, then model estimation occurs. This involves firstly examining the model fit and secondly obtaining and interpreting estimates of the free parameters from the set of observed data (Kline, 2011). The structural equation model uses various techniques to estimate the unknown parameters such that the implied covariance matrix, Σ , is close to the sample covariance matrix, S. There are various fitting functions that are used to minimise the difference between Σ and S (Bollen, 1989). The distributional properties of the variables that are being examined drive the choice of the estimation technique.

Maximum likelihood estimation is the most commonly used technique in many statistical programs. The fit function for maximum likelihood estimation that is given by (Bollen, 1989) is:

$$F_{ML} = \log|\Sigma(\theta)| + tr(S\Sigma^{-1}(\theta)) - \log|S| - p \tag{2.1}$$

where:

θ is the vector that contains the model parameters;

$\Sigma(\theta)$ is the covariance matrix written as a function of θ;

$tr(.)$ is the trace function;

S is the sample covariance matrix; and

p is the number of observed variables.

Lei and Wu (2012) indicated that under the assumption of multivariate normality of the observed variables and correct model specification, the maximum likelihood estimator is asymptotically consistent, efficient, unbiased and normally distributed. Furthermore the model fit statistic is asymptotically distributed as χ^2 with $\frac{1}{2}p(p+1) - t$ degrees of freedom, where t is the number of model parameters estimated. The maximum likelihood estimator however tends to produce unbiased parameter estimates, inflate the χ^2 test which results in a statistically significant model and deflate standard error estimates in cases of non-normality (Lei and Wu, 2012).

According to literature, generalized least squares and unweighted least squares are two additional estimation procedures. The unweighted least squares fitting function as described by Bollen (1989) is given by:

$$F_{ULS} = \left(\frac{1}{2}\right) tr[(S - \Sigma(\theta))^2] \tag{2.2}$$

This function is justified when all variables are measured in the same units. The unweighted least squares estimator is the simplest fitting function and leads to consistent estimates (Schumacker and Lomax, 2004). The disadvantage of the unweighted least squares estimation method is that it has no distributional assumptions and as a result changes in the observed variable scale can yield different estimates or solutions (Schumacker and Lomax, 2004).

The generalized least squares fitting function as described by Bollen (1989) is given by:

$$F_{GLS} = \left(\frac{1}{2}\right) tr\{\left[I - \Sigma(\theta)S^{-1}\right]^2\} \tag{2.3}$$

Both the maximum likelihood estimation and generalized least squares method are scale free thus if one transforms one or more of the observed variables, the transformed and the untransformed variables yield estimates that differ by the

transformation thus they are properly related (Schumacker and Lomax, 2004). Both the generalized least squares estimation method and maximum likelihood estimation method have desirable asymptotic properties that is, large sample properties, such as unbiasedness and minimum variance and both methods assume that the observed variables have multivariate normality (Schumacker and Lomax, 2004).

There are various other estimation methods such as generally weighted least squares, two stage least squares, diagonally weighted least squares and asymptotic distribution-free (ADF) estimation methods which can be used in cases where the methods discussed do not meet some of the assumptions. The weighted least squares estimator which was proposed by Browne (1984) relaxes the distributional assumptions and is referred to as Browne's asymptotically distribution-free method. This method is particularly useful in dealing with non-normal data however recent literature has proven that unless the sample size is in excess of 10 times the number of parameters estimated, the ADF estimator performs very poorly and cannot be trusted (Raykov and Marcoulides, 2000). In some cases, the data does not only include continuous variables, but also includes categorical or ordinal variables. Literature to date supports the notion that when the number of categories is large and the data approximates a normal distribution then failure to address the ordinality of the data is negligible (Muthén and Kaplan, 1985; Byrne, 2010). Furthermore Bentler and Chou (1987) emphasised that in cases of normally distributed categorical variables, continuous methods can be used without concern when a categorical variable has four or more categories. Recent literature has supported these findings and shown that the χ^2 statistic is impacted most by the two category response and becomes less influenced as the number of categories increases (Green, Akey, Fleming, Hershberger and Marquis, 1997).

The estimation methods described are iterative methods since they involve numerous attempts to obtain estimates from free parameters that imply a covariance matrix like the observed one (Hoyle, 1995). Iterative implies that an initial estimate is first made then cycles of calculations are performed usually using a computer to improve the estimates. An improvement occurs when the model-implied covariances based

on the estimates at each step become more similar to the observed ones. This iterative process continues until all the values in the residual matrix cannot be minimised any further and thus the estimation procedure has converged (Suhr, 2006). Once convergence of the estimation procedure has occurred, a single number that summarizes the degree of correspondence between the observed and expected covariances is produced. This number is usually the starting point for constructing indices for model fit and is often called the value of the fitting function (Suhr, 2006). In some cases, non-convergence of the estimation procedure can occur which is usually as a result of poor starting values or poor model specification. This occurs when the interactive process is terminated due to excessive computer time or as a result of the criteria specified such as the maximum number of iterations required. According to Suhr (2006), in any SEM analysis non-converged results should not be trusted and one should always avoid linear dependencies among variables. As a result, variables should be carefully chosen before the model is tested and estimated.

There are four assumptions which need to be met in any SEM analysis:

1. Multivariate normality;
2. Linearity of all relationships;
3. Random sampling of respondents;
4. Independence of the observations (Mohanlal, 1998).

Once these assumptions have been adhered to, the results are accepted and the goodness of fit indices need to be examined for the overall model and then for the structural and measurement models separately. The results are then reviewed for estimated coefficients that exceed acceptable thresholds, called offending estimates. Common examples of offending estimates are those with:

1. Standardised coefficients very close to or exceeding 1;
2. Very large standard errors for estimates;

3. Non-significant or negative error variances for any construct;

4. Covariance or correlation matrices that are not positive definite (Mohanlal, 1998).

The overall model fit is examined after the assumptions are adhered to and there are no offending estimates. Generally one will look at the χ^2 model as a whole as well as a list of fit measures when assessing the model (Blunch, 2008).

2.1.2 Model fit indices

There are three types of goodness of fit indices: absolute fit indices, parsimony fit indices and incremental fit indices. Absolute fit indices are used to determine the degree to which a priori model fits the sample data by comparing a function of discrepancy from the fitted model to a function of discrepancy from the null model and thus determines which model has the best fit (McDonald and Ho, 2002). Absolute fit indices provide the most crucial evidence of how well the model fits the data. This is because they do not use an alternative model as a base for comparison but are simply derived from the fit of the obtained and the implied covariance matrices and the maximum likelihood minimization function. There are several absolute fit indices used in literature: chi-square statistic, goodness of fit index (GFI), adjusted goodness of fit index (AGFI), root mean square error of approximation (RMSEA), root mean square residual (RMR) and standardised root mean square residual (SRMR).

The most traditional measure of examining the overall model fit is the chi-square statistic as it "assesses the magnitude of discrepancy between the sample and fitted covariance matrices" (Hu and Bentler, 1999). A good model fit would indicate a non-significant result at 0.05 level (Barrett, 2007). The chi-square statistic is thus often known in literature as a 'badness of fit' measure (Kline, 2011). Although the chi-square statistic is popularly used in studies to assess the fit of the model, there are some limitations in its use. Firstly the chi-square statistic assumes multivariate

normality thus any departures from this could cause the model to be rejected even in cases where it is properly specified (McIntosh, 2006). Furthermore this test is very sensitive to both large and small sample sizes. If the sample size is small, one will always accept the null hypothesis while one tends to reject the null hypothesis if the sample size is large (Blunch, 2008). In order to minimise the sensitivity to the sample size, the normed chi-square statistic was developed by Wheaton, Muthen, Alwin and Summers (1977). The normed chi-square statistic is the chi-square statistic adjusted for the degrees of freedom and various acceptable ratios have been introduced in literature with some thresholds as low as 2 (Tabachnick and Fidell, 2007) and others as high as 5 (Wheaton *et al.*, 1977).

The root mean square error of approximation (RMSEA) is the second commonly used absolute fit index. This index was developed by Steiger and Lindt in 1980 (Steiger, 1990) and depicts how well the model with unknown, optimally chosen parameters fits the populations covariance matrix. Diamantopoulos and Siguaw (2000) describe this index as one of the most informative fit indices since it is very sensitive to the number of estimated parameters in the model. In the early nineties any cutoff greater than 0.10 was considered a poorly fitted model however in recent years, a cutoff of approximately 0.06 (Hu and Bentler, 1999) or a threshold value of 0.07 (Steiger, 2007) is considered as the cutoff for RMSEA. The main advantage of this statistic is that a confidence interval can be calculated around its value (MacCallum, Browne and Sugawara, 1996).

The goodness of fit index (GFI) is another widely used index that was developed as an alternative to the chi-square statistic (Jöreskog and Sörbom, 1981). This index determines the proportion of the variance accounted for by the estimated population covariance (Tabachnick and Fidell, 2007). This index lies between 0 which depicts poor fit and 1 which depicts perfect fit. The disadvantage of this index is that it has upward bias with larger samples and it increases as the number of parameters increases. Furthermore, it has downward bias when there are a large number of degrees of freedom in relation to the sample size. The adjusted goodness of fit index (AGFI) was developed to adjust the GFI for the degrees of freedom (Jöreskog and Sörbom, 1981). Similarly a value of 0 indicates poor fit for AGFI

while a value of 1 depicts perfect fit. For both the GFI and AGFI, it has generally been accepted that a cutoff of 0.90 is an acceptable model. The AGFI and GFI are both severely influenced by the sample size and as a result are not used as standalone indices.

The root mean square residual (RMR) and standardised root mean square residual (SRMR) depict the square root of the differences between the residual of the sample covariance matrix and the hypothesised covariance model. The RMR is dependent on the scale of each indicator and thus can become difficult to interpret while the SRMR overcomes this difficulty and is easier to understand. The SRMR values lie between 0 and 1 with good fitting models having values less than 0.05 (Diamantopoulos and Siguaw, 2000) however values as large as 0.08 can be considered acceptable (Hu and Bentler, 1999). A disadvantage of SRMR is that it tends to be lower when there are a larger number of parameters in the model or a large sample size.

Incremental fit measures are often referred to as comparative indices since they compare the proposed model to a comparison model. There are three incremental fit measures described in literature: Tucker-Lewis index (TLI), the comparative fit index (CFI) and the normed fit index (NFI). Tucker and Lewis (1973) describe the TLI as combining a measure of parsimony into a relative index between the proposed model and the null model. The null model is often referred to as the independence model since it specifies that all measured variables are uncorrelated while the proposed model is the target model which is the model obtained based on proposed theory. The TLI is expressed as:

$$TLI = \frac{\left[\left(\frac{\chi^2_{null}}{df_{null}}\right) - \left(\frac{\chi^2_{proposed}}{df_{proposed}}\right)\right]}{\left[\left(\frac{\chi^2_{null}}{df_{null}}\right) - 1\right]} \tag{2.4}$$

where:

χ^2_{null} is the chi-square statistic for the null model;

$\chi^2_{proposed}$ is the chi-square statistic for the proposed model;

df_{null} is the degrees of freedom for the null model; and

$df_{proposed}$ is the degrees of freedom for the proposed model.

This index lies between 0 and 1 with a value of 0.90 or more considered acceptable.

NFI was developed by Bentler and Bonnet (1980) in order to compare the χ^2 value of the proposed model to the χ^2 value of the null model. The values of this statistic range from 0 to 1 with a value of 0.90 or more considered acceptable. NFI is expressed as:

$$NFI = \frac{\left[\chi^2_{null} - \chi^2_{proposed}\right]}{\chi^2_{null}} \tag{2.5}$$

This index is very sensitive to the sample size and often underestimates the fit when the sample size is less than 200 (Bentler, 1990) thus it must not solely be relied on (Kline, 2011).

The CFI is a revised form of the NFI which performs well even when there is a small sample size (Tabachnick and Fidell, 2007). The CFI was developed by Bentler (1990) and ranges from 0 which depicts poor fit to 1 which depicts perfect fit. In recent years a CFI value of 0.95 or more is considered as an indication of a good model fit (Hu and Bentler, 1999).

The third type of model fit indices are called model parsimony indices. Parsimony is described by Schumacker and Lomax (2004) as the number of estimated parameters needed to achieve a specific level of fit. In most cases, an overidentified model is compared to a restricted model thus these indices assess whether model fit has been established by overfitting the data with too many coefficients. Two types of parsimony indices called the Parsimonious Normed Fit Index (PNFI) and the Parsimony Goodness-of-Fit Index (PGFI) were developed by Mulaik, James, Van Alstine, Bennet, Lind and Stilwell (1989). The PGFI adjusts the GFI for the loss of degrees of freedom while the PNFI adjusts the NFI for the loss of degrees of freedom. These indices are useful since both penalise for model complexity and thus the values of these indices are much lower than the goodness of fit indices. Mulaik *et al.* (1989) indicate that these indices can be within the 0.50 region while

other goodness of fit indices are much higher. Since no acceptable threshold has been recommended over the years, Mulaik *et al.* (1989) indicate that these indices must be used in conjunction with other goodness of fit indices.

A second type of parsimony indices are called the information criteria indices. There are two well-known indices that have been described in literature, the Akaike Information Criterion (AIC) and the Consistent Version of AIC (CAIC) (Akaike, 1974). These indices have predominantly been used when comparing non-hierarchical models estimated with the same data. AIC values closer to zero indicate a good fit and these indices require a sample size of at least 200 in order for interpretation to be reliable (Diamantopoulos and Siguaw, 2000).

Due to the variety of indices discussed, it often becomes a temptation to use indices that indicate the best model fit however researchers have emphasised that this should be avoided as much as possible. In a review by McDonald and Ho (2002), they indicate that the most widely used indices are NFI, GFI, CFI and the TLI. Kline (2011) indicates that although the model χ^2 has its deficiencies, it should always be reported along with the relevant p value and degrees of freedom. Furthermore Kline (2011) also emphasises the use of the RMSEA, CFI and SRMR to assess model fit. In addition, the Two-Index Presentation Strategy was introduced by Hu and Bentler (1999). This strategy included a combination of rules for the following sets of two indices:

Table 2.1: Two-Index Presentation Strategy

Fit Index Combination	**Combinational Rules**
TLI and SRMR	TLI$\geq$0.96 and SRMR$\leq$0.09
RMSEA and SRMR	RMSEA$\leq$0.06 and SRMR$\leq$0.09
CFI and SRMR	CFI$\geq$0.96 and SRMR$\leq$0.09

Based on the literature over the years, it would thus seem sensible to always include the χ^2 statistic, RMSEA, CFI, SRMR, TLI and potentially one parsimony fit index such as PNFI. These indices have been chosen over other indices as they have been found to be the most insensitive to sample size, model misspecification and parameter estimates (Hooper, Coughlan and Mullen, 2008).

If the model fit is adequate then the estimates of the parameters are assessed. The ratio of each parameter estimate to its standard error is distributed as a z-statistic. This z-statistic is significant at the 0.05 level if the value is greater than 1.96 and at the 0.01 level if the value is greater than 2.56 (Hoyle, 1995). If the model fit is found to be unacceptable then the model is revised in cases where the modifications are meaningful (Suhr, 2006). Ullman (2006) emphasises that the two main reasons for modifying a model are firstly to test hypotheses in the theoretical work and secondly to improve the model fit. Model modification is often referred to in literature as a means to adjust a specified and estimated model by either fixing parameters that were free or freeing parameters that were fixed (Suhr, 2006). The three main forms of model modification are the chi-square difference, the Wald test and the Lagrange multiplier. Each approach is asymptotically equivalent but they approach model modification differently. The chi-square difference test is applied when the initial model is a subset of the new larger model (Ullman, 2006). In these cases, the χ^2 value for the larger model is subtracted from the χ^2 value for the smaller model and the difference is also a χ^2 model with the degrees of freedom equal to the difference between the degrees of freedom in the two models. This however can only be applied when the data are normally distributed. When data are non-normal then the Satorra-Bentler scaled chi-square is applied (Satorra and Bentler, 2001). This approach includes a scaling correction for the χ^2 statistic when distributional assumptions are violated and it has been shown in literature to be the most reliable test statistic for evaluating mean and covariance structure models under various distributions and sample sizes (Byrne, 2010).

The Lagrange multiplier test also compares nested models however it only requires estimation of one model (Ullman, 2006). The Lagrange multiplier provides information about the amount of χ^2 change that would occur if parameters that were formerly fixed were free in a specified model (Hoyle, 1995). As a result, researchers often use this information to conduct a series of model modifications. Hox and Bechger (2007) indicates that at each step a parameter is freed that produces the largest improvement in fit until an adequate model fit is achieved. The Lagrange

multiplier test assesses whether parameters should be added to the model while the Wald test determines whether parameters should be deleted from the model. The Wald test thus provides information regarding the change in χ^2 that would occur if formerly free parameters were now fixed to zero (Hoyle, 1995). Both the Lagrange multiplier test and the Wald test are stepwise procedures thus type I errors are often inflated however there are no available adjustments as yet (Ullman, 2006). In addition, the order in which the parameters are freed can affect the significance of the remaining parameters thus MacCallum (1986) suggested doing the Lagrange multiplier test before the Wald test. A further disadvantage of these tests is that they examine the overall changes in χ^2 not changes in the individual parameter estimates. Lastly if the hypothesised model is incorrect then testing the modification indices by themselves may not be sufficient to reveal the true model (Ullman, 2006).

Once the overall model fit has been assessed, the measurement model fit can be examined by first assessing the indicator loadings for statistical significance then by examining the construct's reliability and variance. Furthermore in order to assess the structural fit of the model, one needs to test the significance of the estimated coefficients. The standard error and t value for each coefficient will need to be examined as well as the coefficient of determination, R^2 , for each endogenous equation. The coefficient of determination indicates the amount of variation of the dependent variable accounted for by the independent variables and thus depicts a measure of fit for each structural equation.

2.1.3 Mathematical Notation

The following section gives a mathematical overview of SEM. SEM procedures minimise the difference between the sample covariances and the covariances predicted by the model. The fundamental hypothesis for these procedures is that the covariance matrix of the observed variables is a function of a set of parameters

(Bollen, 1989). The covariance matrix is generally given by the following form:

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \tag{2.6}$$

where:

Σ_{xy} contains the covariances between x and y;

Σ_{yy} contains the covariances among the y's; and

Σ_{xx} contains the covariances among the x's.

If the model was correct and one knew the parameters then the population covariance matrix would be exactly reproduced. As a result SEM involves testing the following hypothesis:

$$\Sigma = \Sigma(\theta) \tag{2.7}$$

where:

Σ is the population covariance matrix of the observed variable;

θ is a vector that contains the model parameters; and

$\Sigma(\theta)$ is the covariance matrix written as a function of θ.

Suppose that one has a model with m exogenous constructs, n endogenous constructs, p exogenous indicators and q exogenous indicators then the basic form of the structural equation model is given by:

$$\eta = \beta\eta + \Gamma\xi + \zeta \tag{2.8}$$

where:

$\beta_{(n\times n)}$ is a coefficient matrix of latent endogenous variables;

$\eta_{(n\times 1)}$ is a vector of latent endogenous variables;

$\Gamma_{(n\times m)}$ is a coefficient matrix of latent exogenous variables;

$\xi_{(m\times 1)}$ is a vector of latent exogenous variables; and

$\zeta_{(n\times 1)}$ is a random vector of residuals.

The correlations among the exogenous constructs of the structural equation model are represented by a $\Phi_{(m\times m)}$ correlation matrix of ξ while the correlations among the error terms of the endogenous constructs are represented by a $\Psi_{(n\times n)}$ correlation matrix of ζ.

The basic equations for the measurement model are given by:

$$x = \Lambda_x \xi + \delta \tag{2.9}$$

$$y = \Lambda_y \eta + \epsilon \tag{2.10}$$

where:

$x_{(p\times 1)}$ is a vector of observed indicators of ξ;

$y_{(q\times 1)}$ is a vector of observed indicators of η;

$\Lambda_{x(p\times m)}$ is a coefficient matrix relating x to ξ;

$\Lambda_{y_{(q\times n)}}$ is a coefficient matrix relating y to η;

$\delta_{(p\times 1)}$ is a vector of measurement errors of x; and

$\epsilon_{(q\times 1)}$ is a vector of measurement error of y.

The correlations among the error terms of the exogenous indicators in the measurement model are given by a $\Theta_{\delta(p\times p)}$ correlation matrix of δ while the correlations among the error terms of the endogenous indicators in the measurement model are given by a $\Theta_{\epsilon(q\times q)}$ correlation matrix of ϵ.

The main assumptions of a structural equation model are:

1. $\mathrm{E}(\eta) = \mathrm{E}(\xi) = \mathrm{E}(\delta) = \mathrm{E}(\epsilon) = \mathrm{E}(\zeta) = 0$;

2. ξ is uncorrelated with ζ;

3. ϵ is uncorrelated with η and ξ;

4. δ is uncorrelated with η and ξ;

5. δ, ϵ, ζ are mutually uncorrelated; and

6. $I - \beta$ is non-singular where I is the identity matrix.

Further details surrounding the mathematics in SEM are provided by Bollen (1989).

2.1.4 Mediation Analysis

This section gives an overview of the methods proposed in literature to assess the significance of mediated effects in structural equation models. Mediation analysis is often conducted by researchers in order to indirectly examine the effect of an independent variable on a dependent variable through at least one intervening variable (Preacher and Hayes, 2004). Thus mediation explains the process of "why" and "how" a cause and effect occurs (Wu and Zumbo, 2008). Mediation analysis involving one mediator is called simple mediation while mediation that involves multiple mediating variables is called multiple mediation (Preacher and Hayes, 2008).

The following figure depicts the simple mediation model proposed widely in literature (Preacher and Hayes, 2004; Mallinckrodt, Abraham, Wei and Russell, 2006; MacKinnon, Fairchild and Fritz, 2007a; Preacher and Hayes, 2008; Hayes, 2009; Hayes and Scharkow, 2013):

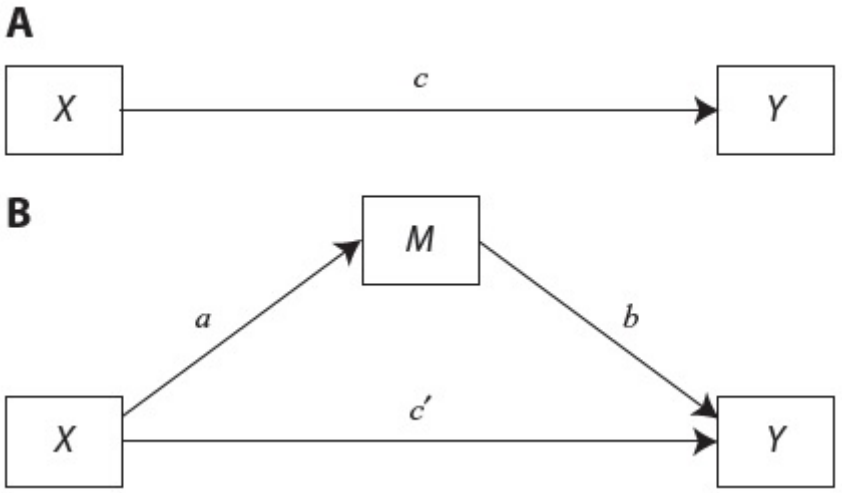


Figure 2.2: Panel A: Illustration of direct effect where X affects Y. Panel B: Illustration of mediation design where X indirectly affects Y through M.

Figure 2.2 illustrates the variables X, M and Y in rectangles and the arrows representing the relationships between these variables (MacKinnon *et al.*, 2007a). The notation widely used in literature (Preacher and Hayes, 2004; Mallinckrodt *et al.*, 2006; MacKinnon *et al.*, 2007a; Preacher and Hayes, 2008; Hayes, 2009; Hayes and Scharkow, 2013) depicts a which denotes the relation of X to M, b which denotes the relation of M to Y adjusted for X and c' which denotes the relation of X to Y adjusted for M (MacKinnon *et al.*, 2007a). As a result, c' illustrates the direct effect of X while the product of a and b denotes the indirect effect of X on Y through M (Hayes, 2009). In addition, c denotes the total effect of X on Y , not controlling for M (Preacher and Hayes, 2004). According to James and Brett (1984), complete mediation occurs when the effect of X on Y decreases to zero when the mediator M is included in the analysis. However, partial mediation occurs when the effect of X on Y decreases by a nontrivial amount, but not to zero (Preacher and Hayes, 2004).

The following figure illustrates the multiple mediation model proposed in literature (Preacher and Hayes, 2008):

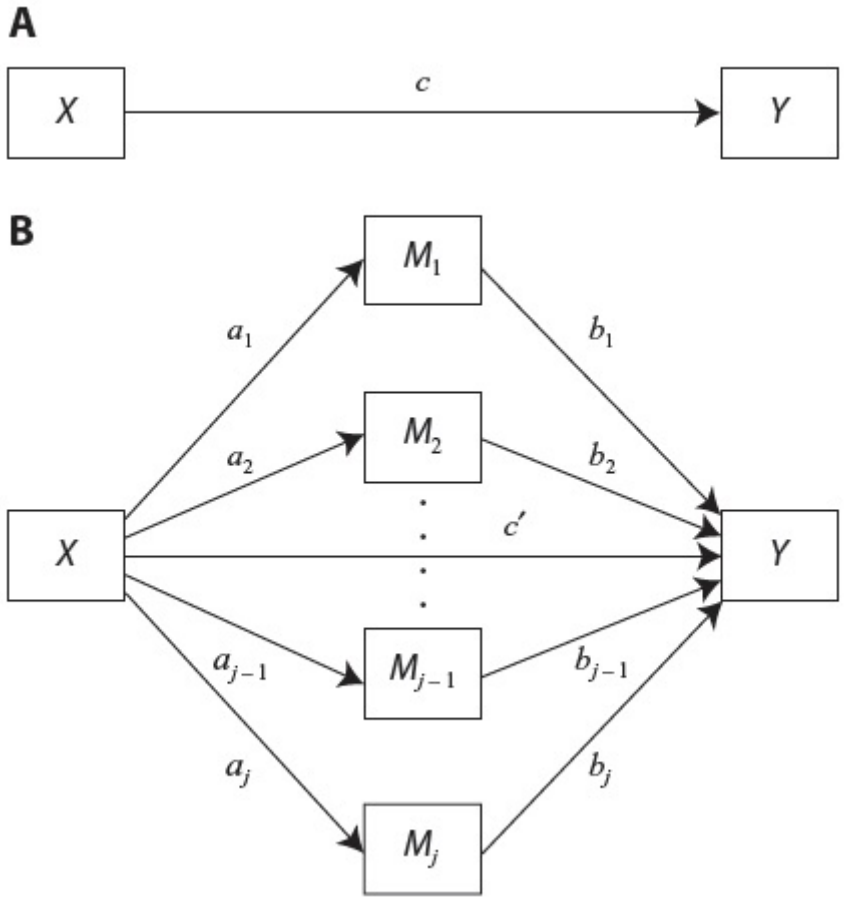


Figure 2.3: Panel A: Illustration of direct effect where X affects Y. Panel B: Illustration of multiple mediation design where X indirectly affects Y through $M_1, M_2, ..., M_j$.

Investigating multiple mediation has been shown in literature to be more complex than investigating simple mediation (Preacher and Hayes, 2008). The specific indirect effect through an intervening variable (say M_1) in multiple mediation design is not the same as the indirect effect through M_1 alone (Preacher and Hayes, 2008). In the multiple mediation model, the specific indirect effect through M_1 illustrates how M_1 mediates the relationship between X and Y conditional on the inclusion of all of the other mediators in the model (Preacher and Hayes, 2008).

Multiple mediation needs to be assessed in two steps:

1. Investigate whether the total indirect effect is significant;

2. Test hypothesis regarding individual intervening variables in the multiple mediation model, that is, assess whether a specific indirect effect is significant (Preacher and Hayes, 2008).

Various different approaches to assess the significance of the indirect effects have been proposed in literature (Leth-Steensen and Gallito, 2016). The most commonly used approach is the causal steps approach developed by Baron and Kenny (1986). Four regression-based conditions are involved in assessing mediation when the causal steps approach is used (Baron and Kenny, 1986). First, a significant relationship between the independent variable (X) and dependent variable (Y) is required. Secondly, a significant relationship between the independent variable (X) and the hypothesised mediator (M) is required. Thirdly, the mediator (M) needs to be significantly associated with the dependent variable (Y) when both the independent variable (X) and the mediator (M) are predictors of the dependent variable (Y). Lastly, the coefficient (c) relating the independent variable (X) to the dependent variable (Y) must be larger than the coefficient (c') relating the independent variable (X) to the dependent variable (Y) when the mediator is present in the model. Various simulation studies in literature comparing the different approaches to assessing the significance of indirect effects have illustrated that the causal steps approach seems to have the lowest statistical power (MacKinnon, Lockwood, Hoffman, West and Sheets, 2002; Fritz and MacKinnon, 2007). Despite the limitations of the causal steps approach it is still widely used in literature due to the simplicity of the approach (Hayes, 2009).

Another technique that is widely used in literature to examine the significance of the indirect effects, is the product of coefficients approach, which is commonly referred to as the Sobel test (Sobel, 1982). The Sobel test depends on the estimate of the standard error of ab. The ratio of ab to its standard error is used to test the null hypothesis that the indirect effect is zero with the p-value derived from the standard normal distribution (Hayes, 2009). The Sobel test has primarily been used in literature in conjunction with the Baron and Kenny approach (Hayes, 2009). The major flaw of the Sobel test is the requirement of normality since the

sampling distribution of ab tends to be asymmetric with nonzero skewness and kurtosis (Bollen and Stine, 1990).

MacKinnon, Fritz, Williams and Lockwood (2007b) developed the distribution of product approach while Shrout and Bolger (2002) developed the currently popular bootstrapping approach. Both of these methods overcome the limitations of the Sobel test as they adequately capture the skewed, asymmetric nature of the sampling distribution (Leth-Steensen and Gallito, 2016). The general consensus arising from simulation work conducted in literature (MacKinnon, Lockwood and Williams, 2004; Preacher and Hayes, 2008; Hayes and Scharkow, 2013) is that the bias corrected bootstrapping approach developed by Shrout and Bolger (2002) is the best method to examine whether the indirect effects are significant.

Bias corrected bootstrapping of the confidence intervals for the indirect effects involves taking multiple repeated samples with replacement from the dataset of interest (Leth-Steensen and Gallito, 2016). For each bootstrapped sample, the path or structural equation model is refit and estimates of all the parameters are retained (Leth-Steensen and Gallito, 2016). Once a sample is constructed, the product of the path coefficients, ab, is noted (Hayes, 2009). This procedure is repeated for a total of k times where k is a number chosen by the researcher. After all of the bootstrapped results are collected, the set of k values obtained for each effect (for example, ab in the case of the indirect effect) are sorted and the lower and upper percentile values are determined (Leth-Steensen and Gallito, 2016). For a standard 95% confidence interval, these values represent the 2.5th and 97.5th percentile values. The bias corrected confidence intervals, however, involve a slight adjustment to these percentile values (Leth-Steensen and Gallito, 2016). Hayes (2009) indicated that a researcher can state that an indirect effect is not zero with (1-α)% confidence if zero does not lie between the lower and upper bound of the confidence interval. This principle is the same as rejecting the null hypothesis that the true indirect effect is zero at the α level of significance.

2.2 Comparison of SEM and multiple linear regression

The purpose of this section is to emphasise the differences and similarities between structural equation modelling and multiple linear regression as these are the two techniques that will be compared in the analysis. Multiple linear regression is one of the most commonly used multivariate dependence techniques that is applicable in various research areas. It is often referred to as a first generation multivariate method as it evaluates constructs and relationships between constructs. Multiple linear regression is a multivariate statistical technique that examines the linear correlations between two or more independent variables and a single dependent variable (Tabachnick and Fidell, 2007).

The regression line for p explanatory variables $X_1, X_2, ..., X_p$ is defined by:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon \tag{2.11}$$

In this equation $\beta_0, \beta_1, ...\beta_p$ are the parameter estimates for the p independent variables, Y_i represents the measurements on the dependent variable and ϵ represents the error term which makes allowance for variability in the sample as a result of variables that have not been entered into the model. When the independent variable X changes by one unit, the value of estimate β_1 indicates the amount the dependent variable, Y, changes while the other independent variables remain fixed (Alavifar, Karimimalayer and Anuar, 2012). Thus the variation in the dependent variable is determined by a weighted combination of the values of the independent variables.

According to Kline (2011) the assumptions of multiple linear regression analysis are stringent and are described as follows:

1. Regression weights describe linear relations only;

2. Statistical tests assume that residuals are normally distributed with equal variances across all levels of predictors;

3. The scores on the predictors are assumed to be perfectly reliable and are measured without error. Any serious violations of this assumption can result in bias which not only affects the regression weights of predictors measured with error but also those of other predictors;

4. Omitted predictors are assumed to be uncorrelated with measured predictors.

Multiple linear regression analysis is often considered to be a special case of SEM since it assumes that all of the explanatory variables are measured without error.

Goldberger (1973) provides three situations where SEM has advantages over regression analysis:

1. When observed variables contain measurement errors;

2. When important explanatory variables have not been measured;

3. When there is interdependence among observed response variables.

Hoyle (1995) emphasises that SEM is more flexible and comprehensive than multiple linear regression since it provides a means of controlling not only the extraneous or confounding variables but also the measurement error. Furthermore Farrel (1994) describes how SEM has an advantage over regression analysis when dealing with longitudinal data which can be easily interpreted and analysed using SEM. SEM is also more flexible as it allows one to statistically test apriori theoretical and measurement assumptions against empirical data. In addition, SEM has another advantage over regression analysis since one can model many dependent (latent or observed) variables in the same research model and analyse all paths simultaneously rather than one at a time. As a result, SEM is often referred to as a second generation multivariate method since simultaneous analysis of all variables in the model is allowed. Moreover, SEM also provides more rigorous variance analysis (Bollen, 1989) and the researcher is allowed to include both the common

variance as well as the specific and error variance in the model (Hair, Anderson, Tatham and Black, 1998).

Although SEM has various advantages over regression analysis, the areas of diagnostics and goodness of fit statistics still require lots of work. Examining the overall model fit for SEM is not as simple as with other multivariate dependence techniques since there is no single statistical test that best describes the strength of the models' predictions (Bentler and Bonnet, 1980). As a result researchers need to use many indices together in order to examine the model fit. Furthermore regression diagnostics are much more superior than the diagnostics applied when using SEM. Multicollinearity can be examined by the tolerance and the variance inflation factor in regression analysis as described by Kleinbaum, Kupper and Muller (1988). Conversely Bollen (1989) indicates that multicollinearity poses great difficulties for measurement models. Furthermore detection of outliers is simple in regression however it becomes problematic in SEM since the residual covariance or correlation matrix is used. Obtaining the intercept in SEM is also more complex than in regression analysis (Bollen, 1989). Moreover, in structural equation models each equation represents a causal link while in regression analysis each equation depicts the conditional mean of a dependent variable as a function of explanatory variables (Goldberger, 1973). According to Mohanlal (1998), regression analysis thus might be incapable of estimating structural equation models and the suitability of each method depends on the research problems that are being analysed.

2.3 Hypertension

In order to apply SEM and multiple linear regression to understand the risk factors associated with hypertension, it is necessary to understand the disease as well as the patterns and determinants of hypertension from previous research.

2.3.1 Definition

Blood pressure is a measure of how one's blood presses against the walls of the arteries as it is pumped around the body by the heart (NICE, 2011). Blood pressure is comprised of two measurements: SBP which is taken when the heart is beating and pumping blood and diastolic blood pressure (DBP) which is taken when the heart is filling up with blood between beats. It is measured in millimetres of mercury (mmHg) and is measured as systolic over diastolic blood pressure. The national health and nutrition survey conducted in South Africa in 2012 stated that hypertension is determined by a SBP $\geq$ 140 mmHg and/or DBP $\geq$ 90 mmHg ($\geq$ 140\90 mmHg). In order to diagnose hypertension, two or more elevated readings within a minimum period of one week are needed (Shisana, Labadarios, Rehle, Simbayi, Zuma, Dhansay, Reddy, Parker, Hoosain, Naidoo, Hongoro, Mchiza, Steyn, Dwane, Makoae, Maluleke, Ramlagan, Zungu, Evans, Jacobs, Faber and SANHANES-1 Team, 2014).

2.3.2 Risk factors associated with hypertension

As with other non-communicable diseases, commonly known as chronic diseases, there are various behavioural and lifestyle risk factors associated with hypertension that are well documented in literature. Due to economic transition and urbanisation, these risk factors have become more evident in the last three decades. The WHO stated in 2012 that up to 25% of the global disease burden is currently affected by lifestyle and behaviour and poorer countries are most at risk (WHO, 2012).

Modifiable bio-behavioural risk factors that affect hypertension prevalence include alcohol consumption, body mass index (BMI), smoking, resting heart rate, physical activity and psychosocial stress. The BMI is a measure of the weight relative to the height and gives an approximation of the total body fat while the resting heart rate refers to the number of heartbeats per minute during periods of inactivity or relaxation. Extensive research has shown a positive relationship between a higher BMI and an elevated blood pressure (Van Rooyen, Kruger, Huisman,

Wissing, Margetts, Venter, and Vorster, 2000; Mishra, Arnold, Semenov, Hong and Mukuria, 2005; Brummett *et al.*, 2011; Papathanasiou, Zerva, Zacharis, Papandreou, Papageorgiou, Tzima, Georgakopoulos and Evangelou, 2015; Keetile, Navaneetham and Letamo, 2015). Similarly the more one consumes alcohol, the greater the risk of an elevated blood pressure (Tsuruta, Adachi, Hirai, Fujiura and Imaizumi, 2000; Steyn, Bradshaw, Norman and Laubscher, 2008; Brummett *et al.*, 2011). Furthermore a positive association between the heart rate and hypertension has repeatedly been observed in literature (Chaix *et al.*, 2010; Brummett *et al.*, 2011; Cois and Ehrlich, 2014). Numerous studies over the years have also shown a significant positive association between psychosocial stress and blood pressure (Jorde, Williams and Kuida, 1986). On the contrary, physical activity tends to be associated with a lower blood pressure (Hu, Barengo, Tuomilehto, Lakka, Nissinen and Jousilahtio, 2004; Papathanasiou *et al.*, 2015). Various studies have shown that smoking either has a small effect or does not have a significant effect on hypertension (Primatesta, Falaschetti, Gupta, Marmot and Poulter, 2001; Papathanasiou *et al.*, 2015) and in some cases smokers have a lower blood pressure than non-smokers (Green, Jucha, Luz and Israel, 1986). These behavioural factors are often disproportionate across socio-economic levels and thus prove to be apt for monitoring the effects of SES on blood pressure.

2.4 Association between socio-economic status and health outcomes

A recent study based on a sample of 30 to 79 year old French adults used the path analysis model to examine the extent to which hypertension risk factors mediated the relationships between individual or neighbourhood socio-economic variables and SBP (Chaix *et al.*, 2010). This study showed that BMI, resting heart rate and other health behaviours often mediate the association between socio-economic variables and SBP.

Brummett *et al.* (2011) performed a cross-sectional study using data of individuals aged 24 to 35 years from Wave IV (2007 to 2009) of the National Longitudinal Study of Adolescent Health. Linear regression models were first applied to the data in order to estimate the independent relationships between the predictor variables and SBP, first adjusted for gender, hypertension medication and age and thereafter adjusted for all the independent variables in the study. Using path models, the researchers found that increased BMI, waist circumference and higher resting heart rate were significant mediators of the association between lower SES and high blood pressure.

Bardenheier *et al.* (2013) conducted a cross-sectional study on 2,230 older adults ($\geq$ 50 years) without diabetes from a sample of the 2001 to 2006 National Health and Nutrition Examination Survey respondents. The aim of their analysis was to use SEM to test a hypothesised model of causal pathways related to pre-diabetes in older adults. Initially BMI and total cholesterol were included in their model however after excluding these variables, their best-fit model included SES, physical exercise and an unhealthy diet. The results from this study concluded that a large waist circumference had the strongest direct effect on pre-diabetes. Furthermore the direct effect of SES on pre-diabetes illustrates that there is some mediation by certain factors that were not considered in the analysis.

Ploubidis *et al.* (2013) used data from a community survey conducted in Kenya on 4,314 participants in order to evaluate the extent to which SES and later life prevalence of diabetes, visual impairment and hypertension are mediated by health-related behaviours. This study depicted that the accumulation of material resources was positively related to diabetes and hypertension however material resources and education had a negative association with the prevalence of visual impairment. Health related behaviour such as alcohol consumption and smoking did not mediate the observed relationships.

A study was conducted in 2013 on 2,266 Japanese workers to investigate the association of SES with nutrient intake and health outcomes (Miyaki *et al.*, 2013). Multiple linear regression was first performed to examine the relationship between

nutrient intake and confounding factors. From this study it was observed that household income and education levels were significantly associated with nutrient intake and depression scales. Path analysis was then performed to demonstrate the impact of nutrient intake on health outcomes. The structural equation model applied depicted that folate intake strongly mediates the effects of education level on the depression scale.

Cois and Ehrlich (2014) researched the degree to which SES is associated with blood pressure and whether bio-behavioural risk factors mediate this relationship. This study was performed on the first phase of the National Income Dynamics Study (NIDS) data conducted in South Africa on more than 15,000 adults in 2008. SEM was applied in this cross-sectional analysis in order to evaluate the effect of SES on SBP and DBP and measure the role of bio-behavioural risk factors in explaining the observed relationships. After adjusting for confounding factors, this study found higher income and education to be independently associated with a higher DBP in males. In females however, a higher education was associated with lower SBP and DBP while a higher income predicted lower SBP. Furthermore BMI was found to be the strongest mediator of the indirect effect of SES on blood pressure. The BMI in conjunction with physical activity, alcohol consumption, resting heart rate and smoking thus substantially contributed to the mediation of the observed relationship in males. Conversely, factors that were not considered in the research appeared to play a more prominent role in females.

Kastorini *et al.* (2015) examined how socio-economic factors influence modifiable cardiovascular disease risk factors and in particular acute coronary syndrome (ACS) or ischaemic stroke presence. One thousand individuals were enrolled in this case-control study; 250 ACS patients and 250 control subjects as well as 250 stroke patients and 250 control subjects. First a logistic regression analysis was performed so that the relative odds of having ACS or stroke according to social and economic factors could be determined. Path analysis using SEM was then performed to reveal the mediating effects of social and economic factors on modifiable cardiovascular risk factors in relation to the outcome. After adjusting for confounding factors, the logistic regression analysis illustrated that financial

status satisfaction and occupation type were inversely associated with ACS or stroke presence while the education level did not have significant associations with ACS or stroke presence. Conversely, the path analysis revealed that the education level influences the financial status satisfaction and occupation type and indirectly affects the probability of developing a cardiovascular disease.

2.5 Summary

This chapter examined the research and literature on the history of SEM and the basics required for understanding and applying SEM. In addition, a comparison of the differences and similarities between multiple linear regression and SEM was reviewed. Furthermore, an explanation of hypertension and the risk factors associated with hypertension was provided. Studies that have focused on understanding the relationship between SES and health outcomes were then reviewed. An overview of the methodology applied in the analysis is discussed in the next chapter.

Chapter 3

Methodology

The following chapter provides an overview of the methodology that will be applied in the analysis. A discussion of the data source and sampling design is provided in section 3.1. The measures considered in the analysis and the hypothesised structural equation model are then explained in section 3.2 while section 3.3 details the techniques applied in the analysis.

3.1 Data source and sampling design

In this study, data from the adult subsample and household information of the third wave of data from the National Income Dynamics Study (NIDS) was used. The NIDS survey uses a combination of household and individual questionnaires in order to gain insights about the household, adults, children and proxies. The household questionnaire was completed by the oldest women in the household and took approximately 39 minutes in non-agricultural households and 50 minutes in agricultural households to complete. The adult questionnaire was completed by all individuals 15 years and older and took approximately 38 minutes per adult to complete. The survey was a face-to-face longitudinal panel survey of individuals residing in South Africa and their households (De Villiers *et al.*, 2013).

The aim of the survey was to provide insights into the conditions of South Africans and how this impacts on their well-being over time. This was measured by capturing information regarding wealth creation, demographic dynamics, social heritage, the impact of life events, social capital and intergenerational developments and access to cash transfers and social services. A stratified two phase cluster sampling plan

was used in sampling households across 400 primary sampling units in the first phase of the survey which was conducted in 2008. The survey process is described in more detail by Leibbrandt, Woolard and De Villiers (2009). Trained fieldworkers conducted the interviews and collected anthropometric data such as height, weight and waist measurements from each of the individuals in the households. The target population of the base survey included all private households across South Africa and residents in convents, monasteries and workers' hostels, excluding collective living quarters such as prison, military barracks, hospitals, old age homes and student hostels. 69% of households responded in the first phase of the study while 93% of the individuals in these households responded in the first phase of the study (Leibbrandt *et al.*, 2009). The third phase of the survey re-interviewed respondents of phase one and phase two and transpired between April and December 2012. The third phase comprised of 32,633 successfully interviewed individuals of which 18,710 individuals were 15 years and older. The NIDS dataset is publicly available for research purposes and was thus used in this study (Southern Africa Labour and Development Research Unit, 2013).

3.2 Measures used in the analysis

3.2.1 Demographic variables

In this study age and gender were the only demographic variables used. (Chaix *et al.*, 2010; Brummett *et al.*, 2011; Ploubidis *et al.*, 2013; Miyaki *et al.*, 2013; Cois and Ehrlich, 2014; Kastorini *et al.*, 2015). Age was treated as a continuous numerical variable while gender was a categorical variable. Each of the analysis were performed separately for males and females.

3.2.2 Socio-economic variables

Kaplan and Keil (1993) stated that SES covers a wide range of measures including education, income, living conditions, occupation, income inequality and various

other socio-economic aspects of life. In this study individual education measured as years of complete schooling was used as a measure of SES. Education is often used in epidemiologic studies as a measure of SES since it is often fixed after young adulthood and is unlikely to be impacted by poor health in adulthood (Kaplan and Keil, 1993). Education was an ordinal variable indicating whether an individual had no schooling, primary education, secondary education or tertiary education. In addition, monthly income was calculated from a wide array of questions related to income as a measure of SES. Although income can decrease as a result of poor health, it is a vital measure as it measures the purchasing capacity of goods and healthcare (Kaplan and Keil, 1993; Colhoun, Hemingway and Poulter, 1998). Income was treated as a continuous numerical variable in the analysis. Similarly employment status was used as a measure of SES in order to provide insight on labour market participation. Occupation is often categorized as a measure of SES in epidemiological studies as it often reflects status, power and prestige however there were only one third of individuals employed thus employment status was instead used as a measure of SES. The association between employment status and health is vital to measure, as it is important to distinguish between those that are unemployed but are searching for work and those that are unemployed but are not searching for work due to health reasons or for various other reasons (Kaplan and Keil, 1993). Employment status was a categorical variable indicating whether an individual was not economically active, unemployed but discouraged (those who were actively seeking employment), unemployed strict (those that were not actively seeking employment) or whether the individual was employed. Ownership of various household goods such as a radio, hifi, sewing machine, motor vehicle, bakkie, motorcycle, bicycle, computer, camera and cellphone were also used as a measure of SES. Although these measures may be highly correlated with education and income, they also emphasise important lifestyle differences (Kaplan and Keil, 1993). The ownership of a radio, hifi, sewing machine, motor vehicle, bakkie, motorcycle, bicycle, computer, camera and cellphone were each dichotomous variables indicating whether an individual owned the good or did not own the household good.

3.2.3 Blood pressure and resting heart rate

Using an automated blood pressure monitor, two readings of blood pressure and the heart rate were measured in the left arm with a five minute rest period between the readings. Measurements were included in the data if the SBP was between 80 and 240 mmHg and the DBP was greater than or equal to 30 mmHg (Cois and Ehrlich, 2014). Heart rate measurements were included if they were greater than or equal to 30 bpm and less than or equal to 150 bpm (Cois and Ehrlich, 2014). Each blood pressure measurement and each resting heart rate measurement was treated as a continuous numerical variable and was used in the study. The use of antihypertensive medication was also assessed from a single interview question asking participants if medication was currently used to control their high blood pressure (Chaix *et al.*, 2010; Brummett *et al.*, 2011; Ploubidis *et al.*, 2013; Cois and Ehrlich, 2014). The use of antihypertensive medication was a dichotomous variable indicating whether an individual was currently using medication to control high blood pressure or not.

3.2.4 Bio-behavioural risk factors

Physical activity, alcohol consumption, emotional well-being, smoker status, resting heart rate and BMI were included as possible mediators in the relationship between SES and blood pressure. Qualified nurses provided training to field workers on how to take anthropometric measurements such as height, weight and waist measurements accurately (Leibbrandt *et al.*, 2009). NIDS fieldworkers took two weight and height measurements per individual. A third measure was taken if the weight measurement was more than one kilogram apart or the height measurement was more than one centimetre apart (De Villiers *et al.*, 2013). In this study, implausible height and weight measurements were excluded (height < 100 cm or > 210 cm, weight < 20 kg or > 200 kg) (De Villiers *et al.*, 2013) and the first two measurements were used to calculate the BMI (weight in kilograms\height in metres^2) based on the first reading and BMI based on the second reading. The

third readings were not considered as very few people had a third reading for weight or height. BMI was treated as a continuous numerical variable in each analysis. Physical activity was assessed based on a question related to the frequency of exercise. Physical activity was an ordinal variable indicating whether an individual had never exercised, had exercised less than once a week, exercised once a week, twice a week or three or more times in a week. The smoker status was an ordinal variable split into four categories: individuals who do not smoke, those that smoke 5 or less cigarettes per day, those that smoke between 6 and 10 cigarettes per day and those that smoke more than 10 cigarettes per day. The assessment of alcohol consumption was based on the number of alcoholic drinks had in a day and the frequency of alcohol usage. Alcohol consumption was thus an ordinal variable indicating whether an individual was a non-drinker, had one or two standard drinks in the day, had three to four standard drinks, had five to six standard drinks or had seven or more standard drinks per day. The emotional well-being of an individual was assessed by the 10 Likert scale questions aimed at measuring the general well-being of individuals during the past week (Chaix *et al.*, 2010; Brummett *et al.*, 2011; Cois and Ehrlich, 2014; Kastorini *et al.*, 2015). Each of the emotional wellbeing questions were treated as ordinal variables.

3.2.5 Hypothesised model for structural equation model

Based on epidemiological and biological evidence (Zhang and Kesteloot, 1999; Appel, 2003), the model in Figure 3.1 depicts each of the variables that are considered to be related in the proposed SEM analysis:

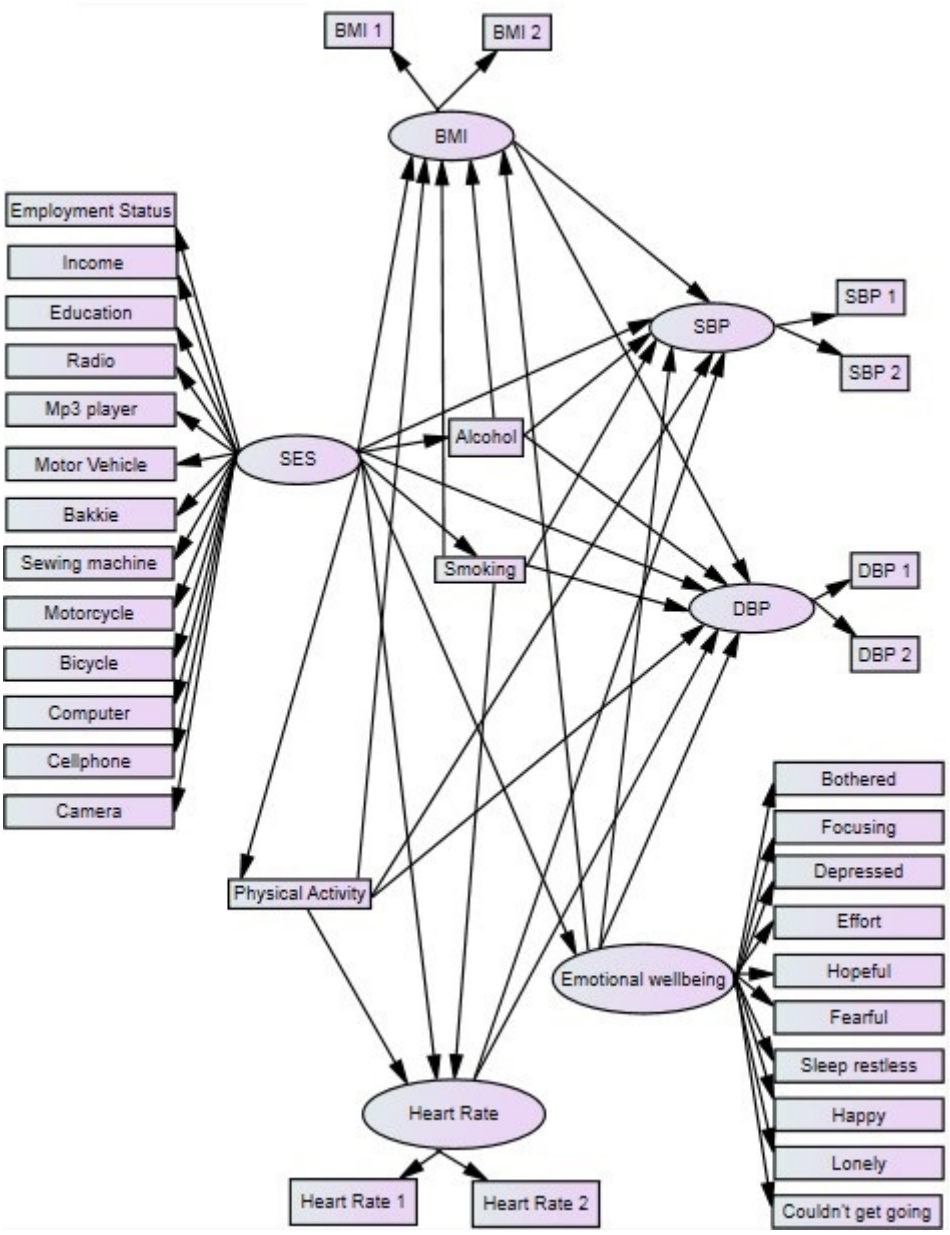


Figure 3.1: Hypothesised pathways between SES and blood pressure

In the proposed model, squares and circles depict the observed and latent variables respectively, and the arrows denote the hypothesised effects. The use of antihypertensive medication and age have not been included in the diagram but will be included as possible confounders in the model (Miyaki *et al.*, 2013; Ploubidis *et al.*, 2013; Bardenheier *et al.*, 2013; Cois and Ehrlich, 2014; Keetile *et al.*, 2015). Confounding variables are variables that may affect the dependent variable however we are not particularly interested in them (Becker, 2005). As a result when one controls for a variable, the aim is to balance its effect across subjects so that it can

be ignored and the relationships of interest can be studied. BMI, SES, emotional well-being, heart rate and blood pressure will be introduced as latent variables in the model to reduce the bias caused by measurement error. For the blood pressure, heart rate and BMI latent variables, the observed multiple readings will be considered as the manifest variables or indicators while each of the 10 Likert scale questions measuring the general well-being of individuals during the past week will be considered as indicators of emotional well-being. The income, education, employment status and ownership of various household goods will be considered as indicators of SES (Kaplan and Keil, 1993). Latent variables are not measured directly but are analytically inferred from other directly measured variables. In the case of blood pressure measurements, this approach has been shown, under broad assumptions to be more effective than the common practice of averaging multiple readings (Batista-Foguet, Coenders and Ferragud, 2001).

The proposed model suggests an effect of SES on DBP and SBP mediated by BMI, physical activity, heart rate, smoking, alcohol consumption and emotional well-being. Mediation occurs when a dependent variable is indirectly affected by a predictor through at least one intervening variable (Preacher and Hayes, 2008). A mediator is thus a third variable that links a cause and effect and explains the process of "why" and "how" a cause and effect occurs (Wu and Zumbo, 2008). BMI is also assumed to be affected by alcohol consumption, smoking, emotional well-being and physical activity while the heart rate is affected by smoker status and physical activity. Estimated model coefficients will be used to decompose the total effects of SES on blood pressure into mediated or indirect effects and unexplained or direct effects. Unexplained effects are effects unrelated to variation in the considered mediators and are depicted by the direct paths connecting the SES indicators to blood pressure. The mediated effects are the effects explained by variations in BMI, heart rate, physical activity, smoking, alcohol consumption and emotional well-being which are represented by the indirect paths connecting SES to the blood pressure through the different factors. The magnitude of the mediated effects will be measured as the product of the regression coefficients in the considered paths (Preacher and Kelley, 2011). The model coefficients will be

estimated before and after adjusting for possible confounders in the model such as age and the use of antihypertensive medication (Ploubidis *et al.*, 2013; Miyaki *et al.*, 2013; Bardenheier *et al.*, 2013; Cois and Ehrlich, 2014; Keetile *et al.*, 2015).

3.3 Techniques applied in analysis

Statistical analysis was performed in the SPSS version 23, SPSS AMOS version 23 (Arbuckle, 2014) as well as SAS version 9.4 (SAS Institute Inc., 2008). Sample characteristics of the dataset were described as the median and interquartile range for continuous variables and frequency or percentage for categorical variables. The median and interquartile range measure the central tendency and spread, respectively, but are robust against outliers and non-normal data. As a result, these statistics will only be used to describe the continuous variables as some continuous variables have an asymmetrical distribution (Altman, Gore, Gardner and Pocock, 1983). The data was first screened for the following characteristics: influential outliers, missing data, homoscedasticity, linearity, normality and multicollinearity by using SPSS version 23 before performing any analysis. The data was assessed for multivariate normality as this is a critically important assumption when conducting any SEM analysis (Arbuckle, 2014). Furthermore, the data was assessed for multivariate outliers, which depict cases that are vastly different from other cases in the dataset. A univariate outlier has an extreme score on a single variable while a multivariate outlier has an extreme score on two or more variables (Kline, 2011). The squared Mahalanobis distance was computed for each case in order to detect multivariate outliers. According to Byrne (2010), this statistic measures the distance in standard deviation units between a set of scores for one case and the sample means for all cases.

Once the assumptions for SEM were assessed as discussed in subsection 2.1.1, SEM was applied to examine the relationship between SES and blood pressure. Firstly, the direct relationship between SES and blood pressure was assessed. Secondly, behavioural risk factors were included in the model as mediators to determine if they

can explain the observed relationship between SES and systolic and diastolic blood pressure. This model was specified as described in Figure 3.1. ADF estimation was used to build the model and to perform model modification. In our study, the data did not only include continuous variables, but also included categorical or ordinal variables. Literature to date supports the notion that when the number of categories is large and the data approximates a normal distribution then failure to address the ordinality of the data is negligible (Muthén and Kaplan, 1985; Byrne, 2010). Furthermore Bentler and Chou (1987) emphasised that in cases of normally distributed categorical variables, continuous methods can be used without concern when a categorical variable has four or more categories. Recent literature has supported these findings and shown that the χ^2 statistic is impacted most by the two category response and becomes less influenced as the number of categories increases (Green *et al.*, 1997). As a result, in each SEM analysis, each of the ordinal variables were treated as continuous variables.

The χ^2 statistic along with the relevant p value and degrees of freedom were then examined to assess the overall fit of the model. The ratio of the chi-square to the degrees of freedom (χ^2/df) was also used to reduce the sensitivity of the χ^2 statistic to the sample size. In addition, the RMSEA, CFI, SRMR, TLI and PNFI were used to examine the model fit as these indices are most insensitive to model misspecification, sample size and parameter estimates.

SEM was performed in SPSS AMOS. Mediation analysis was conducted using the bootstrapping method with 1000 bootstrap resamples and 95% bias-corrected confidence estimates (MacKinnon *et al.*, 2004; Preacher and Hayes, 2004). Bootstrapping is a non-parametric method that involves sampling repeatedly from the dataset and estimating the indirect effect in each resampled dataset (Preacher and Hayes, 2008). This process is repeated thousands of times and is used to construct confidence intervals for the indirect effect (Preacher and Hayes, 2008). The p-value of this confidence interval is used to determine whether the indirect effect is significant. When the p-value is less than 0.05 then one is able to reject the null hypothesis that the true indirect effect is zero (Hayes, 2009). SPSS AMOS however only provides the significance and the size of the sum of all the indirect

effects through multiple mediators (Arbuckle, 2014). Therefore testing the significance of the individual indirect effects is not allowed in SPSS AMOS. As a result a user-defined estimand function, which enables one to estimate quantities that SPSS AMOS does not usually estimate, was developed by Gaskin (2016). This user-defined estimand function enables one to examine the significance of each individual indirect effect in the multiple mediation context and was used in the analysis.

In order to assess the relationship between various behavioural, demographic and socio-economic variables and systolic and diastolic blood pressure, a multiple linear regression analysis was then implemented. Two separate multiple linear regression models were analysed for both genders since the burden of hypertension and its risk factors often vary across genders (Wong *et al.*, 2015). SBP was considered as a dependent variable in one of the regression models analysed while DBP was considered as a dependent variable in a subsequent model analysed since it has been shown in literature that socio-economic variables affect DBP and SBP differently (Cois and Ehrlich, 2014). The following factors were considered as independent variables in each regression analysis: age, the use of antihypertensive medication, income, education, employment status, the ownership of household goods, BMI, resting heart rate, smoker status, alcohol consumption, physical activity and the emotional well-being of individuals. In the regression analysis, each of the categorical or ordinal variables were treated as such and were not regarded as continuous variables. The R^2 value, known as the coefficient of determination, which measures the proportion of the variation in the response variable that is explained by the predictors (Montgomery, Peck and Vining, 2015) and the mean square error(MSE) of the model were used to examine the fit of the regression model. The MSE measures the average of the square of the errors and can be written as the sum of the variance of the estimator and the squared bias of the estimator (Hair *et al.*, 1998). Only parameter estimates that have a p-value less than 0.05 will give an indication of the demographic, behavioural and socio-economic variables that have a significant association with blood pressure. The regression analysis was performed in SAS version 9.4. The results obtained from the multiple linear

regression analysis and SEM were then compared with respect to their ability to answer the research questions. The comparison of numerical estimates between SEM and multiple linear regression was not possible as ordinal variables were treated as continuous variables in the SEM analysis.

3.4 Summary

This chapter discussed the methodology applied in this study. Firstly, the source of the data and sampling design was discussed. Thereafter, the demographic variables, socio-economic variables and bio-behavioural risk factors that were used in the analysis were discussed in detail. The hypothesised structural equation model that would be analysed in the study was then reviewed. This proposed model suggested an effect of SES on DBP and SBP mediated by BMI, physical activity, heart rate, smoking, alcohol consumption and emotional well-being. Next, a discussion of SEM and its application to examine the relationship between SES and blood pressure was reviewed. Lastly mediation analysis and multiple linear regression were discussed. The next chapter provides a description of the data and a discussion of the analysis.

Chapter 4

Data analysis and results

This analysis consists of three parts, namely SEM, multiple linear regression and the evaluation and comparison of the results. Section 4.1 first describes the sample characteristics while section 4.2 details the data screening applied to the dataset. Subsequent sections provide insights into the two analytic approaches applied to the dataset and a comparison of the results. All of the SEM analysis was performed separately for males and females as one is interested in the separate effects of SES on SBP and DBP by gender. For multiple linear regression, the data was subdivided into a training set, validation set and test set for both males and females. The training set comprised of 40% of the data while the test and validation sets each comprised of 30% of the data (Harrell, 2015).

4.1 Sample characteristics

The descriptive statistics was conducted using SPSS version 23 (Arbuckle, 2014). The unweighted sample demographic and SES characteristics of the individuals that participated in the study are described in Table B.1 in Appendix B. Sample characteristics of the dataset are described as the median and interquartile range for continuous variables and frequency or percentage for categorical variables. The median and interquartile range measure the central tendency and spread, respectively, but are robust against outliers and non-normal data.

Table B.1 in Appendix B illustrates that 44% of participants in the sample were under 30 years old, 31% of participants were between 30 and 49, 16% of participants were between 50 and 64 and only 9% of participants were 65 or older. Female

participants accounted for 60% of the sample while males constituted 40% of the sample. Most of the participants in the study were African (82%) followed by Coloureds (14%). White participants (3%) had a low response rate and were under-represented relative to the South African population (Leibbrandt *et al.*, 2009). Only 34% of the participants were employed while 50% of the participants were not economically active. Unemployment proved to be a major contributor to the availability of the respondents to participate in the survey (Leibbrandt *et al.*, 2009). Unemployment was differentiated into discouraged and strict unemployment by distinguishing those who desired a job and were actively searching for work from those who were not actively searching (De Villiers *et al.*, 2013). The discouraged unemployed participants comprised of 2% of the sample while the participants who were strict unemployed comprised of 14% of the sample. The median individual income for participants in the sample was R800 with the individual income ranging from R0 to R956,000. The interquartile range for income was between R0 and R1740 which highlights that the income variable is skewed since the median is closer to the first quartile than the third quartile. Only 10% of individuals had no schooling while 22% completed primary schooling. The majority of the participants in the study completed secondary schooling (57%) and a further 11% completed tertiary education. Most of the participants owned a cellphone (77%), followed by a radio (39%) and a hifi (21%).

Table B.2 in Appendix B illustrates the bio-behavioural statistics for the participants in the study. Table B.2 illustrates that 72% of participants in the sample did not perform any physical exercise while only 12% of individuals exercised three or more times in a week. Furthermore 77% of participants were non-drinkers, 13% of individuals consumed between one and four alcoholic drinks in a day while 10% of participants consumed five or more alcoholic drinks in a day. The majority of participants in the study were non-smokers (81%). In addition, 10% of individuals smoked 5 or less cigarettes in a day, a further 6% smoked between 6 and 10 cigarettes per day and only 3% smoked more than 10 cigarettes in a day. In the majority of the cases, the individuals depicted a positive outlook to their emotional well-being. Most of the individuals indicated that they were rarely bothered (69%)

while only 2% of participants indicated that they were bothered all of the time. Furthermore 62% of individuals indicated that they had trouble focusing on a rare occasion while only 2% had trouble focusing all of the time. In addition, 56% of participants in the study were depressed on a rare occasion while 3% were depressed all of the time. Moreover, 50% of participants rarely felt that everything was an effort while 11% of individuals felt that everything was an effort between five and seven days of the week. In addition, 33% of participants were hopeful about the future between five and seven days of the week while only 26% were rarely hopeful about the future in the past week. Furthermore 64% of the participants in the study did not feel fearful in the past week while only 2% of individuals felt fearful all of the time. The majority of participants did not have a restless sleep in the past week (56%) while only 3% had a restless sleep between five and seven days of the week. Furthermore 68% of individuals were happy between three and seven days of that week while only 14% of individuals were rarely happy. Similarly 61% of participants felt lonely on a rare occasion while only 3% of individuals felt lonely all of the time. Furthermore 69% of individuals could not get going on a rare occasion while only 2% of participants could not get going all of the time. Each of these responses emphasise that the individuals depicted a positive outlook to their emotional well-being in the majority of cases.

Table B.3 in Appendix B depicts the blood pressure, the use of antihypertensive medication and the resting heart rate for the participants in the sample. Table B.3 illustrates that the BMI, SBP, DBP and heart rate readings were very similar. Since the distribution for BMI reading 1 and BMI reading 2 were almost identical and were only slightly skew it was decided to use the average of the readings and to treat BMI as an observed variable instead of a latent variable in the SEM (Cois and Ehrlich, 2014). In the sample, the median BMI was 25 kg\m^2, the median SBP was approximately 120 mmHg, the median DBP was approximately 80 mmHg and the median heart rate was approximately 74 bpm. Overall the use of antihypertensive medication was reported by 14% of participants. Current use of antihypertensive medication was noted by 8% of males and 18% of females. Using SBP greater than or equal to 140 mmHg and/or DBP greater than or equal to 90 mmHg (Shisana

et al., 2014) as cutoffs, 30% of males and 31% of females would be classified as hypertensive based on the first reading. This is much higher than the prevalence of high blood pressure in adults which was approximately 22% in 2014 (WHO, 2014). Approximately 11% of males were diagnosed with high blood pressure while 22% of females were diagnosed with high blood pressure. It is evident that a larger proportion of females currently use antihypertensive medication and have been diagnosed with high blood pressure however based on the cutoffs from the SBP and DBP a similar proportion of males and females would be classified as hypertensive.

4.2 Data screening

Before performing any analysis, the data were screened for missing data, outliers, normality, homoscedasticity, linearity and multicollinearity (Kline, 2011). A description of each of the variables used in the analysis is provided in Appendix A.

4.2.1 Missing data

For all questions in the survey, there was an option for individuals to either respond that they did not know the answer, refuse to answer, choose the not applicable option or choose not to respond to the question. For the purposes of the analysis, each of these responses were coded as missing as they were non-responses.

Missing data is a problem in any statistical analysis as most standard statistical methods presume complete information of all the variables included in the analysis. Few absent observations on certain variables can dramatically reduce the sample size. Consequently the precision of confidence intervals is affected, the statistical power is weakened and the parameter estimates may be biased. Dealing with missing data can often prove to be challenging as it requires careful examination of the data to determine the type and pattern of missingness and a clear understanding how each of the imputation methods work. Kline (2011) indicated that there is

little concern if there are less than 5% of missing cases on a single variable in a large sample where the reason for data loss is accidental or ignorable not systematic.

Before the data was made available for research purposes, the underlying variables related to income were imputed (De Villiers *et al.*, 2013). Single imputation which replaces missing values with a single calculated score was run in cases where there were 100 or more valid cases for a variable and the extent of missingness did not exceed 40%. After observing the dataset, it was evident that only 3 (7.32%) variables had complete data, while 17,596 (94.05%) cases were complete and 761,980 (99.33%) of values were complete. In this study, most of the variables used for the analysis had less than 2% of missing cases. The variable SBP_2 depicting the second SBP reading, had most of the cases missing, which accounted for 2.2% of the data. Since a maximum of 2.2% of data was missing on a single variable in a large sample, the missing values were of little concern (Kline, 2011) thus listwise deletion was applied to the dataset. For all further analysis, the full dataset included 17,596 complete cases. This was then subdivided into two datasets including 7,088 males and 10,508 females.

4.2.2 Univariate and Multivariate normality

Appendix C highlights the results after assessing univariate and multivariate normality. Univariate normality was assessed by examining the skewness and the kurtosis of the variables entered into the SEM analyses. Skewness is described as a measure of symmetry of the distribution while kurtosis is a measure of the flatness or peakedness of the distribution (Hair *et al.*, 1998). When the skewness and kurtosis values are close to zero, a distribution is said to be perfectly normal Tabachnick and Fidell (2007). Kline (2011) indicated that variables with the absolute value of the skewness index greater than 3 are extremely skewed while a variable with the absolute value of the kurtosis index greater than 10 suggests a problem. Table C.1 depicts that income, whether you own a sewing machine, bakkie, motorcycle, bicycle, computer or camera are severely skewed and kurtotic for males thus indicating non-normality for these variables. Similarly Table C.2

depicts that income, whether you own a sewing machine, private vehicle, bakkie, motorcycle, bicycle, computer or camera are severely skewed and kurtotic for females thus indicating non-normality for these variables. The distribution of the income variable is highly skewed for males and females with approximately 90% of individuals earning less than R4,000 per month while the remaining 10% of individuals had a monthly income between R4,000 and R956,000. For analysis purposes, the natural log-transformation of the income variable was used. This was based on the reasonable assumption that the impact on an individual's life of a given income increment decreases as income increases and does not stay constant as implied by the untransformed variable (Cois and Ehrlich, 2014). Furthermore, a log transformation reduces the dependency of the estimated regression coefficients on extreme values thus avoiding the influence of a few subjects with a large income. The other variables such as the ownership of a sewing machine, bakkie, motorcycle, bicycle, computer and camera that indicated non-normality were excluded from all further analysis for males. Similarly these variables were excluded from all further analysis for females. In addition, the variable indicating the ownership of a private vehicle was also excluded for females as it was extremely kurtotic and skewed. Very few individuals in the dataset indicated ownership of these goods thus resulting in the non-normality of these variables. Smoker status was also skewed and kurtotic for females however it was still included in the model for females in order to assess its effect as a mediator between SES and SBP and SES and DBP (Table C.4 in Appendix C).

Byrne (2010) indicated that since SEM is based on the analysis of covariance structures, multivariate kurtosis is a concern in any SEM analysis. As a result, Mardia's test which was developed in 1970, was used to assess multivariate normality (Mardia, 1970). These results are displayed in Table C.3 in Appendix C for males and in Table C.4 in Appendix C for females. The last two columns of these tables depict the univariate kurtosis value and its critical ratio for each of the remaining variables considered in the analysis. The index of multivariate kurtosis and its critical ratio are both placed at the bottom of the kurtosis and critical ratio columns respectively. This multivariate critical ratio value depicts Mardia's normalized

estimate of multivariate kurtosis (Byrne, 2010). In practice, Mardia's normalized estimate values exceeding 5 indicate that the data is non-normally distributed (Bentler, 2005). In our study, the multivariate critical ratio value of 139.67 for males and 195.17 for females is highly suggestive of multivariate non-normality. Interpretations based on maximum likelihood estimation become problematic since the data reveals evidence of multivariate kurtosis, thus an alternative approach such as ADF estimation was used (Browne, 1984).

4.2.3 Outliers

Outliers are observations with a unique combination of characteristics such that they are completely different from the other observations (Hair *et al.*, 1998). Box-plots were assessed to determine if there are mild or extreme univariate outliers for the variables included in the regression analysis. The IQR depicts the interquartile range, which is the middle 50% of the scores. Mild outliers are any scores more than 1.5*IQR from the rest of the scores, and are indicated by open dots while extreme outliers are any score more than 3*IQR from the rest of the scores. In other words, an outlier is determined by comparison to the bulk of the scores in the middle. From subsection 3.2.3 and 3.2.4, it was evident that only plausible SBP, DBP, heart rate, height and weight measurements were retained in the analysis. The cutoffs that were retained for each measurement were based on literature guidelines (Cois and Ehrlich, 2014). Age and income were then examined for univariate outliers as all other variables were either categorical variables, dichotomous variables or Likert scale questions, thus one would not expect to observe outliers for these variables as there were specific options that needed to be selected by participants of the study. From Figure D.1(a) in Appendix D, it was clear that there were no extreme univariate outliers for age for males however there were 12 mild outliers. Similarly from Figure D.1(b) in Appendix D, it was clear that there were no extreme univariate outliers for age for females however there were five mild outliers. From Figure D.2(a) in Appendix D, it was evident that there were 21 extreme univariate outliers for income for males and from Figure D.2(b) in Appendix D, it

was evident that there were 13 extreme univariate outliers for income for females. Each of the other survey responses for each of these individuals were examined to determine whether the income seemed plausible and in all cases, this appeared to be the case.

Since SEM is a multivariate analysis technique, multivariate outliers were then examined. The squared Mahalanobis distance was computed for each case in order to detect the multivariate outliers. These results are displayed in Table D.1 in Appendix D for males and in Table D.2 in Appendix D for females. There were seven cases detected as multivariate outliers for males and females as the squared Mahalanobis distance values stood distinctively apart from all other squared Mahalanobis distance values. As a result each of these multivariate outliers were removed from the males and females datasets so that the analysis was not severely affected by these outliers (Byrne, 2010).

4.2.4 Linearity and Homoscedasticity

Linearity and homoscedasticity are two important assumptions required for any multiple linear regression analysis. Linearity refers to a linear relationship between the variables while homoscedasticity pertains to the assumption that the variance around the regression line is the same for all values of the independent variables (Hair *et al.*, 1998). Linearity was assessed by determining whether each of the continuous independent variables and the dependent variable were related. Each of the scatterplots for the training, validation and testing dataset for males and females revealed a linear relationship between BMI, age, the natural logarithm of income and heart rate with SBP and with DBP. Homoscedasticity was assessed by observing a scatterplot of the residual and the predicted value for each dependent variable. Based on the inspection of each of the scatterplots for the training, validation and testing dataset for males and females, one observes that the variance is homogenously distributed for SBP and DBP. This appears to be the case since the residuals are distributed around zero which implies that the assumption of

homoscedasticity is not violated. Furthermore the residuals do not appear to exhibit any particular pattern which implies independence of the residuals.

4.2.5 Multicollinearity

Multicollinearity depicts the extent to which a variables effect can be taken into account by other independent variables in the analysis (Hair *et al.*, 1998). One of the problems that may arise when highly correlated predictor variables are present in the dataset, is that the regression equation does not make sense with some of the coefficients of the regression equation having seemingly incorrect signs. In addition, the sizes and significances of the variables may change greatly as variables enter and leave the regression. According to Kline (2011), the variance inflation factor (VIF) and the tolerance values can be used as an indication of multicollinearity. The tolerance indicates the proportion of total standardised variance that is unique while the VIF depicts the ratio of the total standardised variance over the tolerance (Kline, 2011). A tolerance value less than 0.10 and VIF greater than 10 indicates multicollinearity (Kline, 2011). From each of the parameter estimates tables for the models in section 4.4.1 below, one observes that the tolerance values and VIF do not exceed these thresholds thus multicollinearity is not a problem in the dataset.

4.3 Structural equation modelling

4.3.1 Cronbach's alpha and Exploratory factor analysis

SPSS version 23 was first used to determine the Cronbach alpha value in order to assess the internal consistency of the items measured on a Likert scale to determine whether the scale is reliable (Kline, 2011). These results are found below for males and in Appendix E for females. Initially it was proposed that 10 items would load onto the emotional well-being latent variable, however after examining the Cronbach alpha value, it was decided to remove emohope, emohappy and emoeff

(variables described in Table A.1 in Appendix A) from the items as these variables would result in a lower Cronbach alpha for both males and females. The following table depicts the final Cronbach alpha value for items loading onto emotional well-being for males:

Table 4.1: Reliability Statistics for 7 variables loading onto emotional well-being for males

Cronbach's Alpha	**Cronbach's Alpha Based on Standardised Items**	**Items**
0.80	0.80	7

From Table 4.1, it is evident that the final Cronbach alpha value was 0.80 which indicates a "very good" reliability coefficient (Kline, 2011). Similarly the final Cronbach alpha value for emotional well-being for females was 0.80 and this is depicted in Table E.1 in Appendix E.

Table 4.2: Inter-Item Correlation Matrix for 7 variables loading onto emotional well-being for males

	Emobth	**Emomnd**	**Emodep**	**Emofear**	**Emoslp**	**Emolone**	**Emogo**
Emobth	1.00	0.51	0.41	0.33	0.33	0.30	0.31
Emomnd	0.51	1.00	0.42	0.40	0.39	0.35	0.38
Emodep	0.41	0.42	1.00	0.38	0.38	0.34	0.34
Emofear	0.33	0.40	0.38	1.00	0.32	0.35	0.36
Emoslp	0.33	0.39	0.38	0.32	1.00	0.34	0.34
Emolone	0.30	0.35	0.34	0.35	0.34	1.00	0.35
Emogo	0.31	0.38	0.34	0.36	0.34	0.35	1.00

From Table 4.2 and Table E.2 (in Appendix E), it is clear that all of the correlations for the variables loading onto emotional well-being for males and females are positive and are in a similar range with no very high or very low correlations (Robinson, Shaver and Wrightsman, 1991).

Table 4.3: Item-Total Statistics for 7 variables loading onto emotional well-being for males

	Scale Mean if Item Deleted	**Scale Variance if Item Deleted**	**Corrected Item-Total Correlation**	**Squared Multiple Correlation**	**Cronbach's Alpha if Item Deleted**
Emobth	8.87	8.80	0.53	0.32	0.77
Emomnd	8.77	8.33	0.60	0.39	0.76
Emodep	8.68	8.25	0.55	0.31	0.77
Emofear	8.83	8.83	0.52	0.27	0.77
Emoslp	8.70	8.48	0.51	0.26	0.78
Emolone	8.76	8.65	0.49	0.24	0.78
Emogo	8.86	8.85	0.50	0.26	0.78

Table 4.3 for males and Table E.3 for females (in Appendix E) highlight that there are no other items that load onto emotional well-being that would cause a higher Cronbach alpha if they were deleted. It was also observed that the minimum of the corrected item-total correlation, which is the correlation between the item and the sum of the rest of the items, exceeded 0.4 which is often used as a rule of thumb (Blunch, 2008) when assessing the reliability coefficients for items.

After examining the Cronbach alpha for items loading onto emotional well-being, exploratory factor analysis was performed with varimax rotation on all variables that load onto latent variables. Varimax rotation is a commonly used orthogonal method of rotation that produces factors that are uncorrelated (Costello and Osborne, 2005). This was done to ensure that the variables loading onto the latent variables only load onto one factor. These results can be found in Appendix F. The EFA was performed in SPSS version 23. In the first iteration for males, factor analysis was performed on each of the following variables: SBP_1, SBP_2, DBP_1, DBP_2, HR_1, HR_2, emobth, emomnd, emodep, emofear, emoslp, emolone, emogo, empl_stat, educ, ln_inc, ownrad, ownhif, ownvehpri and owncel (described in Table A.1 in Appendix A). The results of the rotated component matrix in Table F.1 depict that ownvehpri loads onto multiple factors for males. As a result, this variable was

then excluded from the factor analysis and the results were examined (Costello and Osborne, 2005). The EFA for females excluded the ownvehpri variable as it was severely skewed and kurtotic and all of the remaining variables were taken into account in the factor analysis.

The following table depicts the Kaiser-Meyer-Olkin (KMO) and Bartlett's measure for males in the final iteration:

Table 4.4: KMO and Bartlett's Test for males iteration 2

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.73
Bartlett's Test of Sphericity	Approx. Chi-Square	62004.56
	df	171
	Sig.	0.000

The Kaiser-Meyer-Olkin (KMO) and Bartlett's measure were examined to determine whether it is appropriate to apply factor analysis. These results are displayed in Table 4.4 for males and Table F.2 for females (in Appendix F). Generally the KMO statistic varies between zero and one and a value greater than 0.5 is considered acceptable (Kaiser, 1974). In our case since the KMO is greater than 0.5 for both males and females, it was acceptable to apply factor analysis. Bartlett's measure tests the null hypothesis that the original correlation matrix is an identity matrix. A significance level less than 0.05 indicates that the R-matrix is not the identity matrix and there are some relationships between the variables (Bartlett, 1950). For our data, Bartlett's test was highly significant for both males and females therefore factor analysis was appropriate. The following table depicts the total variance explained when applying factor analysis for males:

Table 4.5: Variance explained when eigenvalues exceed 1 for males iteration 2

	Initial Eigenvalues			**Rotation Sums Squared Loadings**		
Component	**Total**	**% Variance**	**Cum %**	**Total**	**% Variance**	**Cum %**
1	3.73	19.61	19.61	3.35	17.64	17.64
2	3.17	16.68	36.29	3.19	16.78	34.42
3	2.14	11.26	47.55	2.16	11.38	45.79

Component	Initial Eigenvalues			Rotation Sums Squared Loadings		
	Total	% Variance	Cum %	Total	% Variance	Cum %
4	1.81	9.53	57.08	1.92	10.11	55.90
5	1.03	5.40	62.47	1.25	6.57	62.47
Extraction Method: Principal Component Analysis and Cum = Cumulative						

Table 4.5 highlights that 62.47% of the variance can be explained by the variables for males while 62.33% of the variance can be explained by the variables for females (Table F.3 in Appendix F).

The following table depicts the rotated component matrix for males:

Table 4.6: Rotated Component matrix for males iteration 2

	Component				
	1	2	3	4	5
SBP_1	0.92				
SBP_2	0.92				
HR_1				0.97	
HR_2				0.97	
DBP_1	0.89				
DBP_2	0.89				
Emobth		0.68			
Emomnd		0.74			
Emodep		0.70			
Emofear		0.66			
Emoslp		0.65			
Emolone		0.63			
Emogo		0.64			
Empl_stat			0.71		
Educ					0.84
Ln_inc			0.77		
Ownrad_r			0.71		
Ownhif_r			0.66		

	Component				
	1	2	3	4	5
Owncel_r					0.68
Rotation Method: Varimax with Kaiser Normalization Rotation converged in 5 iterations					

From Table 4.6 it is clear that each of the items load onto one factor for males. Similarly from Table F.4 in Appendix F it is clear that each of the items load onto one factor for females. SBP_1, SBP_2, DBP_1 and DBP_2 (described in Table A.1 in Appendix A) load onto one factor however for the purposes of our analysis, SBP_1 and SBP_2 will load onto the SBP latent variable and DBP_1 and DBP_2 will load onto the DBP latent variable. This will be done since we are interested in the separate effects of behavioural and socio-economic factors on SBP and DBP. Table 4.6 also highlights that education and the ownership of a cellphone load onto one factor while employment status, income and the ownership of radio and hi-fi load onto another factor however for the purposes of the CFA we will examine all of these factors loading onto the SES latent variable for males as all of these variables are perceived to be a measure of SES (Kaplan and Keil, 1993). Similarly Table F.4 depicts that the ownership of radio and the ownership of hi-fi load onto one factor while employment status, education, income and the ownership of a cellphone load onto another factor however for the purposes of the CFA we will examine all of these factors loading onto the SES latent variable for females (Kaplan and Keil, 1993).

4.3.2 Discussion of SEM results

SEM was used to test the hypothesised conceptual framework using SPSS AMOS version 23. SEM is a combination of factor (measurement model) and path (structural model) analyses. In this study, the relationships among several factors influencing SBP and DBP were analysed through two steps. CFA was first performed to assess the validity of the latent factors while the impact of the predictors influencing SBP and DBP was modelled in the second step.

CFA was first performed in order to determine the validity of the emotional well-being and SES latent variables. SBP, DBP and HR each had two observed variables loading on them which depicted the first and second respective reading thus CFA was not necessary for these latent variables. For each CFA model, measures of overall fit were first examined by considering various indices and the model was modified if it was necessary. Thereafter parameter estimates were assessed to ensure that they were of reasonable size and in the right direction (Byrne, 2010).

In the second step of the analysis, two different models were built for males and females to assess the research questions proposed in this study. The first model estimated the direct relationship between SES and SBP and SES and DBP while the second model investigated whether certain bio-behavioural risk factors mediate the relationship between SES and blood pressure. Since the data reveals evidence of multivariate non-normality, ADF estimation was used to build all the models, to perform model modification and to evaluate the model (Browne, 1984). The model fit was evaluated according to the threshold values discussed in subsection 2.1.1.

4.3.2.1 Confirmatory factor analysis for emotional well-being

The first baseline CFA model that was built investigated whether each of the observed variables loading onto the emotional well-being latent construct were satisfactory. The tested baseline measurement model for emotional well-being is presented in Figure G.1 for males and females in Appendix G. Various model fit indices were first examined to determine the overall model fit for the baseline CFA model for emotional well-being. The model chi-square, normed chi-square, RMSEA, SRMR, CFI, PNFI and TLI fit indices were inspected in this study. These fit indices for the baseline model are found in Table G.2 for males and Table G.7 for females in Appendix G. The chi-square value was significant ($\chi^2(14) = 140.271$, $p < 0.001$) depicting that the baseline model for males predicted relations that were significantly different from the relations observed in the sample. The normed chi-square ($\chi^2 \backslash \text{DF}$) was used to correct for the sample size sensitivity of the model chi-square. The normed chi-square value, which was 10.019, depicted a poor fit as values close to 5 have been recommended to demonstrate reasonable fit (Wheaton *et al.*, 1977). In contrast, the CFI = 0.925, SRMR = 0.0257, RMSEA = 0.036, PNFI = 0.612 and TLI = 0.888 which indicated a close approximate fit (Mulaik

et al., 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000) of the baseline emotional well-being CFA model for males. Similarly the chi-square value for females was significant ($\chi^2(14) = 222.8$, $p < 0.001$) which depicted that the baseline model predicted relations that were significantly different from the relations observed in the sample. Furthermore, the normed chi-square which was 15.914 for the baseline emotional well-being model for females, indicated a poor fit (Wheaton *et al.*, 1977). Since the chi-square value and the normed chi-square value for the baseline models for both males and females suggested that the model fit is not entirely adequate, the modification indices were then investigated (Byrne, 2010). Modification indices are very useful for inspecting local misfit. They give an indication of how much the chi-square value of a model would drop if the parameter were free instead of constrained thus giving an indication of how much the model would improve (Byrne, 2010). The modification indices for the baseline models for males and females (Table G.1 and Table G.6 in Appendix G) revealed that the incorporation of the error covariance between having trouble focusing and being emotionally bothered made a substantially large improvement to the model fit. Jamison, Sbrocco and Parris (1989) validated this relationship when they examined the influence of problems with concentration and memory on emotional distress and daily activities in chronic pain patients. As a result, the baseline models for both males and females were then modified to incorporate this covariance.

The tested modified measurement model for emotional well-being for both males and females is presented below:

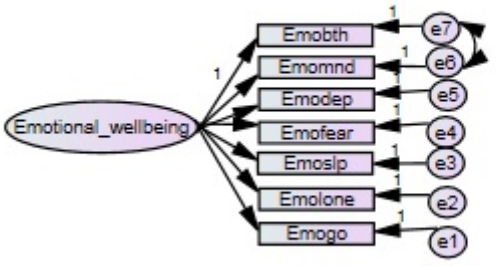


Figure 4.1: Modified model: Emotional well-being CFA model for males and females

The model fit indices for the modified models for males and females are found in Table G.4

and Table G.9 in Appendix G. The chi-square value for males ($\chi^2(14) = 56.65$, $p < 0.001$) was still significant however there was a vast improvement with a reduction in the chi-square value. The normed chi-square value for males was 4.358 which now indicated a reasonable model fit (Wheaton *et al.*, 1977). Furthermore the CFI = 0.974, SRMR = 0.015, RMSEA = 0.022, PNFI = 0.599 and TLI = 0.958 all indicated a close approximate fit (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000) of the modified emotional well-being CFA model for males. Similarly the normed chi-square value for females was 5.824 which depicted a reasonable model fit (Wheaton *et al.*, 1977). In addition, the CFI = 0.976, SRMR = 0.0145, RMSEA = 0.021, PNFI = 0.602 and TLI = 0.962 all indicated a reasonable model fit of the modified emotional well-being CFA model for females (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). After inspection of the modification indices for the modified measurement models for emotional well-being for both males and females (Table G.3 and Table G.8 in Appendix G), there was no evidence of model misspecification. As a result, these modified models (Figure 4.1) represented the final best fitting and most parsimonious models for emotional well-being.

In addition to the model fit indices, path coefficients were assessed to interpret the results of the CFA. Unstandardised estimates of the path coefficients are interpreted as unstandardised regression coefficients that estimate the direct effects of the factors on indicators. The unstandardised path coefficients of the modified CFA model for emotional well-being for males and females are presented in Table G.5 and Table G.10 in Appendix G. The unstandardised coefficients for both males and females showed that all the indicators' loadings on the emotional well-being latent variable were statistically significant. Standardised path coefficients are equivalent to beta weights in regression. The standardised path coefficient values less than 0.10 depict small effects; values around 0.30 depict medium effects; and values greater than 0.50 depict large effects (Kline, 2011). In the current study, standardised factor loadings for emotional well-being for males were all statistically significant and ranged from 0.565 (large) to 0.658 (large). Similarly, standardised factor loadings for emotional well-being for females were all statistically significant and ranged from 0.56 (large) to 0.677 (large). The path coefficients for both males and females demonstrated that the indicators measuring emotional well-being were adequate.

4.3.2.2 Confirmatory factor analysis for SES

The next baseline CFA model that was built investigated whether each of the observed variables loading onto the SES latent construct were satisfactory. The tested baseline measurement model for SES is presented in Figure G.2 for males and females in Appendix G. The model fit indices were first inspected to assess the overall model fit for the baseline CFA model for SES. These fit indices for the baseline model can be examined in Table G.12 for males and Table G.17 for females in Appendix G. The chi-square value was significant ($\chi^2(9) = 1258.992$, $p < 0.001$) indicating that the baseline model for males predicted relations that were significantly different from the relations observed in the sample. The normed chi-square value, which was 139.888, indicated a poor fit (Wheaton *et al.*, 1977). Similarly the CFI = 0.821, SRMR = 0.0905, RMSEA = 0.14, PNFI = 0.492 and TLI = 0.702 indicated a poor fit of the baseline SES CFA model for males (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). Similarly the chi-square value for females was significant ($\chi^2(9) = 2634.737$, $p < 0.001$) and the normed chi-square which was 292.749 indicated a poor fit of the model (Wheaton *et al.*, 1977). In addition the CFI = 0.632, SRMR = 0.1145, RMSEA = 0.167, PNFI = 0.379 and TLI = 0.386 which indicated a poor fit of the baseline SES CFA model for females (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). Since each of the fit indices for the baseline models for both males and females indicated that the fit of the model was not adequate, the modification indices were then investigated (Byrne, 2010). The modification indices for the baseline models for males and females (Table G.11 and Table G.16 in Appendix G) revealed that the incorporation of certain error covariances made a substantially large improvement to the model fit. Each of the error covariances that made a large improvement to the model fit were incorporated into the model one by one and the modification indices were investigated at each point to determine if the model could be improved further.

The following figures illustrate the tested modified measurement model for SES for males and females:

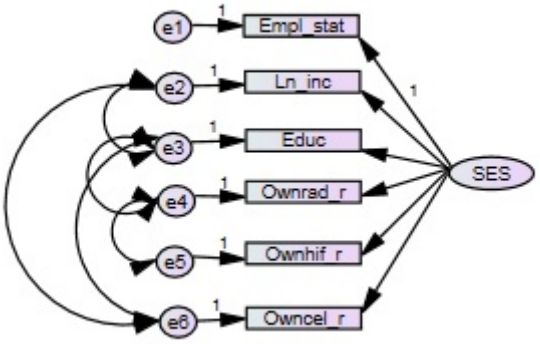


Figure 4.2: Modified model: SES CFA model for males

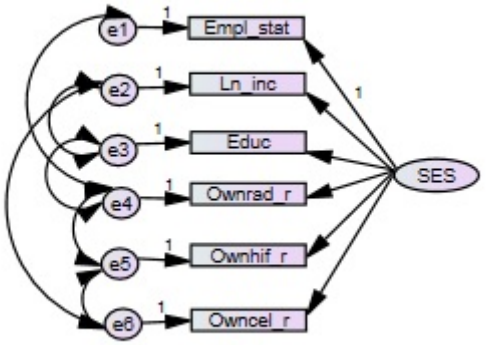


Figure 4.3: Modified model: SES CFA model for females

After the investigation of the modification indices for the baseline model for males, error covariances between income and education, between education and the ownership of a radio, between the ownership of a radio and the ownership of a hifi, between education and the ownership of a cellphone and between income and the ownership of a cellphone were incorporated into the model. After the evaluation of the modification indices for the baseline model for females, error covariances between employment status and the ownership of a radio, between income and education, between education and the ownership of a radio, between the ownership of a radio and the ownership of a hifi, between income and the ownership of a cellphone and between the ownership of a hifi and the ownership of a cellphone were incorporated into the model. Some of these relationships were found in previous literature (De Gregorio and Lee, 2002; Pikhart, Bobak, Rose and Marmot, 2003) which validated the inclusion of these error covariances.

The model fit indices for the modified models for males and females are found in Table G.14 and Table G.19 in Appendix G. The chi-square value for males ($\chi^2(4) = 76.286$, $p < 0.001$) was still significant however there was a vast improvement with a reduction in the chi-square value. The normed chi-square value for males was 19.071 which still indicated a relatively poor model fit (Wheaton *et al.*, 1977). However the CFI = 0.99, SRMR = 0.0209, RMSEA = 0.051 and TLI = 0.961 all indicated a substantially good model fit of the modified SES CFA model for males (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). Similarly the CFI = 0.993, SRMR = 0.0131, RMSEA = 0.04 and TLI = 0.964 all indicated a substantially good model fit of the modified SES CFA model for females (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). After inspection of the modification indices for the modified measurement models for SES for both males and females (Table G.13 and Table G.18 in Appendix G), there was no evidence of model misspecification (Byrne, 2010). As a result, these modified models (Figure 4.2 and Figure 4.3) represented the final best fitting and most parsimonious models for SES.

In addition to the model fit indices, the unstandardised path coefficients of the modified CFA model for emotional SES for males and females are presented in Table G.15 and Table G.20 in Appendix G. The unstandardised coefficients for both males and females showed that all the indicators' loadings on the SES latent variable were statistically significant. The standardised factor loadings for SES for males were all statistically significant and ranged from 0.337 (medium) to 0.81 (large) (Kline, 2011). Similarly, standardised factor loadings for SES for females were all statistically significant and ranged from 0.221 (medium) to 0.75 (large) (Kline, 2011). The path coefficients for both males and females demonstrated that the indicators measuring SES were adequate.

4.3.2.3 SEM Model 1: Investigating the direct relationship between SES and blood pressure

One of the objectives of this study was to determine whether there is a direct relationship between SES and SBP and SES and DBP. The tested baseline model used to investigate this relationship is presented in Figure H.1 for males and Figure H.2 for females in Appendix H. Non-causal correlations were allowed between each pair of blood pressure measurements and between the latent variables representing SBP and DBP. Allowing

these correlations to be different from zero means accepting the plausible hypotheses that factors related to the specific conditions of the measurement (for example cuff positioning or the procedure used by the fieldworker) affect the measured values of systolic and diastolic blood pressure in the same reading, creating a correlation which is not completely explained by the "true" values of the blood pressure (Cois and Ehrlich, 2014).

The model fit indices were first inspected to examine the overall fit of the baseline models investigating the direct relationship between SES and blood pressure. The fit indices for the baseline models can be examined in Table H.1 for males and Table H.9 for females in Appendix H. The chi-square value was significant ($\chi^2(25) = 441.152$, $p < 0.001$) indicating that the baseline model for males predicted relations that were significantly different from the relations observed in the sample. However, as stated in subsection 2.1.1, many problems have been reported related to the chi-square value as a fit statistic. Therefore, various other model fit indices were investigated to assess their consistency with each other. The normed chi-square value, which was 17.646, indicated a poor fit (Wheaton *et al.*, 1977). However the CFI = 0.963, SRMR = 0.0556, RMSEA = 0.048, PNFI = 0.534 and TLI = 0.933 all indicated a very good fit of SEM model 1 for males (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). Similarly the chi-square value for females was significant ($\chi^2(24) = 1565.06$, $p < 0.001$) and the normed chi-square which was 65.211 indicated a poor fit (Wheaton *et al.*, 1977). However the CFI = 0.945, SRMR = 0.0846, RMSEA = 0.058, PNFI = 0.466 and TLI = 0.916 all indicated a reasonable fit of the SEM model 1 for females (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). Since most of the fit indices for the baseline models for both males and females indicated that the fit of the model was adequate, the parameter estimates were then investigated.

The unstandardised path coefficients of the SEM model 1 for males and females are presented in Table H.2 and Table H.10 in Appendix H. The unstandardised coefficients for both males and females showed that all the indicators' loadings on their respective SES latent variable were statistically significant. In addition, the covariances are presented in Table H.3 and Table H.11 in Appendix H. After evaluating the covariances for both males and females, the error covariance between SBP reading 2 and DBP reading 2 was found to be not significant. This error covariance was then removed for males and females. The following figures present the modified models for males and females:

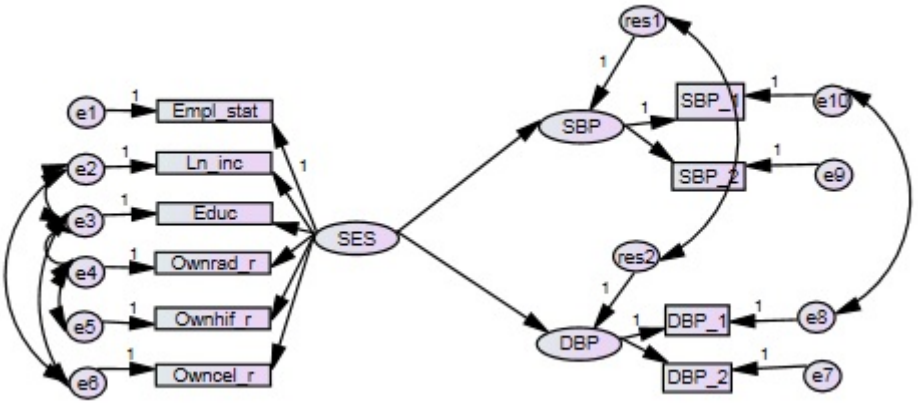


Figure 4.4: Modified model: Model investigating direct relationship of SES and blood pressure for males

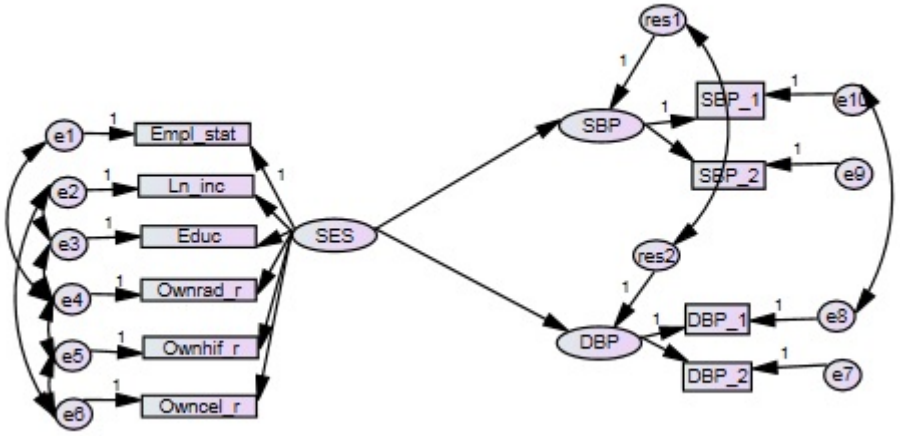


Figure 4.5: Modified model: Model investigating direct relationship of SES and blood pressure for females

The model fit indices are displayed in Table H.4 and Table H.12 in Appendix H. Each of the fit indices for the modified models depicted almost identical results to the baseline models for both males and females thus showing that the model fit was adequate. The unstandardised path coefficients of the modified SEM model 1 for males and females are presented in Table H.5 and Table H.13 in Appendix H. The unstandardised coefficients for both males and females showed that all the indicators' loadings on their respective latent variable were statistically significant and that the relationship between SES and SBP and SES and DBP was statistically significant. Furthermore, the results for males depicted that the standard errors of the parameters (S.E.) ranged from 0.006 to 0.241

and were thus reasonably determined. After examining the regression weights for males, it is observed that the SBP increases by 4.452 for every one unit increase in SES while the DBP increases by 3.215 for every one unit increase in SES. The squared multiple correlation for males which is presented in Table H.8 in Appendix H depicts that SES explains only 6.3% of the variance in SBP and 7.1% of the variance in DBP. Similarly the results for females depicted that the standard errors of the parameters (S.E.) ranged from 0.006 to 0.267 and were thus reasonably determined. The regression weights for females highlight that the SBP increases by 2.44 for every one unit increase in SES while the DBP increases by 2.163 for every one unit increase in SES. The squared multiple correlation for females which is presented in Table H.16 in Appendix H depicts that SES explains only 1.4% of the variance in SBP and 2.3% of the variance in DBP. In addition, the covariances which are presented in Table H.6 and Table H.14 in Appendix H were all statistically significant for males and females. Similarly the variances which are displayed in Table H.7 and Table H.15 in Appendix H were all statistically significant for males and females. These results revealed that there is a significant positive relationship between SES and SBP and SES and DBP for both males and females.

4.3.2.4 SEM Model 2: Investigating whether bio-behavioural risk factors mediate the relationship between SES and blood pressure

Since there is a direct relationship between SES and blood pressure, the next aim was to investigate whether certain bio-behavioural risk factors mediate this relationship. SPSS AMOS version 23 was used to test all of the components of the mediation model at the same time (Arbuckle, 2014). SPSS AMOS however only provides the significance and the size of the sum of all indirect effects through multiple mediators and does not allow for testing the significance of individual effects or comparing the strengths of individual indirect effects (Arbuckle, 2014). As a result, a user-defined estimand function developed by Gaskin (2016) was used, as it enables one to examine the significance of each individual indirect effect using bootstrapping resampling in SPSS AMOS. Thus, the significance of the sum of all indirect effects as well as the significance of the individual effects were assessed (Preacher and Hayes, 2008).

The following table depicts each of the different iterations that were run to investigate the indirect relationship between SES and blood pressure through the use of mediators:

Table 4.7: Description of mediation models for males and females

Model	Description	Estimation
2.1	Without controls: use of antihypertensive medication and age	ADF estimation
2.2	With controls: use of antihypertensive medication and age	ADF estimation

Two separate models were built for males and females to determine the effect of age and the use of antihypertensive medication as controls. These variables were considered as controls as they are potentially confounding variables which may have an effect on the dependent variables (Miyaki *et al.*, 2013; Ploubidis *et al.*, 2013; Bardenheier *et al.*, 2013; Cois and Ehrlich, 2014; Keetile *et al.*, 2015). Controlling for these variables enables one to balance their effects across subjects and groups so that one can ignore them and just study the association between the independent and dependent variables of interest (Becker, 2005). Multiple criteria were used to interpret the SEM model with mediation. In order to interpret the overall fit of the hypothesised model to the data, various model fit indices such as the chi-square, normed chi-square, CFI, SRMR, RMSEA, PNFI and TLI were examined. In addition, parameter estimates were assessed to interpret the effects on the endogenous variables from SES, which has been shown to directly predict them in section 4.3.2.3. Thereafter, the indirect and total effects were assessed to interpret the effects of the mediator variables in the relationship between SES and blood pressure. Lastly, the squared multiple correlation coefficients were examined to determine the amount of variance in SBP and DBP that was explained by the model.

The following tested models were used to investigate the mediation relationship for males and females excluding the controls:

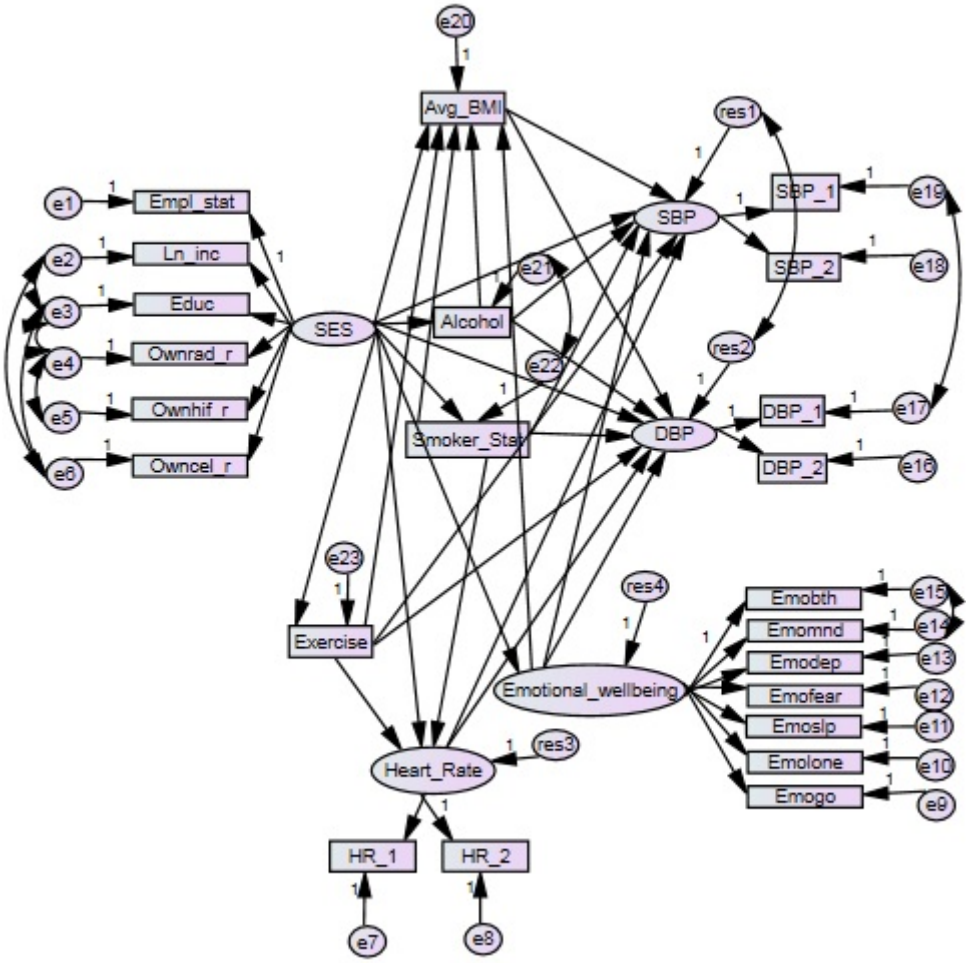


Figure 4.6: Model 2.1 investigating mediated relationship of SES and blood pressure for males excluding controls

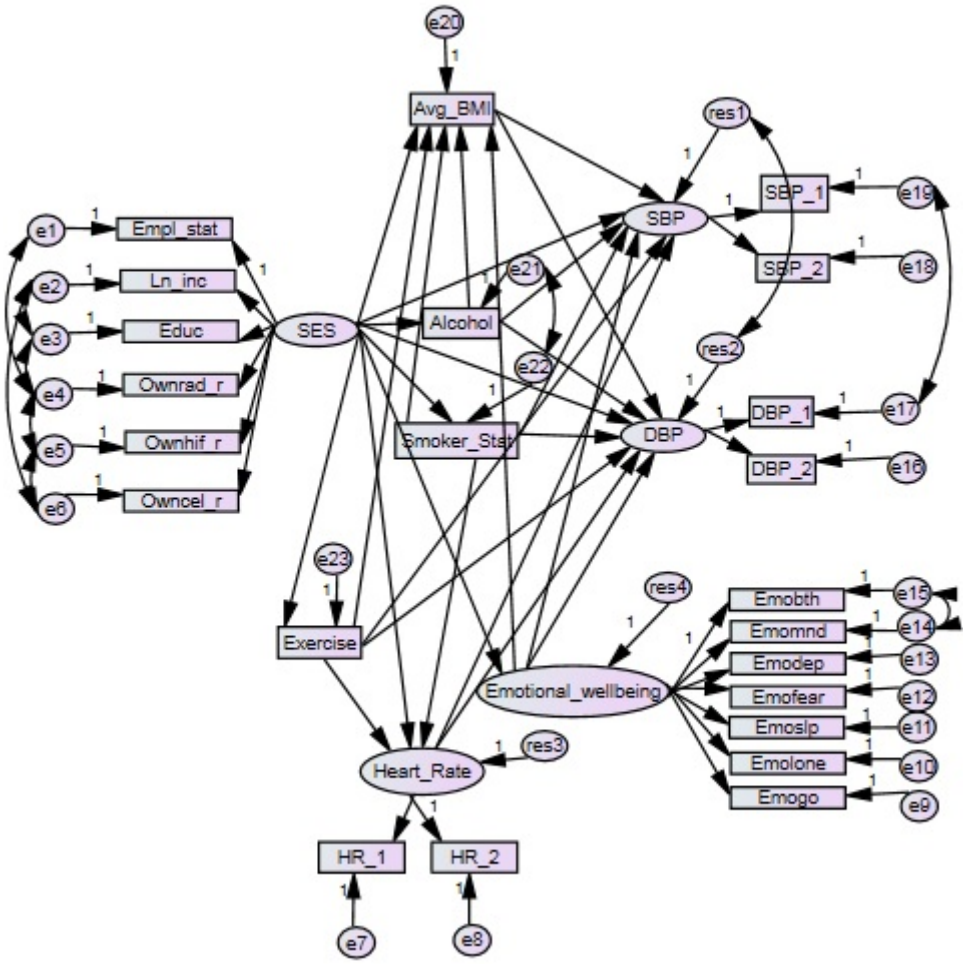


Figure 4.7: Model 2.1 investigating mediated relationship of SES and blood pressure for females excluding controls

The CFA portion of the model for latent variables SES and emotional well-being were built as specified in the final modified CFA models for males and females. The error covariances between SBP reading 1 and DBP reading 1 and the error covariance between SBP and DBP were included in the model as specified in section 4.3.2.3. In addition an error covariance between alcohol consumption and smoker status was included in the model for males and females as it was believed that there is a relationship between these two mediators (Batel, Pessione, Maître and Rueff, 1995).

The following tables depict the model fit indices for the mediated models excluding age and the use of antihypertensive medication for males and females:

Table 4.8: Model fit indices for SEM model 2.1 for males

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
1847.782	9.285	0.938	9.847	0.0444	0.034	0.706	0.913

Table 4.9: Model fit indices for SEM model 2.1 for females

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
3001.724	15.16	0.924	21.764	0.0585	0.037	0.67	0.906

The chi-square value for males was significant ($\chi^2(199) = 1847.782$, $p < 0.001$) depicting that the mediated model for males predicted relations that were significantly different from the relations observed in the sample. The normed chi-square value, which was 9.285, indicated a poor fit however it was relatively close to the threshold of 5 (Wheaton *et al.*, 1977). The CFI = 0.938, SRMR = 0.0444, RMSEA = 0.034, PNFI = 0.706 and TLI = 0.913 indicated a good fit of SEM model 2.1 for males (Hu and Bentler, 1999; Mulaik *et al.*, 1989; Diamantopoulos and Siguaw, 2000). Similarly the chi-square value for females was significant ($\chi^2(198) = 3001.724$, $p < 0.001$) and the normed chi-square which was 15.16 indicated a poor fit (Wheaton *et al.*, 1977). However the CFI = 0.924, SRMR = 0.0585, RMSEA = 0.037, PNFI = 0.67 and TLI = 0.906 indicated a relatively good fit of the SEM model 2.1 for females (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). The modification indices for both males and females did not depict any further parameters which were substantially meaningful or relevant for inclusion in the models. Since most of the fit indices for the mediated models for both males and females indicated that the fit of the model was adequate, the parameter estimates were then investigated.

The unstandardised path coefficients of the mediated SEM model 2.1 for males and females are presented in Table I.1 and Table I.8 in Appendix I. The unstandardised coefficients for both males and females showed that all the indicators' loadings on their respective latent variable were statistically significant at the 5% level of significance. The standardised parameter estimates for both the measurement portion and the structural portion of the model for males and females are presented in Table I.2 and Table I.9 in Appendix I. These results indicated that all factor loadings of the measurement portion ranged from 0.258 (medium effect) to 0.973 (large effect) for males and ranged from 0.312

(medium effect) to 0.982 (large effect) for females (Kline, 2011). In addition, each of the indicators had a positive loading on their respective latent variable. The standardised parameter estimates for the structural portion of the model for males revealed 22 out of 26 statistically significant coefficients. This indicated that 22 of the 26 presumed direct effects on endogenous variables from other variables were statistically significant at the 5% level of significance. The statistically significant coefficients for the structural portion of the model for males ranged from 0.03 (small effect) to 0.372 (medium effect) (Kline, 2011). Similarly, the standardised parameter estimates for the structural portion of the model for females revealed 22 out of 26 statistically significant coefficients. This depicted that 22 of the 26 presumed direct effects on endogenous variables from other variables were statistically significant at the 5% level of significance. The statistically significant coefficients for the structural portion of the model for females ranged from 0.024 (small effect) to 0.309 (medium effect) (Kline, 2011). The four nonsignificant coefficients for males were the direct effects from SES to exercise, SES to heart rate, exercise to DBP and heart rate to SBP. However the four nonsignificant coefficients for females were the direct effects from smoker status to BMI, emotional well-being to DBP, alcohol to SBP and emotional well-being to SBP.

The parameter estimates for males revealed that smoker status, exercise, alcohol consumption, emotional well-being, SES and BMI all had a significant direct effect on SBP at the 5% level of significance. The direct effects of emotional well-being ($\beta = 0.032$, $p = 0.039$), exercise ($\beta = 0.037$, $p = 0.002$), alcohol consumption ($\beta = 0.051$, $p < 0.001$), smoker status ($\beta = 0.09$, $p < 0.001$) and SES($\beta = 0.14$, $p < 0.001$) on SBP were very small for males. These results demonstrated that the more cigarettes a male smokes and the more alcohol he consumes, the higher his SBP. Similarly the more occasions that a male is not feeling emotionally well, the higher his SBP. In addition, when males have a higher SES or exercise more, they have a higher SBP. The direct effect of BMI ($\beta = 0.253$, $p < 0.001$) on SBP for males was the largest of all the parameters that were tested in the model. This result reveals that BMI has the greatest direct effect on the SBP for males (Byrne, 2010) and males with a higher BMI tend to have a higher SBP.

Similarly, the parameter estimates for males revealed that smoker status, heart rate, alcohol consumption, emotional well-being, SES and BMI all had a significant direct effect on DBP at the 5% level of significance. The direct effects of emotional well-being ($\beta = 0.035$, $p = 0.01$), smoker status ($\beta = 0.039$, $p = 0.003$), alcohol consumption

($\beta = 0.05$, $p < 0.001$) on DBP were very small for males. In addition, heart rate ($\beta = 0.141$, $p < 0.001$) and SES ($\beta = 0.175$, $p < 0.001$) also had a relatively small direct effect on DBP for males. These results demonstrated that the more cigarettes a male smokes and the more alcohol he consumes, the higher his DBP. Similarly the more occasions that a male is not feeling emotionally well, the higher his DBP. In addition, when males have a higher SES or higher resting heart rate, they have a higher DBP. The direct effect of BMI ($\beta = 0.228$, $p < 0.001$) on DBP for males was the largest of all the parameters that were tested in the model. This result reveals that BMI has the greatest direct effect on the DBP for males (Byrne, 2010) and males with a higher BMI tend to have a higher DBP.

The parameter estimates for females demonstrated that smoker status, exercise, heart rate, SES and BMI all had a statistically significant direct effect on SBP at the 5% level of significance. The direct effects of heart rate ($\beta = -0.064$, $p < 0.001$), exercise ($\beta = 0.071$, $p < 0.001$), smoker status ($\beta = 0.078$, $p < 0.001$) and SES ($\beta = 0.074$, $p < 0.001$) on SBP were very small for females. These results revealed that the more cigarettes a female smokes and the lower her heart rate is, the higher her SBP. In addition, when females have a higher SES or exercise more, they have a higher SBP. The direct effect of BMI ($\beta = 0.273$, $p < 0.001$) on SBP for females was the largest of all the parameters that were tested in the model. This result reveals that BMI has the greatest direct effect on the SBP for females (Byrne, 2010) and females with a higher BMI tend to have a higher SBP.

Furthermore, the parameter estimates for females revealed that smoker status, heart rate, alcohol consumption, exercise, SES and BMI all had a statistically significant direct effect on DBP at the 5% level of significance. The direct effects of alcohol consumption ($\beta = 0.038$, $p < 0.001$), heart rate ($\beta = 0.041$, $p < 0.001$), smoker status ($\beta = 0.047$, $p < 0.001$) and exercise ($\beta = 0.047$, $p < 0.001$) on DBP were very small for females. In addition, SES ($\beta = 0.116$, $p < 0.001$) also had a relatively small direct effect on DBP for females. These results demonstrated that the more cigarettes a female smokes, the more alcohol she consumes and the more she exercises, the higher her DBP. In addition, when females have a higher SES or higher resting heart rate, they have a higher DBP. The direct effect of BMI ($\beta = 0.273$, $p < 0.001$) on DBP for females was the largest of all the parameters that were tested in the model. This result reveals that BMI has the greatest

direct effect on the DBP for females (Byrne, 2010) and females with a higher BMI tend to have a higher DBP.

Besides the direct effects, the indirect effects of SES on SBP and SES on DBP were of primary interest in this study. Bootstrapping resampling was conducted to derive the 95% bias-corrected confidence estimates (MacKinnon *et al.*, 2004; Preacher and Hayes, 2004). The standardised indirect effect is the difference between the standardised total and standardised direct effects through the five mediators (Preacher and Hayes, 2008). The standardised indirect effects of the mediated SEM model 2.1 for males and females are presented in Table I.4 and Table I.11 in Appendix I. The p-value for the 95% bias-corrected confidence interval for males and females is displayed in Table I.5 and Table I.12 in Appendix I.

These results demonstrate that the standardised total indirect effect of SES on SBP for males is 0.118 with p=0.003 for the bias-corrected bootstrap method. Similarly the standardised total indirect effect of SES on DBP for males is 0.101 with p=0.005 for the bias-corrected bootstrap method. It is thus evident that the total indirect effect of SES on SBP and SES on DBP for males is significantly different from zero at $p < 0.05$ (two tailed). Thus, we can reject the null hypothesis that the total indirect effect is zero (Hayes, 2009). This implies that the bio-behavioural risk factors do mediate the effect of SES on SBP and SES on DBP for males.

In multiple mediation models, one is not only concerned about the total indirect effect of SES on SBP or SES on DBP but also with the specific indirect effects (Preacher and Hayes, 2008). The specific indirect effects were derived in SPSS AMOS by using the user-defined estimand function developed by Gaskin (2016). The specific indirect effects for males are presented in Table I.6 in Appendix I. After examining the specific indirect effects for males, it is evident that BMI ($\beta = 0.0941$, $p = 0.001$), alcohol consumption ($\beta = 0.0113$, $p = 0.009$), smoker status ($\beta = 0.0193$, $p = 0.001$) and emotional well-being ($\beta = -0.0011$, $p = 0.049$) are the only mediators in the relationship between SES and SBP. This is evident since the p-value for the 95% bias-corrected confidence interval for males is less than 0.05. Furthermore in each of these cases, the confidence interval does not contain zero. Neither exercise nor heart rate contributes to the indirect effect of SES on SBP for males since the specific indirect effects of SES on SBP through exercise and heart rate were not significantly different from zero. Similarly, it is evident that BMI

($\beta = 0.0848$, $p = 0.002$), alcohol consumption ($\beta = 0.0111$, $p = 0.008$), smoker status ($\beta = 0.0083$, $p = 0.007$) and emotional well-being ($\beta = -0.0012$, $p = 0.049$) are the only mediators in the relationship between SES and DBP. This is evident since the p-value for the 95% bias-corrected confidence interval for males is less than 0.05. Furthermore in each of these cases, the confidence interval does not contain zero. Neither exercise nor heart rate contributes to the indirect effect of SES on DBP for males since the specific indirect effects of SES on DBP through exercise and heart rate were not significantly different from zero.

Furthermore the standardised total indirect effect of SES on SBP for females is 0.094 with p=0.004 for the bias-corrected bootstrap method. Similarly the standardised total indirect effect of SES on DBP for females is 0.087 with p=0.005 for the bias-corrected bootstrap method. It is thus evident that the total indirect effect of SES on SBP and SES on DBP for females is significantly different from zero at $p < 0.05$ (two tailed). Thus, we can reject the null hypothesis that the total indirect effect is zero (Hayes, 2009). This implies that the bio-behavioural risk factors do mediate the effect of SES on SBP and SES on DBP for females.

The specific indirect effects for females are presented in Table I.13 in Appendix I. After examining the specific indirect effects for females, it is evident that BMI ($\beta = 0.0844$, $p = 0.005$), smoker status ($\beta = 0.0030$, $p = 0.008$), exercise ($\beta = 0.0044$, $p = 0.001$) and heart rate ($\beta = 0.0033$, $p = 0.005$) are the only mediators in the relationship between SES and SBP. This is evident since the p-value for the 95% bias-corrected confidence interval for females is less than 0.05. Furthermore in each of these cases, the confidence interval does not contain zero. Neither alcohol consumption nor emotional well-being contributes to the indirect effect of SES on SBP for females since the specific indirect effects of SES on SBP through alcohol consumption and emotional well-being were not significantly different from zero. Similarly, it is evident that BMI ($\beta = 0.0844$, $p = 0.007$), alcohol consumption ($\beta = 0.0023$, $p = 0.005$), smoker status ($\beta = 0.0018$, $p = 0.008$), exercise ($\beta = 0.0029$, $p = 0.001$) and heart rate ($\beta = -0.0021$, $p = 0.006$) are mediators in the relationship between SES and DBP. This is evident since the p-value for the 95% bias-corrected confidence interval for females is less than 0.05. Furthermore in each of these cases, the confidence interval does not contain zero. Emotional well-being does not contribute to the indirect effect of SES on DBP for females since the specific indirect

effect of SES on DBP through emotional well-being was not significantly different from zero.

The total standardised effect of a variable is the sum of its total direct effects and the indirect effects. In other words, according to Kline (2011), total effects are the amount of effects through all presumed pathways. The total standardised effects for males and females are presented in Table I.7 and Table I.14 in Appendix I. Since SBP and DBP are the outcome variables in the study, the total effect on each of these variables is of primary interest in the study. The total standardised effects revealed that BMI ($\beta = 0.253$, $= 0.002$) and SES ($\beta = 0.258$, $p = 0.012$) had the biggest statistically significant contribution to the prediction of SBP through all the presumed pathways for males. Similarly BMI ($\beta = 0.228$, $p = 0.003$) and SES ($\beta = 0.277$, $p = 0.007$) had the largest statistically significant contribution to the prediction of DBP through all the presumed pathways for males. Furthermore the total standardised effects demonstrated that BMI ($\beta = 0.273$, $p = 0.007$) and SES ($\beta = 0.168$, $p = 0.005$) had the biggest statistically significant contribution to the prediction of SBP through all the presumed pathways for females. Similarly BMI ($\beta = 0.273$, $p = 0.008$) and SES ($\beta = 0.203$, $p = 0.005$) had the largest statistically significant contribution to the prediction of DBP through all the presumed pathways for females. Exercise, emotional well-being, alcohol consumption and heart rate were not statistically significant contributors to the prediction of SBP for males through all of their presumed pathways. Furthermore, alcohol consumption and emotional well-being were not statistically significant contributors to the prediction of SBP for females through all of their presumed pathways. Emotional well-being did not have a statistically significant contribution to the prediction of DBP for males through all of their presumed pathways while alcohol consumption and emotional well-being did not have a statistically significant contribution to the prediction of DBP for females through all of their presumed pathways.

The squared multiple correlation coefficients were inspected in order to assess the amount of variance in SBP and DBP that were explained by the model (Table I.3 and Table I.10 in Appendix I). These results showed that the hypothesised model for males explained 13.1% of the variation in SBP and explained 14.7% of the variation in DBP. Similarly the squared multiple correlation coefficients revealed that the hypothesised model for females explained 10.9% of the variation in SBP and explained 11.6% of the variation in DBP.

The next step in the analysis was to take the confounding variables such as, age and the use of antihypertensive medication into account. The following tested models were used to investigate the mediation relationship for males and females including age and the use of antihypertensive medication:

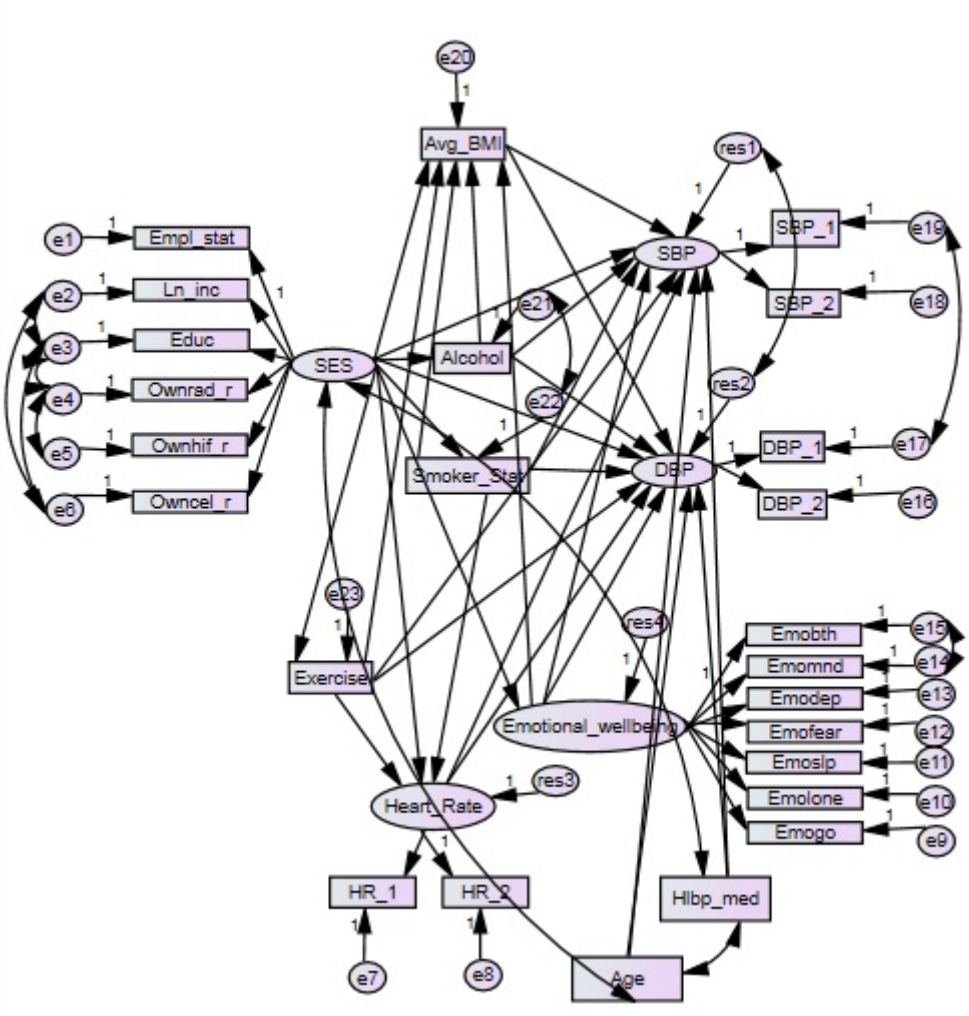


Figure 4.8: Model investigating mediated relationship of SES and blood pressure for males including controls

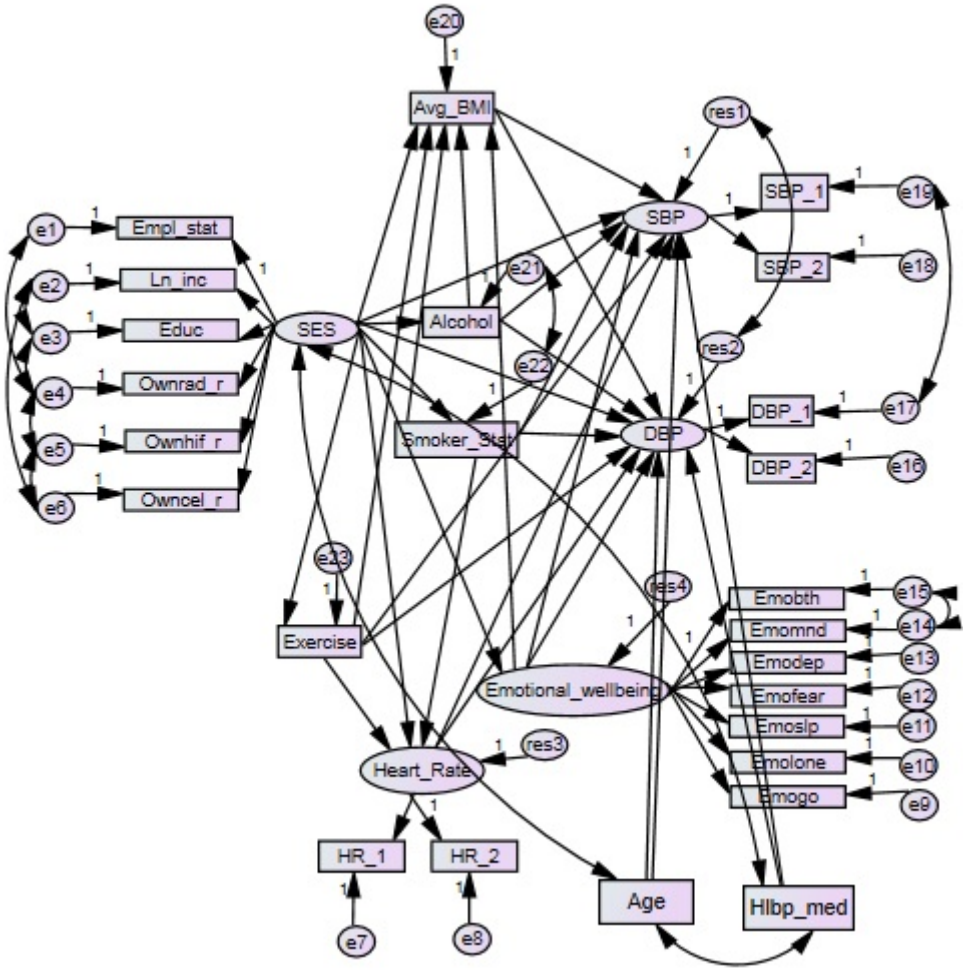


Figure 4.9: Model investigating mediated relationship of SES and blood pressure for females including controls

These models now include a direct effect between age and DBP and age and SBP. Similarly a direct effect between the use of antihypertensive medication and DBP and the use of antihypertensive medication and SBP has been included. This is since these confounding variables are expected to have an effect on SBP and DBP thus these effects need to be controlled so that the relationships of interest can be explored (Ploubidis *et al.*, 2013; Bardenheier *et al.*, 2013; Miyaki *et al.*, 2013; Cois and Ehrlich, 2014; Keetile *et al.*, 2015). In addition, a covariance has been included between age and SES, age and the use of antihypertensive medication and medication and SES as these controls are expected to have an effect on the SES of an individual.

The following tables highlight the model fit indices for the mediated models including age and the use of antihypertensive medication for males and females:

Table 4.10: Model fit indices for SEM model 2.2 for males

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
3317.032	13.879	0.922	13.555	0.0476	0.043	0.68	0.907

Table 4.11: Model fit indices for SEM model 2.2 for females

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
5645.656	23.721	0.906	15.905	0.0564	0.047	0.622	0.90

When controlling for the use of antihypertensive medication and age, the chi-square value for males was significant ($\chi^2(239) = 3317.032$, $p < 0.001$) indicating that the mediated model for males predicted relations that were significantly different from the relations observed in the sample. The normed chi-square value, which was 13.879, indicated a poor fit (Wheaton *et al.*, 1977). The CFI = 0.922, SRMR = 0.0476, RMSEA = 0.043, PNFI = 0.68 and TLI = 0.907 indicated a good fit of SEM model 2.2 for males (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). Similarly when taking the controls into account, the chi-square value for females was significant ($\chi^2(238) = 5645.656$, $p < 0.001$) and the normed chi-square which was 23.721 indicated a poor fit (Wheaton *et al.*, 1977). However the CFI = 0.906, SRMR = 0.0564, RMSEA = 0.047, PNFI = 0.622 and TLI = 0.90 depicted a relatively good fit of the SEM model 2.2 for females (Mulaik *et al.*, 1989; Hu and Bentler, 1999; Diamantopoulos and Siguaw, 2000). The modification indices for both males and females did not depict any further parameters which were meaningful or relevant for inclusion in the models. Since most of the fit indices for the mediated models, when controlling for the use of antihypertensive medication and age, for both males and females indicated that the fit of the models were adequate, the parameter estimates were then investigated.

The unstandardised path coefficients of the mediated SEM model 2.2 for males and females are presented in Table I.15 and Table I.22 in Appendix I. The unstandardised coefficients for both males and females showed that all the indicators' loadings on their respective latent variable were statistically significant at the 5% level of significance. The standardised parameter estimates for both the measurement and the structural portion of the model for males and females, when controlling for the use of antihypertensive medication and age, are presented in Table I.16 and Table I.23 in Appendix I. These

results indicated that all factor loadings of the measurement portion ranged from 0.062 (small effect) to 0.979 (large effect) for males and ranged from 0.199 (small effect) to 0.979 (large effect) for females. In addition, each of the indicators had a positive loading on their respective latent variable for males and females however education had a negative loading on SES for females. The standardised parameter estimates for the structural portion of the model for males revealed 23 out of 30 statistically significant coefficients. This indicated that 23 of the 30 presumed direct effects on endogenous variables from other variables were statistically significant at the 5% level of significance. The statistically significant coefficients for the structural portion of the model including controls ranged from 0.027 (small effect) to 0.459 (medium effect) for males. However the standardised parameter estimates for the structural portion of the model for females revealed 22 out of 30 statistically significant coefficients. This implied that 22 of the 30 presumed direct effects on endogenous variables from other variables were statistically significant at the 5% level of significance. The statistically significant coefficients for the structural portion of the model for females ranged from 0.023 (small effect) to 0.481 (large effect). The seven nonsignificant coefficients for males when including the controls were the direct effects from smoker status to SBP, smoker status to DBP, exercise to DBP, heart rate to SBP, emotional well-being to SBP, SES to DBP and SES to SBP. Since the direct effects of SES to DBP and SES to SBP are not statistically significant for males when including the mediators, this proves that the relationship between SES and SBP and SES and DBP is fully mediated by the statistically significant bio-behavioural risk factors when controlling for the use of antihypertensive medication and age (James and Brett, 1984). Furthermore, the eight nonsignificant coefficients for females were the direct effects from emotional well-being to DBP, emotional well-being to SBP, emotional well-being to BMI, exercise to DBP, smoker status to DBP, smoker status to heart rate, alcohol to SBP and SES to SBP. Since the direct effects of SES to SBP are not statistically significant for females when including the mediators, this proves that the relationship between SES and SBP is fully mediated by the statistically significant bio-behavioural risk factors when controlling for the use of antihypertensive medication and age (James and Brett, 1984). However the direct effects of SES to DBP are partially mediated by statistically significant bio-behavioural risk factors for females when taking the controls into account (Preacher and Hayes, 2004).

The standardised parameter estimates for males, when controlling for the use of antihypertensive medication and age, revealed that exercise, alcohol consumption, BMI, the use of antihypertensive medication and age all had a statistically significant direct effect on SBP at the 5% level of significance. The direct effects of exercise ($\beta = 0.065$, $p < 0.001$) and alcohol consumption ($\beta = 0.054$, $p < 0.001$) on SBP were very small for males. In addition, the direct effects BMI ($\beta = 0.134$, $p < 0.001$), the use of antihypertensive medication ($\beta = -0.105$, $p < 0.001$) and age ($\beta = 0.338$, $p < 0.001$) on SBP were higher for males. These results demonstrated that age has the largest effect on SBP for males (Byrne, 2010) and older males tend to have a higher SBP. In addition, the more alcohol a male consumes or the more he exercises, the higher his SBP. The direct effect of BMI on SBP for males was the largest of all the mediators that were tested in the model (Byrne, 2010). This result reveals that males with a higher BMI tend to have a higher SBP. Furthermore, males that use antihypertensive medication have a higher SBP.

Similarly, the parameter estimates for males, when controlling for the use of antihypertensive medication and age, revealed that alcohol consumption, heart rate, emotional well-being, BMI, the use of antihypertensive medication and age all had a statistically significant direct effect on DBP at the 5% level of significance. The direct effects of emotional well-being ($\beta = 0.027$, $p = 0.044$), alcohol consumption ($\beta = 0.044$, $p < 0.001$) and the use of antihypertensive medication ($\beta = -0.083$, $p < 0.001$) on DBP were very small for males. In addition, heart rate ($\beta = 0.127$, $p < 0.001$), BMI ($\beta = 0.164$, $p < 0.001$) and age ($\beta = 0.277$, $p < 0.001$) had larger direct effects on DBP for males (Byrne, 2010). These results demonstrated that the more alcohol a male consumes or the more occasions he is not feeling emotionally well, the higher his DBP. In addition, males that have a higher resting heart rate or use antihypertensive medication, have a higher DBP. Age has the largest effect on DBP for males (Byrne, 2010) and older males tend to have a higher DBP. The direct effect of BMI on DBP for males was the largest of all the mediators that were tested in the model, after controlling for the use of antihypertensive medication and age (Byrne, 2010). This result reveals that males with a higher BMI tend to have a higher DBP.

The standardised parameter estimates for females, when taking the controls into account, revealed that smoker status, exercise, heart rate, BMI, the use of antihypertensive medication and age all had a statistically significant direct effect on SBP at the 5% level of significance. The direct effects of exercise ($\beta = 0.026$, $p = 0.003$), smoker status

($\beta = 0.027$, $p = 0.04$) and heart rate ($\beta = -0.045$, $p < 0.001$) on SBP were very small for females. In addition, the direct effects of BMI ($\beta = 0.149$, $p < 0.001$), medication ($\beta = -0.201$, $p < 0.001$) and age ($\beta = 0.387$, $p < 0.001$) on SBP were higher for females. These results demonstrated that age has the largest effect on SBP for females and older females tend to have a higher SBP. In addition, the more cigarettes a female smokes or the more she exercises, the higher her SBP. However, females with a lower resting heart rate tend to have a higher SBP. The direct effect of BMI on SBP for females was the largest of all the mediators that were tested in the model (Byrne, 2010). This result reveals that females with a higher BMI tend to have a higher SBP. Furthermore, females that use antihypertensive medication have a higher SBP.

Moreover, the parameter estimates for females, when controlling for the use of antihypertensive medication and age, revealed that heart rate, alcohol consumption, SES, BMI, the use of antihypertensive medication and age all had a statistically significant direct effect on DBP at the 5% level of significance. The direct effects of alcohol consumption ($\beta = 0.023$, $p = 0.022$) and heart rate ($\beta = 0.042$, $p < 0.001$) on DBP were very small for females. In addition, SES ($\beta = 0.114$, $p < 0.001$), the use of antihypertensive medication (β=-0.16, $p < 0.001$), BMI ($\beta = 0.174$, $p < 0.001$) and age ($\beta = 0.215$, $p < 0.001$) had larger direct effects on DBP for females. These results demonstrated that the more alcohol a female consumes or the higher her resting heart rate is, the higher her DBP. In addition, when females have a higher SES or use antihypertensive medication, they have a higher DBP. The direct effect of BMI on DBP for females was the largest of all the mediators that were tested in the model (Byrne, 2010). This result reveals that females with a higher BMI tend to have a higher DBP. Furthermore, age had the largest overall direct effect on DBP for females (Byrne, 2010) and DBP tends to increase with age.

Thereafter, the indirect effects of SES on SBP and SES on DBP, after controlling for the use of antihypertensive medication and age, were examined as they were of primary interest in this study. The standardised indirect effects of the mediated SEM model 2.2 for males and females are presented in Table I.18 and Table I.25 in Appendix I. The p-value for the 95% bias-corrected confidence interval for males and females is displayed in Table I.19 and Table I.26 in Appendix I.

These results demonstrate that the standardised total indirect effect of SES on SBP for males, when controlling for the use of antihypertensive medication and age, is 0.068

with p=0.006 for the bias-corrected bootstrap method. Similarly the standardised total indirect effect of SES on DBP for males, when controlling for the use of antihypertensive medication and age, is 0.085 with p=0.005 for the bias-corrected bootstrap method. It is thus evident that the total indirect effect of SES on SBP and SES on DBP for males, when controlling for the use of antihypertensive medication and age, is significantly different from zero at $p < 0.05$ (two tailed). Thus, we can reject the null hypothesis that the total indirect effect is zero (Hayes, 2009). This implies that the bio-behavioural risk factors do mediate the effect of SES on SBP and SES on DBP for males when taking confounding factors such as the use of antihypertensive medication and age into account.

The specific indirect effects for males, when controlling for the use of antihypertensive medication and age, are presented in Table I.20 in Appendix I. After examining the specific indirect effects for males, it is evident that BMI ($\beta = 0.0615$, $p = 0.007$), alcohol consumption ($\beta = 0.0139$, $p = 0.006$) and exercise ($\beta = -0.0075$, $p = 0.004$) are the only mediators in the relationship between SES and SBP when controlling for the use of antihypertensive medication and age. This is evident since the p-value for the 95% bias-corrected confidence interval for males is less than 0.05. Furthermore in each of these cases, the confidence interval does not contain zero. Smoker status, heart rate and emotional well-being do not contribute to the indirect effect of SES on SBP for males, when controlling for the use of antihypertensive medication and age. This is evident since the specific indirect effects of SES on SBP through smoker status, heart rate and emotional well-being were not significantly different from zero. Similarly, it is clear that BMI ($\beta = 0.0753$, $p = 0.003$), alcohol consumption ($\beta = 0.0114$, $p = 0.008$), heart rate ($\beta = 0.0085$, $p = 0.001$) and emotional well-being ($\beta = 0.0010$, $p = 0.049$) are the only mediators in the relationship between SES and DBP when controlling for the use of antihypertensive medication and age. This is evident since the p-value for the 95% bias-corrected confidence interval for males is less than 0.05. Furthermore in each of these cases, the confidence interval does not contain zero. Neither smoker status nor exercise contributes to the indirect effect of SES on DBP for males since the specific indirect effects of SES on DBP through smoker status and exercise were not significantly different from zero.

Furthermore the standardised total indirect effect of SES on SBP for females, when controlling for the use of antihypertensive medication and age, is 0.076 with p=0.006 for the bias-corrected bootstrap method. Similarly the standardised total indirect effect of

SES on DBP for females, when controlling for the use of antihypertensive medication and age, is 0.081 with p=0.008 for the bias-corrected bootstrap method. It is thus evident that the total indirect effect of SES on SBP and SES on DBP for females, when taking the controls into account, is significantly different from zero at $p < 0.05$ (two tailed). Thus, we can reject the null hypothesis that the total indirect effect is zero (Hayes, 2009). This implies that the bio-behavioural risk factors do mediate the effect of SES on SBP and SES on DBP for females when taking confounding factors such as the use of antihypertensive medication and age into account.

The specific indirect effects for females, when controlling for the use of antihypertensive medication and age, are presented in Table I.27 in Appendix I. After examining the specific indirect effects for females, it is evident that BMI ($\beta = 0.0717$, $p = 0.006$), exercise ($\beta = -0.0012$, $p = 0.018$) and heart rate ($\beta = 0.0035$, $p = 0.007$) are the only mediators in the relationship between SES and SBP when controlling for the use of antihypertensive medication and age. This is evident since the p-value for the 95% bias-corrected confidence interval for females is less than 0.05. Furthermore in each of these cases, the confidence interval does not contain zero. Alcohol consumption, smoker status and emotional well-being do not contribute to the indirect effect of SES on SBP for females since the specific indirect effects of SES on SBP through alcohol consumption, smoker status and emotional well-being were not significantly different from zero. Similarly, it is evident that BMI ($\beta = 0.0837$, $p = 0.011$), alcohol consumption ($\beta = 0.0012$, $p = 0.038$) and heart rate ($\beta = -0.0032$, $p = 0.003$) are the only mediators in the relationship between SES and DBP when controlling for the use of antihypertensive medication and age. This is evident since the p-value for the 95% bias-corrected confidence interval for females is less than 0.05. Furthermore in each of these cases, the confidence interval does not contain zero. Smoker status, exercise and emotional well-being do not contribute to the indirect effect of SES on DBP for females since the specific indirect effect of SES on DBP through smoker status, exercise and emotional well-being were not significantly different from zero.

The total standardised effects for males and females in the models, including the controls, were then examined. The total standardised effects for males and females are presented in Table I.21 and Table I.28 in Appendix I. Since SBP and DBP are the outcome variables in the study, the total effect on each of these variables is of primary interest in the study. The total standardised effects revealed that after taking the controls into

account, BMI ($\beta = 0.134$, $p = 0.008$) had the biggest statistically significant contribution to the prediction of SBP through all the presumed pathways for males (Byrne, 2010). In addition, BMI ($\beta = 0.164$, $p = 0.003$), SES ($\beta = 0.129$, $p = 0.006$) and heart rate ($\beta = 0.127$, $p = 0.001$) had the largest statistically significant contribution to the prediction of DBP through all the presumed pathways for males (Byrne, 2010). The total standardised effects demonstrated that BMI (β=0.149, $p = 0.006$) had the biggest statistically significant contribution to the prediction of SBP through all the presumed pathways for females after controlling for the use of antihypertensive medication and age (Byrne, 2010). However BMI ($\beta = 0.174$, $p = 0.009$) and SES ($\beta = 0.196$, $p = 0.013$) had the largest statistically significant contribution to the prediction of DBP through all the presumed pathways for females after controlling for the use of antihypertensive medication and age (Byrne, 2010).

Smoker status, emotional well-being and heart rate were not statistically significant contributors to the prediction of SBP for males through all of their presumed pathways. Furthermore, exercise, alcohol consumption, smoker status and emotional well-being were not statistically significant contributors to the prediction of DBP for males through all of their presumed pathways. As a result, after taking age and the use of antihypertensive medication into account, SES, exercise, alcohol consumption and BMI all had a statistically significant total effect on SBP for males through all of their presumed pathways. These results revealed that males with a higher SES, a higher BMI, that consume more alcohol or exercise more frequently have a higher SBP. Similarly, after controlling for the use of antihypertensive medication and age, SES, BMI and heart rate had significant total effects on DBP for males through all of their presumed pathways. These results demonstrated that males with a higher SES, have a higher BMI or have a higher resting heart rate tend to have a higher DBP. For females, exercise, emotional well-being, alcohol consumption and smoker status were not statistically significant contributors to the prediction of SBP and DBP for females through all of their presumed pathways. As a result, BMI, heart rate and SES had a statistically significant contribution to the prediction of blood pressure for females through all of their presumed pathways.

The squared multiple correlation coefficients were inspected to examine the amount of variance in SBP and DBP that were explained by the model when controlling for the use of antihypertensive medication and age (Table I.17 and Table I.24 in Appendix I). These results showed that after taking the controls into account, the hypothesised model for

males explained 21% of the variation in SBP and explained 20.8% of the variation in DBP. Similarly the squared multiple correlation coefficients after controlling for the use of antihypertensive medication and age revealed that the hypothesised model for females explained 32.5% of the variation in SBP and explained 24.9% of the variation in DBP.

4.4 Multiple linear regression

One of the aims of this analysis was to assess the relationships between demographic, behavioural, socio-economic variables and blood pressure. This research question was addressed by using multiple linear regression. The data was first subdivided into a training set, validation set and test set for both males and females. Simple random sampling was applied to the data in order to subdivide it into the training set, validation set and test set (Reitermanova, 2010). This is a cross validation approach that enables one to balance between minimal bias and minimal variance (Reitermanova, 2010). The training set comprised of 40% of the data while the test and validation sets each comprised of 30% of the data (Hair *et al.*, 1998; Harrell, 2015). This was done so that the results from the training, validation and testing datasets could be compared in terms of the MSE in order to choose the best fitting model. The best fitting models would then be compared against the results from SEM.

Two separate multiple linear regression models were analysed for both genders since the burden of hypertension and its risk factors often vary across genders (Wong *et al.*, 2015). SBP was considered as a dependent variable in one of the regression models analysed while DBP was considered as a dependent variable in a subsequent model analysed. The following factors were considered as predictor variables in the regression analysis: age, the use of antihypertensive medication, income, education, employment status, ownership of cellphone, ownership of radio, ownership of hifi, BMI, resting heart rate, smoker status, alcohol consumption, physical activity and the emotional well-being of individuals. The average of each of the seven variables that loaded onto the emotional well-being latent variable in SEM was used as the emotional well-being variable in the regression analysis. The average of SBP reading 1 and SBP reading 2 were used to form the SBP variable (Silva, Scheines, Glymour and Spirtes, 2006). Similarly the average of DBP reading 1 and DBP reading 2 formed the DBP variable while the average of heart rate reading 1

and heart rate reading 2 formed the heart rate variable for the regression analysis (Silva *et al.*, 2006).

Age and the use of antihypertensive medication are potentially confounding variables which may have an effect on the dependent variables (Miyaki *et al.*, 2013; Bardenheier *et al.*, 2013; Ploubidis *et al.*, 2013; Cois and Ehrlich, 2014; Keetile *et al.*, 2015). Controlling for these variables enables one to balance their effects across subjects and groups so that one can ignore them and just study the association between the independent and dependent variables of interest (Becker, 2005). In regression analysis, confounding variables are treated as additional independent variables in the analysis (McNamee, 2005; Pourhoseingholi, Baghestani and Vahedi, 2012). The comparison of the models including confounding variables and excluding confounding variables can give an indication of how much the confounders in the model distort the relationship between the independent and dependent variables (McNamee, 2005; Pourhoseingholi *et al.*, 2012). As a result, the use of antihypertensive medication and age were first excluded as predictors from the models in order to assess the difference in the model fit when including and excluding these variables as predictors. Therefore, 24 different multiple linear regression models were built. Interaction effects are included in a regression analysis when a moderator variable or a second independent variable changes the form of the relationship between another independent variable and the dependent variable (Hair *et al.*, 1998). Since there were no variables that were moderators in the analysis, each regression model only included the main effects.

The following table displays a description of each multiple linear regression model built:

Table 4.12: Description of each model built

Model	**Gender**	**Data used**	**Description**	**Dependent variable**
1	Males	Training	Model without age and medication	SBP
2	Males	Validation	Model without age and medication	SBP
3	Males	Testing	Model without age and medication	SBP
4	Males	Training	Model with age and medication	SBP
5	Males	Validation	Model with age and medication	SBP
6	Males	Testing	Model with age and medication	SBP
7	Males	Training	Model without age and medication	DBP
8	Males	Validation	Model without age and medication	DBP

Model	Gender	Data used	Description	Dependent variable
9	Males	Testing	Model without age and medication	DBP
10	Males	Training	Model with age and medication	DBP
11	Males	Validation	Model with age and medication	DBP
12	Males	Testing	Model with age and medication	DBP
13	Females	Training	Model without age and medication	SBP
14	Females	Validation	Model without age and medication	SBP
15	Females	Testing	Model without age and medication	SBP
16	Females	Training	Model with age and medication	SBP
17	Females	Validation	Model with age and medication	SBP
18	Females	Testing	Model with age and medication	SBP
19	Females	Training	Model without age and medication	DBP
20	Females	Validation	Model without age and medication	DBP
21	Females	Testing	Model without age and medication	DBP
22	Females	Training	Model with age and medication	DBP
23	Females	Validation	Model with age and medication	DBP
24	Females	Testing	Model with age and medication	DBP

Each of the categorical variables were recoded as dummy or indicator variables so that they could be included in the regression analysis. Each level of the categorical variables was represented by either a one or zero. According to Hair *et al.* (1998), one dummy variable less than the number of levels of the categorical variable is required for any regression analysis. The omitted category is then termed the reference category (Hair *et al.*, 1998). The coding of each of the categorical variables into dummy variables is presented in Table J.1 in Appendix J.

Stepwise regression was applied to each model to enable the choice of the predictive model to be carried out by an automatic procedure in the form of a sequence of F-tests. Stepwise regression differs from the forward-selection method in that variables already in the model do not necessarily remain in the model (Hair *et al.*, 1998). Variables are added one at a time to the model as in the forward-selection method, and the F-statistic for a variable must be significant in order to be added to the model. After a variable has been added to the model, the stepwise method then looks at all the variables that

are in the model already and deletes any variable that does not result in a significant F-statistic (Hair *et al.*, 1998). Backward elimination is a method that starts by including all of the predictors in the model and then eliminating variables that do not make a significant contribution to the prediction (Hair *et al.*, 1998). Thus stepwise regression is a combination of both backward elimination and forward-selection and is often the preferred approach among researchers since one is able to both add and delete variables at each stage (Hair *et al.*, 1998). Due to the efficiency of selecting independent variables that maximize predictive accuracy and the ability to both add and delete variables at each stage, stepwise regression was chosen as the preferred method for the regression analysis (Hair *et al.*, 1998).

Table J.2 in Appendix J highlights each of the independent variables used after stepwise regression was applied. In each regression model that was built, BMI was significant in the prediction of SBP and DBP. Furthermore, when the use of antihypertensive medication and age were included in the regression model, each of these independent variables were significant predictors of SBP and DBP.

4.4.1 Discussion of multiple linear regression results

The following table depicts a summary of the fit of each multiple linear regression model:

Table 4.13: Fit of each multiple linear regression model

Model	R square	Adjusted R square	Root MSE	MSE
1	0.146	0.141	18.429	339.634
2	0.228	0.223	18.104	327.762
3	0.133	0.129	**17.386**	**302.256**
4	0.225	0.221	17.547	307.893
5	0.333	0.330	16.816	282.788
6	0.226	0.226	**16.423**	**269.718**
7	0.131	0.126	12.092	146.210
8	0.206	0.201	11.761	138.317
9	0.149	0.142	**11.690**	**136.645**
10	0.168	0.162	11.842	140.241
11	0.259	0.255	**11.358**	**129.009**

Model	R square	Adjusted R square	Root MSE	MSE
12	0.181	0.174	11.470	131.563
13	0.205	0.202	19.958	398.324
14	0.216	0.213	20.658	426.755
15	0.235	0.232	**18.905**	**357.400**
16	0.320	0.318	18.440	340.017
17	0.404	0.401	18.022	324.798
18	0.365	0.362	**17.231**	**296.910**
19	0.144	0.142	12.447	154.934
20	0.183	0.180	12.558	157.715
21	0.181	0.178	**12.002**	**144.037**
22	0.188	0.185	12.130	147.130
23	0.274	0.274	11.851	140.449
24	0.250	0.245	**11.504**	**132.339**

Table 4.13 depicts some instances where the model with the lowest MSE does not have the highest adjusted R-square value. It has been shown in literature that the R-squared value is not always a reliable guide to model quality thus the lowest MSE will instead be used as a measure of the model quality (Kvalseth, 1985). The root MSE (RMSE) is a good indication of how accurately the model predicts the response, and is the most important criterion for fit if the main purpose of the model is prediction (Hair *et al.*, 1998).

When assessing the models excluding age and the use of antihypertensive medication as predictor variables for males, one observes that the testing dataset (model 3) produces the lowest MSE in the prediction of SBP thus indicating the best model fit. Similarly after the inclusion of age and the use of antihypertensive medication as predictor variables for males, one observes that the testing dataset (model 6) produces the lowest MSE in the prediction of SBP. Furthermore, the testing dataset (model 9) produces the lowest MSE in the prediction of DBP for males when the use of antihypertensive medication and age are excluded from the model. However, when the use of antihypertensive medication and age are included in the model, the validation dataset (model 11) produces the lowest MSE in the prediction of DBP for males. When assessing the models excluding

age and the use of antihypertensive medication as predictor variables for females, one observes that the testing dataset (model 15) produces the lowest MSE in the prediction of SBP. Similarly after the inclusion of age and the use of antihypertensive medication as predictor variables for females, one observes that the testing dataset (model 18) produces the lowest MSE in the prediction of SBP. Furthermore, the testing dataset (model 21) produces the lowest MSE in the prediction of DBP for females when the use of antihypertensive medication and age are excluded from the model. Similarly, when the use of antihypertensive medication and age are included in the model, the testing dataset (model 24) produces the lowest MSE in the prediction of DBP for females. The detailed results for model 3, 6, 9, 11, 15, 18, 21 and 24 will be discussed in more detail below. These results will then be compared against the SEM results in terms of their ability to answer the proposed research questions.

Before accepting a regression model, it is important to examine the influence and fit diagnostics to determine whether the model might be unduly influenced by a few observations and whether the data supports the assumptions that underlie the linear regression. Figures K.1 to K.8 in Appendix K illustrate a panel of diagnostic plots. The graphs in the upper left of each figure depicts the relationship between the residuals and the predicted value. As mentioned in section 4.2, the residuals for each model are distributed around zero which implies that the assumption of homoscedasticity is not violated for any of the linear regression models that were analysed (Tabachnick and Fidell, 2007). Furthermore the residuals do not appear to exhibit any particular pattern which implies independence of the residuals. The plot of the RSTUDENT residuals shows the externally studentized residuals that take into account heterogeneity in the variability of the residuals. RSTUDENT residuals that exceed the threshold values of ± 2 often indicate outlying observations (Hair *et al.*, 1998). According to Stevens (2002), one expects only 5% of standardised residuals to be $> |2|$ if the linear model is appropriate. The plot of the externally studentized residuals by leverage values reveals that there are a few observations with a high leverage that might be overly influencing the fit produced for each of the models analysed. Since less than 5% of the residuals exceeded the threshold value, the linear models were deemed appropriate (Stevens, 2002). The plot of Cook's distance (Cook, 1977) by observation is very useful for identifying influential points. It measures the combined influence of the point being a outlier on the dependent variable and on the set of predictors (Cook, 1977). According to Cook and Weisberg

(1982), a Cook's distance greater than one would generally be considered large. Since none of the models had a Cook's distance greater than one, this implies that there were no highly influential points in any of the regression models analysed (Cook and Weisberg, 1982). The normal-probability Q-Q plot in the second row of each panel shows that there is a linear trend with a slight deviation at the tails for each model, which implies that the assumption of normality is satisfied for each of the regression models that were analysed (Hair *et al.*, 1998). The histogram of the residuals depicts that the assumption of normality is only slightly violated thus the interpretation of the regression results are appropriate for each of the models that were analysed (Hair *et al.*, 1998).

4.4.1.1 SBP multiple linear regression results for males

The following table highlights the stepwise regression results for the model excluding age and the use of antihypertensive medication:

Table 4.14: Test dataset: Stepwise regression results for model 3

Step	Group Entered	Group Removed	Number Vars In	Model R-Square	F Value	Pr > F
1	Avg_BMI		1	0.0695	158.49	<0.0001
2	Education		4	0.0984	22.65	<0.0001
3	Ownrad_r_d2		5	0.1166	43.79	<0.0001
4	Ln_inc		6	0.1233	16.21	<0.0001
5	Alcohol		10	0.1316	5.04	0.0005
6	Emotional_well-being		11	0.1333	4.02	0.0452

Table 4.14 highlights that BMI, education, the ownership of a radio, income, alcohol consumption and emotional well-being are significant predictors of SBP for males. Each of these independent variables have a p-value less than 0.05 indicating sufficient evidence for predicting the SBP.

Table 4.15: Test dataset: Overall summary of results for model 3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	98195	8926.80858	29.53	<0.0001
Error	2113	638666	302.25571		
Corrected Total	2124	736861			
Root MSE	17.3855	**R-Square**	0.1333		
Dependent Mean	124.42094	**Adj R-Sq**	0.1287		
Coeff Var	13.97313				

From Table 4.15 it is evident that the global F-test in the ANOVA table which has a p-value < 0.0001 depicts that the model is significant in the prediction of SBP for males based on the group of independent variables in the model. The value of the R-square is 0.1333 which means that 13.33% of the variation in SBP is explained by the independent variables. The RMSE or standard deviation is 17.39 which implies that approximately 95% of the sampled SBP values are within 2 standard deviations of their respective predicted values.

Table 4.16: Test dataset: Parameter estimates for model 3

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tol	Variance Inflation
Intercept	92.84	2.73	34.01	<0.0001	.	0.00
Emotional_well-being	1.64	0.82	2.00	0.05	0.97	1.03
Ln_inc	0.39	0.11	3.67	0.0002	0.84	1.19
Avg_BMI	0.92	0.09	10.71	<0.0001	0.90	1.11
Ownrad_r_d2	4.13	0.83	5.00	<0.0001	0.87	1.15
Educ_d1	12.25	1.81	6.78	<0.0001	0.60	1.66
Educ_d2	4.35	1.47	2.96	0.003	0.38	2.60
Educ_d3	1.62	1.31	1.24	0.21	0.35	2.89
Alcohol_d1	2.76	1.33	2.07	0.04	0.93	1.07
Alcohol_d2	4.05	1.44	2.81	0.01	0.96	1.04
Alcohol_d3	2.88	1.23	2.35	0.02	0.94	1.07

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tol	Variance Inflation
Alcohol_d4	3.82	1.32	2.90	0.004	0.93	1.08

Based on the t-test with the significant level α equals 0.05, the p-values for the BMI, education, the ownership of a radio, income, alcohol consumption and the emotional well-being of an individual are less than 0.05 indicating that they each make a significant contribution to the predictive power of the model. The constant value (92.84 mmHg) represents the intercept, which is the predicted SBP for males when each independent variable is zero. Each β_i parameter represents the mean change in the SBP for every one unit increase in the corresponding X_i when all the other X's are kept constant. Thus the SBP increases by 0.92 mmHg for every one unit increase in the BMI for males when all other independent variables remain constant. Furthermore the more occasions that a male is not feeling emotional well, the higher the SBP. In addition, as long as the percentage increase in income is fixed, one will observe the same difference in the SBP regardless of where the baseline income is. For example, for a 10% increase in income for males, the difference in the expected mean SBP will always be 0.39443*ln(1.1) = 0.0376 mmHg. Moreover, a male who owns a radio has on average 4.13 mmHg higher SBP than a male who does not own a radio when all other independent variables remain constant. The coefficients of indicator variables represent the difference in the SBP between being in the specified category and being in the reference category when the other variables are all controlled (Hair *et al.*, 1998). After controlling for all other independent variables in the model, there is no statistically significant evidence that males with a secondary education have a different SBP to males with a tertiary education. However on average males with no schooling have 12.25 mmHg higher SBP than males with a tertiary education while males with a primary education have on average 4.35 mmHg higher SBP than males with a tertiary education. Furthermore males who consume any alcoholic drinks irrespective of the number of drinks have a significantly higher SBP than males who do not consume alcoholic drinks.

The following table highlights the stepwise regression results for the model including age and the use of antihypertensive medication:

Table 4.17: Test dataset: Stepwise regression results for model 6

Step	Group Entered	Group Removed	Number Vars In	Model R-Square	F Value	Pr > F
1	Age		1	0.182	472.44	<0.0001
2	Avg_BMI		2	0.2077	68.77	<0.0001
3	Hlbp_med_d1		3	0.2194	31.87	<0.0001
4	Alcohol		7	0.2244	3.41	0.0087

Table 4.17 highlights that age, BMI, the use of antihypertensive medication and alcohol consumption are significant predictors of SBP for males. Each of these independent variables have a p-value less than 0.05 indicating sufficient evidence for predicting the SBP. When the use of antihypertensive medication and age are included in the model, one observes that the BMI and alcohol consumption remain significant predictors however education, the ownership of a radio, income and emotional well-being are no longer significant predictors of SBP for males.

Table 4.18: Test dataset: Overall summary of results for model 6

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	166139	20767	77	<0.0001
Error	2116	570722	269.71758		
Corrected Total	2124	736861			
Root MSE	16.42308	**R-Square**	0.2255		
Dependent Mean	124.42094	**Adj R-Sq**	0.2225		
Coeff Var	13.19961				

Table 4.18 highlights that the global F-test which has a p-value < 0.0001 depicts that the model is significant in the prediction of SBP for males based on the group of independent variables in the model. The value of the R-square is 0.2255 which means that 22.55% of the variation in SBP is explained by the predictor variables. This implies that the model including age and the use of antihypertensive medication explains much more

of the variation in SBP. The RMSE or standard deviation is 16.42 which implies that approximately 95% of the sampled SBP values are within 2 standard deviations of their respective predicted values. Since the RMSE for the model including age and the use of antihypertensive medication is lower, this indicates that the model produces a better fit in the prediction of SBP for males.

Table 4.19: Test dataset: Parameter estimates for model 6

Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr** $>\|t\|$	**Tol**	**Variance Inflation**
Intercept	97.83	2.74	35.76	<0.0001	.	0.00
Avg_BMI	0.61	0.08	7.48	<0.0001	0.90	1.11
Alcohol_d1	2.68	1.25	2.14	0.03	0.94	1.06
Alcohol_d2	3.52	1.36	2.59	0.01	0.96	1.04
Alcohol_d3	1.96	1.16	1.69	0.09	0.94	1.07
Alcohol_d4	3.04	1.23	2.46	0.01	0.95	1.06
Age	0.42	0.03	16.42	<0.0001	0.82	1.22
Hlbp_med_d1	9.76	1.72	5.68	<0.0001	0.86	1.16

Based on the t-test with the significant level α equals 0.05, the p-values for the BMI, age, the use of antihypertensive medication and alcohol consumption are less than 0.05 indicating that they each make a significant contribution to the predictive power of the model. The constant value (97.83 mmHg) represents the intercept, which is the predicted SBP for males when each independent variable is zero. The SBP increases by 0.61 mmHg for every one unit increase in the BMI for males when all other independent variables remain constant. Furthermore the SBP increases by 0.42 mmHg for every one unit increase in age for males. In addition, a male who is taking antihypertensive medication has on average 9.76 mmHg higher SBP than a male who is not taking antihypertensive medication when all other independent variables remain constant. After controlling for all other independent variables in the model, there is no statistically significant evidence that males who consume three or four alcoholic drinks have a different SBP to males who do not consume alcoholic drinks. However on average males who consume seven or more alcoholic drinks have 2.68 mmHg higher SBP than males who do not consume alcoholic drinks. Similarly males who consume five or six alcoholic drinks have 3.52 mmHg higher

SBP than males who do not consume alcoholic drinks. In addition, males who consume one or two alcoholic drinks have 3.04 mmHg higher SBP than males who do not consume alcoholic drinks.

4.4.1.2 DBP multiple linear regression results for males

The following table highlights the stepwise regression results for the model excluding age and the use of antihypertensive medication:

Table 4.20: Test dataset: Stepwise regression results for model 9

Step	**Group Entered**	**Group Removed**	**Number Vars In**	**Model R-Square**	**F Value**	**Pr > F**
1	Avg_BMI		1	0.066	150.06	<0.0001
2	Heart_Rate		2	0.0916	59.67	<0.0001
3	Ownrad_r_d2		3	0.1078	38.63	<0.0001
4	Alcohol		7	0.1217	8.36	<0.0001
5	Education		10	0.134	10	<0.0001
6	Exercise		14	0.1431	5.6	0.0002
7	Ln_inc		15	0.1472	10.14	0.0015

Table 4.20 highlights that BMI, resting heart rate, the ownership of a radio, alcohol consumption, education, exercise and income are significant predictors of DBP for males. Each of these independent variables have a p-value less than 0.05 indicating sufficient evidence for predicting the DBP.

Table 4.21: Test dataset: Overall summary of results for model 9

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	16	50259	3141.2152	22.99	<0.0001
Error	2108	288049	136.64542		
Corrected Total	2124	338308			
Root MSE	11.68954	**R-Square**	0.1486		

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Dependent Mean	80.22071	**Adj R-Sq**	0.1421		
Coeff Var	14.57173				

Table 4.21 illustrates that the global F-test which has a p-value < 0.0001 implies that the model is significant in the prediction of DBP for males based on the group of independent variables in the model. The value of the R-square is 0.1486 which means that 14.86% of the variation in DBP is explained by the independent variables. The RMSE or standard deviation is 11.69 which implies that approximately 95% of the sampled DBP values are within 2 standard deviations of their respective predicted values.

Table 4.22: Test dataset: Parameter estimates for model 9

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tol	Variance Inflation
Intercept	50.70	2.29	22.13	<0.0001	.	0.00
Heart_Rate	0.13	0.02	6.38	<0.0001	0.96	1.04
Ln_inc	0.23	0.07	3.21	0.001	0.84	1.19
Avg_BMI	0.57	0.06	9.79	<0.0001	0.89	1.12
Ownrad_r_d2	2.48	0.56	4.45	<0.0001	0.87	1.15
Educ_d1	3.24	1.24	2.61	0.01	0.58	1.73
Educ_d2	0.27	1.00	0.27	0.79	0.38	2.66
Educ_d3	-0.71	0.88	-0.80	0.42	0.34	2.91
Alcohol_d1	1.91	0.90	2.13	0.03	0.92	1.08
Alcohol_d2	4.03	0.97	4.15	<0.0001	0.95	1.05
Alcohol_d3	2.23	0.83	2.69	0.01	0.93	1.08
Alcohol_d4	3.19	0.89	3.59	0.0003	0.92	1.08
Exercise_d1	3.15	0.69	4.55	<0.0001	0.55	1.80
Exercise_d2	1.73	1.07	1.63	0.10	0.75	1.34
Exercise_d3	2.80	1.20	2.34	0.02	0.80	1.25
Exercise_d4	2.00	1.10	1.82	0.07	0.77	1.30

Based on the t-test with the significant level α equals 0.05, the p-values for the BMI,

resting heart rate, the ownership of a radio, alcohol consumption, education, exercise and income of an individual are less than 0.05 indicating that they each make a significant contribution to the predictive power of the model. The constant value (50.70 mmHg) represents the intercept, which is the predicted DBP for males when each independent variable is zero. The DBP increases by 0.13 mmHg for every one unit increase in the resting heart rate for males when all other independent variables remain constant. Similarly the DBP increases by 0.57 mmHg for every one unit increase in the BMI for males. In addition, as long as the percentage increase in income is fixed, one will observe the same difference in the DBP regardless of where the baseline income is. For example, for a 10% increase in income for males, the difference in the expected mean DBP will always be 0.23248*ln(1.1) = 0.0222 mmHg. Moreover, a male who owns a radio has on average 2.48 mmHg higher DBP than a male who does not own a radio when all other independent variables remain constant. After controlling for all other independent variables in the model, there is no statistically significant evidence that males with a secondary education or primary education have a different DBP to males with a tertiary education. However on average males with no schooling have 3.24 mmHg higher DBP than males with a tertiary education. Furthermore males who consume any alcoholic drinks irrespective of the number of drinks have a significantly higher DBP than males who do not consume alcoholic drinks. In addition, males who never exercise and who exercise once a week have a significantly higher DBP than males who exercise three or more times a week. However, there is no statistically significant evidence that males who exercise less than once a week or who exercise twice a week have a different DBP to males who exercise three or more times a week.

The following table highlights the stepwise regression results for the model including age and the use of antihypertensive medication:

Table 4.23: Validation dataset: Stepwise regression results for model 11

Step	**Group Entered**	**Group Removed**	**Number Vars In**	**Model R-Square**	**F Value**	**Pr > F**
1	Age		1	0.1916	503.03	<0.0001
2	Avg_BMI		2	0.2251	91.72	<0.0001
3	Employment_Status		5	0.2433	16.98	<0.0001
4	Heart_Rate		6	0.2498	18.27	<0.0001

Step	Group Entered	Group Removed	Number Vars In	Model R-Square	F Value	Pr > F
5	Hlbp_med_d1		7	0.2534	10.13	0.0015
6	Emotional_well-being		8	0.2557	6.49	0.0109
7	Smoker_Status		11	0.2585	2.73	0.0427

Table 4.23 highlights that age, BMI, employment status, heart rate, the use of antihypertensive medication, emotional well-being and smoker status are significant predictors of DBP for males. Each of these independent variables have a p-value less than 0.05 indicating sufficient evidence for predicting the DBP. When age and the use of antihypertensive medication are included in the model, one observes that the BMI and heart rate remain significant predictors however alcohol consumption, the ownership of a radio, exercise, education and income are no longer significant predictors of DBP for males. In addition, the employment status, emotional well-being and smoker status are now significant predictors of DBP for males.

Table 4.24: Validation dataset: Overall summary of results for model 11

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	95005	8636.78799	66.95	<0.0001
Error	2112	272466	129.0085		
Corrected Total	2123	367471			
Root MSE	11.35819	**R-Square**	0.2585		
Dependent Mean	78.92444	**Adj R-Sq**	0.2547		
Coeff Var	14.39122				

Table 4.24 highlights that the global F-test which has a p-value < 0.0001 implies that the model is significant in the prediction of DBP for males based on the group of independent variables in the model. The value of the R-square is 0.2585 which means that 25.85% of the variation in DBP is explained by the predictor variables. This implies that the model including age and the use of antihypertensive medication explains much more of the variation in DBP. The RMSE or standard deviation is 11.36 which implies that

approximately 95% of the sampled DBP values are within 2 standard deviations of their respective predicted values. Since the RMSE for the model including age and the use of antihypertensive medication is lower, this indicates that the model produces a better fit in the prediction of DBP for males.

Table 4.25: Validation dataset: Parameter estimates for model 11

Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Tol**	**Variance Inflation**
Intercept	61.14	2.39	25.64	<0.0001	.	0.00
Heart_Rate	0.08	0.02	4.03	<0.0001	0.97	1.03
Emotional_well-being	1.22	0.52	2.36	0.02	0.98	1.02
Avg_BMI	0.42	0.06	7.36	<0.0001	0.78	1.29
Smoker_Stat_d1	-3.29	1.18	-2.78	0.01	0.21	4.76
Smoker_Stat_d2	-2.97	1.28	-2.33	0.02	0.29	3.40
Smoker_Stat_d3	-3.45	1.36	-2.53	0.01	0.38	2.60
Empl_stat_d1	-4.21	0.61	-6.85	<0.0001	0.67	1.50
Empl_stat_d2	-5.56	2.01	-2.76	0.01	0.95	1.05
Empl_stat_d3	-2.12	0.99	-2.14	0.03	0.84	1.19
Age	0.21	0.02	11.41	<0.0001	0.59	1.69
Hlbp_med_d1	3.53	1.09	3.25	0.00	0.76	1.32

Based on the t-test with the significant level α equals 0.05, the p-values for age, BMI, employment status, heart rate, the use of antihypertensive medication, emotional well-being and the smoker status are less than 0.05 indicating that they each make a significant contribution to the predictive power of the model. The constant value (61.14 mmHg) represents the intercept, which is the predicted DBP for males when each independent variable is zero. The DBP increases by 0.42 mmHg for every one unit increase in the BMI for males when all other independent variables remain constant. Similarly the DBP increases by 0.08 mmHg for every one unit increase in the resting heart rate for males. Furthermore the DBP increases by 0.21 mmHg for every one unit increase in age for males. In addition, a male who is taking antihypertensive medication has on average 3.53 mmHg higher DBP than a male who is not taking antihypertensive medication when all other independent variables remain constant. Moreover the more occasions that a

male is not feeling emotional well, the higher the DBP. After controlling for all other independent variables in the model, males who consume less than 10 cigarettes per day or do not consume cigarettes have a significantly lower DBP than males who consume more than 10 cigarettes per day. Furthermore the males who are unemployed or not economically active have a significantly lower DBP than the males who are employed.

4.4.1.3 SBP multiple linear regression results for females

The following table highlights the stepwise regression results for the model excluding age and the use of antihypertensive medication:

Table 4.26: Test dataset: Stepwise regression results for model 15

Step	Group Entered	Group Removed	Number Vars In	Model R-Square	F Value	Pr > F
1	Education		3	0.1328	160.52	<0.0001
2	Avg_BMI		4	0.1881	214.46	<0.0001
3	Ownrad_r_d2		5	0.2101	87.43	<0.0001
4	Ln_inc		6	0.2199	39.71	<0.0001
5	Emotional_well-being		7	0.2241	16.81	<0.0001
6	Smoker_Status		10	0.229	6.67	0.0002
7	Employment_Status		13	0.2333	5.92	0.0005
8	Heart_Rate		14	0.2346	5.16	0.0232

Table 4.26 highlights that education, BMI, the ownership of a radio, income, emotional well-being, smoker status, employment status and heart rate are significant predictors of SBP for females. Each of these independent variables have a p-value less than 0.05 indicating sufficient evidence for predicting the SBP.

Table 4.27: Test dataset: Overall summary of results for model 15

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	344883	22992	64.33	<0.0001
Error	3134	1120093	357.40033		

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Corrected Total	3149	1464975			
Root MSE	18.90503	**R-Square**	0.2354		
Dependent Mean	119.9181	**Adj R-Sq**	0.2318		
Coeff Var	15.76496				

From Table 4.27 it is evident that the global F-test in the ANOVA table which has a p-value < 0.0001 depicts that the model is significant in the prediction of SBP for females based on the group of independent variables in the model. The value of the R-square is 0.2354 which means that 23.54% of the variation in SBP is explained by the independent variables. The RMSE or standard deviation is 18.90 which implies that approximately 95% of the sampled SBP values are within 2 standard deviations of their respective predicted values.

Table 4.28: Test dataset: Parameter estimates for model 15

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tol	Variance Inflation
Intercept	86.16	5.12	16.82	<0.0001	.	0.00
Heart_Rate	-0.06	0.03	-2.32	0.02	0.99	1.01
Emotional_well-being	2.70	0.72	3.75	0.0002	0.97	1.03
Ln_inc	0.81	0.12	6.56	<0.0001	0.66	1.52
Avg_BMI	0.69	0.06	12.13	<0.0001	0.89	1.13
Ownrad_r_d2	6.70	0.79	8.49	<0.0001	0.88	1.13
Smoker_Stat_d1	3.17	3.97	0.80	0.42	0.11	9.30
Smoker_Stat_d2	10.03	4.26	2.36	0.02	0.14	6.93
Smoker_Stat_d3	8.61	4.63	1.86	0.06	0.28	3.59
Educ_d1	21.40	1.63	13.09	<0.0001	0.47	2.13
Educ_d2	12.25	1.43	8.56	<0.0001	0.38	2.63
Educ_d3	2.83	1.24	2.27	0.02	0.32	3.17
Empl_stat_d1	3.11	0.98	3.16	0.002	0.47	2.11
Empl_stat_d2	2.37	2.23	1.06	0.29	0.87	1.15

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tol	Variance Inflation
Empl_stat_d3	-0.24	1.16	-0.21	0.84	0.63	1.60

Based on the t-test with the significant level α equals 0.05, the p-values for education, BMI, the ownership of a radio, income, emotional well-being, smoker status, employment status and heart rate are less than 0.05 indicating that they each make a significant contribution to the predictive power of the model. The constant value (86.16 mmHg) represents the intercept, which is the predicted SBP for females when each independent variable is zero. The SBP increases by 0.69 mmHg for every one unit increase in the BMI for females when all other independent variables remain constant. Furthermore the more occasions that a female is not feeling emotional well, the higher the SBP. In addition, as long as the percentage increase in income is fixed, one will observe the same difference in the SBP regardless of where the baseline income is. For example, for a 10% increase in income for females, the difference in the expected mean SBP will always be 0.81295*ln(1.1) = 0.0775 mmHg. Moreover, a female who owns a radio has on average 6.70 mmHg higher SBP than a female who does not own a radio when all other independent variables remain constant. Furthermore the SBP for females decreases by 0.06 mmHg for every one unit increase in the resting heart rate. After controlling for all other independent variables in the model, there is no statistically significant evidence that females who are non-smokers and who smoke between 6 and 10 cigarettes have a significantly different SBP to females with who smoke more than 10 cigarettes per day. However on average females who smoke less than 5 cigarettes per day have a significantly higher SBP than females who smoke more than 10 cigarettes per day. Furthermore females who have no schooling, primary education or secondary education have a significantly higher SBP than females who have a tertiary education. There is no statistically significant evidence that females who are unemployed strict or unemployed discouraged have a different SBP to females who are employed. However on average females who are not economically active have a significantly higher SBP than females who are employed.

The following table highlights the stepwise regression results for the model including age and the use of antihypertensive medication:

Table 4.29: Test dataset: Stepwise regression results for model 18

Step	Group Entered	Group Removed	Number Vars In	Model R-Square	F Value	Pr > F
1	Age		1	0.3279	1535.93	<0.0001
2	Hlbp_med_d1		2	0.3475	94.26	<0.0001
3	Avg_BMI		3	0.3509	16.82	<0.0001
4	Education		6	0.3571	10.04	<0.0001
5	Smoker_Status		9	0.3598	4.4	0.0043
6	Emotional_well-being		10	0.361	6.01	0.0142
7	Ownrad_r_d2		11	0.3623	6.1	0.0136
8	Ownhif_r_d2		12	0.3634	5.57	0.0183

Table 4.29 highlights that age, the use of antihypertensive medication, BMI, education, smoker status, emotional well-being, the ownership of a radio and the ownership of a hifi are significant predictors of SBP for females. Each of these independent variables have a p-value less than 0.05 indicating sufficient evidence for predicting the SBP. When the use of antihypertensive medication and age are included in the model, one observes that the BMI, education, smoker status, emotional well-being and the ownership of a radio remain significant predictors however the employment status, resting heart rate and income are no longer significant predictors of SBP for females. In addition, the ownership of a hifi is now a significant predictor of SBP for females.

Table 4.30: Test dataset: Overall summary of results for model 18

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	534460	35631	120	<0.0001
Error	3134	930516	296.90992		
Corrected Total	3149	1464975			
Root MSE	17.23107	**R-Square**	0.3648		
Dependent Mean	119.9181	**Adj R-Sq**	0.3618		
Coeff Var	14.36904				

Table 4.30 highlights that the global F-test which has a p-value < 0.0001 depicts that the model is significant in the prediction of SBP for females based on the group of independent variables in the model. The value of the R-square is 0.3648 which means that 36.48% of the variation in SBP is explained by the predictor variables. This implies that the model including age and the use of antihypertensive medication explains much more of the variation in SBP for females. The RMSE or standard deviation is 17.23 which implies that approximately 95% of the sampled SBP values are within 2 standard deviations of their respective predicted values. Since the RMSE for the model including age and the use of antihypertensive medication is lower, this indicates that the model produces a better fit in the prediction of SBP for females.

Table 4.31: Test dataset: Parameter estimates for model 18

Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Tol**	**Variance Inflation**
Intercept	78.97	4.18	18.91	<0.0001	.	0.00
Emotional_well-being	1.68	0.66	2.56	0.01	0.97	1.03
Avg_BMI	0.29	0.05	5.34	<0.0001	0.82	1.22
Ownhif_r_d2	-2.40	1.00	-2.41	0.02	0.75	1.33
Ownrad_r_d2	2.67	0.81	3.29	0.001	0.69	1.45
Smoker_Stat_d1	5.30	3.62	1.46	0.14	0.11	9.30
Smoker_Stat_d2	10.05	3.88	2.59	0.01	0.14	6.93
Smoker_Stat_d3	8.88	4.22	2.10	0.04	0.28	3.59
Educ_d1	6.20	1.61	3.84	0.0001	0.40	2.50
Educ_d2	5.78	1.32	4.37	<0.0001	0.37	2.71
Educ_d3	3.63	1.13	3.20	0.001	0.32	3.17
Age	0.57	0.03	20.65	<0.0001	0.45	2.20
Hlbp_med_d1	9.44	1.10	8.55	<0.0001	0.77	1.30

Based on the t-test with the significant level α equals 0.05, the p-values for the age, the use of antihypertensive medication, BMI, education, smoker status, emotional well-being, the ownership of a radio and the ownership of a hifi are less than 0.05 indicating that they each make a significant contribution to the predictive power of the model. The constant value (78.97 mmHg) represents the intercept, which is the predicted SBP for

females when each independent variable is zero. The SBP increases by 0.29 mmHg for every one unit increase in the BMI for females when all other independent variables remain constant. Similarly the SBP increases by 0.57 mmHg for every one unit increase in age for females. In addition, a female who is taking antihypertensive medication has on average 9.44 mmHg higher SBP than a female who is not taking antihypertensive medication when all other independent variables remain constant. Moreover the more occasions that a female is not feeling emotional well, the higher the SBP. A female who owns a radio has on average 2.67 mmHg higher SBP than a female who does not own a radio when all other independent variables remain constant. However a female who owns a hifi has on average 2.40 mmHg lower SBP than a female who does not own a hifi. After controlling for all other independent variables in the model, there is no statistically significant evidence that females who are non-smokers have a different SBP to females with who smoke more than 10 cigarettes per day. However on average females who smoke between 1 and 10 cigarettes per day have a significantly higher SBP than females who smoke more than 10 cigarettes per day. Furthermore females who have no schooling, primary education or secondary education have a significantly higher SBP than females who have a tertiary education.

4.4.1.4 DBP multiple linear regression results for females

The following table highlights the stepwise regression results for the model excluding age and the use of antihypertensive medication:

Table 4.32: Test dataset: Stepwise regression results for model 21

Step	**Group Entered**	**Group Removed**	**Number Vars In**	**Model R-Square**	**F Value**	**Pr > F**
1	Avg_BMI		1	0.0763	260.2	<0.0001
2	Education		4	0.1409	78.84	<0.0001
3	Ln_inc		5	0.1578	62.96	<0.0001
4	Ownrad_r_d2		6	0.1687	41.24	<0.0001
5	Smoker_Status		9	0.1778	11.54	<0.0001
6	Emotional_well-being		10	0.1803	9.64	0.0019

Table 4.32 highlights that BMI, education, income, the ownership of a radio, smoker status and emotional well-being are significant predictors of DBP for females. Each of these independent variables have a p-value less than 0.05 indicating sufficient evidence for predicting the DBP.

Table 4.33: Test dataset: Overall summary of results for model 21

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	11	99992	9090.2143	63.11	<0.0001
Error	3138	451990	144.03749		
Corrected Total	3149	551982			
Root MSE	12.00156	**R-Square**	0.1812		
Dependent Mean	79.9573	**Adj R-Sq**	0.1783		
Coeff Var	15.00996				

From Table 4.33 it is evident that the global F-test in the ANOVA table which has a p-value < 0.0001 depicts that the model is significant in the prediction of DBP for females based on the group of independent variables in the model. The value of the R-square is 0.1812 which means that 18.12% of the variation in DBP is explained by the independent variables. The RMSE or standard deviation is 12.00 which implies that approximately 95% of the sampled DBP values are within 2 standard deviations of their respective predicted values.

Table 4.34: Test dataset: Parameter estimates for model 21

Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Tol**	**Variance Inflation**
Intercept	59.05	2.90	20.40	<0.0001	.	0.00
Emotional_well-being	1.40	0.46	3.06	0.002	0.98	1.02
Ln_inc	0.43	0.07	6.22	<0.0001	0.85	1.17
Avg_BMI	0.47	0.04	13.17	<0.0001	0.89	1.12
Ownrad_r_d2	3.39	0.50	6.78	<0.0001	0.89	1.13
Smoker_Stat_d1	0.61	2.51	0.24	0.81	0.11	9.24

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tol	Variance Inflation
Smoker_Stat_d2	6.02	2.70	2.23	0.03	0.14	6.91
Smoker_Stat_d3	4.45	2.94	1.51	0.13	0.28	3.59
Educ_d1	9.23	1.00	9.23	<0.0001	0.50	1.98
Educ_d2	5.97	0.89	6.73	<0.0001	0.40	2.51
Educ_d3	1.23	0.78	1.58	0.11	0.33	3.07

Based on the t-test with the significant level α equals 0.05, the p-values for the BMI, education, income, the ownership of a radio, smoker status and emotional well-being are less than 0.05 indicating that they each make a significant contribution to the predictive power of the model. The constant value (59.05 mmHg) represents the intercept, which is the predicted DBP for females when each independent variable is zero. The DBP increases by 0.47 mmHg for every one unit increase in the BMI for females when all other independent variables remain constant. Furthermore the more occasions that a female is not feeling emotional well, the higher the DBP. In addition, as long as the percentage increase in income is fixed, one will observe the same difference in the DBP regardless of where the baseline income is. For example, for a 10% increase in income for females, the difference in the expected mean DBP will always be 0.4298*ln(1.1) = 0.0410 mmHg. Moreover, a female who owns a radio has on average 3.39 mmHg higher DBP than a female who does not own a radio when all other independent variables remain constant. After controlling for all other independent variables in the model, there is no statistically significant evidence that females who are non-smokers and who consume between 6 and 10 cigarettes per day have a different DBP to females with who smoke more than 10 cigarettes per day. However on average females who smoke between 1 and 5 cigarettes per day have a significantly higher DBP than females who smoke more than 10 cigarettes per day. Furthermore there is no statistically significant evidence that females who have a secondary education have a different DBP to females who have a tertiary education. However on average females who have no schooling or who have primary education have a significantly higher DBP than females who have a tertiary education.

The following table highlights the stepwise regression results for the model including age and the use of antihypertensive medication:

Table 4.35: Test dataset: Stepwise regression results for model 24

Step	Group Entered	Group Removed	Number Vars In	Model R-Square	F Value	Pr > F
1	Age		1	0.2038	805.7	<0.0001
2	Avg_BMI		2	0.2226	76.24	<0.0001
3	Hlbp_med_d1		3	0.2315	36.47	<0.0001
4	Smoker_Status		6	0.2383	9.3	<0.0001
5	Education		9	0.241	3.76	0.0103
6	Ownrad_r_d2		10	0.2431	8.47	0.0036
7	Emotional_well-being		11	0.2441	4.2	0.0405
8	Ownhif_r_d2		12	0.2451	4.15	0.0416
9	Employment_Status		15	0.247	2.65	0.0471

Table 4.35 highlights that age, BMI, the use of antihypertensive medication, smoker status, education, the ownership of a radio, emotional well-being, the ownership of a hifi and employment status are significant predictors of DBP for females. Each of these independent variables have a p-value less than 0.05 indicating sufficient evidence for predicting the DBP. When the use of antihypertensive medication and age are included in the model, one observes that the BMI, education, smoker status, emotional well-being and the ownership of a radio remain significant predictors however income is no longer a significant predictor of DBP for females. In addition, the ownership of a hifi and employment status are now significant predictors of DBP for females.

Table 4.36: Test dataset: Overall summary of results for model 24

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	137894	6894.6929	52.1	<0.0001
Error	3129	414088	132.33881		
Corrected Total	3149	551982			
Root MSE	11.50386	**R-Square**	0.2498		
Dependent Mean	79.9573	**Adj R-Sq**	0.245		
Coeff Var	14.38751				

Table 4.36 highlights that the global F-test which has a p-value < 0.0001 implies that the model is significant in the prediction of DBP for females based on the group of independent variables in the model. The value of the R-square is 0.2498 which means that 24.98% of the variation in DBP is explained by the predictor variables. This implies that the model including age and the use of antihypertensive medication explains much more of the variation in DBP for females. The RMSE or standard deviation is 11.50 which implies that approximately 95% of the sampled DBP values are within 2 standard deviations of their respective predicted values. Since the RMSE for the model including age and the use of antihypertensive medication is lower, this indicates that the model produces a better fit in the prediction of DBP for females.

Table 4.37: Test dataset: Parameter estimates for model 24

Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Tol**	**Variance Inflation**
Intercept	55.14	3.12	17.67	<0.0001	.	0.00
Emotional_well-being	0.98	0.44	2.24	0.03	0.97	1.03
Avg_BMI	0.30	0.04	8.45	<0.0001	0.81	1.23
Ownhif_r_d2	-1.55	0.67	-2.32	0.02	0.75	1.34
Ownrad_r_d2	1.79	0.54	3.30	0.00	0.69	1.45
Smoker_Stat_d1	2.30	2.43	0.95	0.34	0.11	9.41
Smoker_Stat_d2	6.17	2.60	2.38	0.02	0.14	6.96
Smoker_Stat_d3	4.58	2.82	1.62	0.10	0.28	3.60
Educ_d1	2.53	1.08	2.35	0.02	0.40	2.51
Educ_d2	3.01	0.89	3.39	0.00	0.37	2.72
Educ_d3	1.50	0.76	1.97	0.05	0.31	3.18
Empl_stat_d1	-1.32	0.53	-2.48	0.01	0.59	1.68
Empl_stat_d2	-0.08	1.32	-0.06	0.95	0.92	1.09
Empl_stat_d3	-1.44	0.67	-2.17	0.03	0.70	1.42
Age	0.25	0.02	13.60	<0.0001	0.45	2.21
Hlbp_med_d1	4.30	0.74	5.83	<0.0001	0.77	1.30

Based on the t-test with the significant level α equals 0.05, the p-values for age, BMI, the use of antihypertensive medication, smoker status, education, the ownership of a

radio, emotional well-being, the ownership of a hifi and employment status are less than 0.05 indicating that they each make a significant contribution to the predictive power of the model. The constant value (55.14 mmHg) represents the intercept, which is the predicted DBP for females when each independent variable is zero. The SBP increases by 0.30 mmHg for every one unit increase in the BMI for females when all other independent variables remain constant. Similarly the SBP increases by 0.25 mmHg for every one unit increase in age for females. In addition, a female who is taking antihypertensive medication has on average 4.30 mmHg higher DBP than a female who is not taking antihypertensive medication when all other independent variables remain constant. Moreover the more occasions that a female is not feeling emotional well, the higher the DBP. A female who owns a radio has on average 1.79 mmHg higher DBP than a female who does not own a radio when all other independent variables remain constant. However a female who owns a hifi has on average 1.55 mmHg lower DBP than a female who does not own a hifi. After controlling for all other independent variables in the model, there is no statistically significant evidence that females who are non-smokers and who consume between 6 and 10 cigarettes per day have a different DBP to females who smoke more than 10 cigarettes per day. However on average females who smoke between 1 and 5 cigarettes per day have a significantly higher DBP than females who smoke more than 10 cigarettes per day. Furthermore females who have no schooling, primary education or secondary education have a significantly higher DBP than females who have a tertiary education. There is no statistically significant evidence that females who are unemployed discouraged have a different DBP to females who are employed. However on average females who are unemployed strict or not economically active have a significantly lower DBP than females who are employed.

4.5 Comparison of SEM and multiple linear regression results

The structural equation models, included five latent variables which explain the concepts of SES, SBP, DBP, heart rate and emotional well-being. The SBP, DBP and heart rate components were each measured by their first and second respective reading. The SES component was measured by income, education, employment status, the ownership of

a radio, the ownership of a cellphone and the ownership of a hifi (Kaplan and Keil, 1993). The emotional well-being component was measured by the number of occasions an individual felt bothered, had trouble focusing, felt depressed, felt fearful, couldn't sleep, felt lonely and could not get going. Each of the indicators loaded highly onto their respective latent variable.

The SEM analysis investigated two of the research questions proposed for both males and females. The first set of models estimated the direct relationship between SES and SBP and SES and DBP. These models are presented in Figure 4.4 and Figure 4.5. These models depicted that there is a significant positive association between SES and SBP and SES and DBP for both males and females. These results revealed that SES explains 6.3% of the variance in SBP and 7.1% of the variance in DBP for males while SES explains only 1.4% of the variance in SBP and 2.3% of the variance in DBP for females. Furthermore the model fit indices revealed that the model fit was adequate for both males and females. In addition all of the indicators' loading onto SES were statistically significant for both males and females. These results demonstrated that an increasing SES was associated with increased blood pressure (Addo, Smeeth and Leon, 2009; Ploubidis *et al.*, 2013; Fikadu and Lemma, 2016).

The second set of models investigated whether certain bio-behavioural risk factors mediate the relationship between SES and blood pressure. These models are presented in Figure 4.6, Figure 4.7, Figure 4.8 and Figure 4.9. These models demonstrated that all of the indicators in the model were explained by their corresponding factors significantly for both males and females. The structural and measurement regression models fitted the data well when the controls such as the use of antihypertensive medication and age were included in the analysis and in the cases where the controls were not taken into account.

When the controls were not included in the analysis, the direct effects from SES to DBP and SES to SBP were statistically significant for both males and females indicating that the behavioural risk factors partially mediate the relationship between SES and blood pressure for males and females (Preacher and Hayes, 2004). Partial mediation occurs since the effect of SES on SBP or SES on DBP for both males and females decreased by a nontrivial amount, but not to zero, when the mediators were included in the analysis (Preacher and Hayes, 2004). Furthermore, it was evident that the total indirect effect of SES on SBP and SES on DBP for males and females was significantly different from zero

at $p < 0.05$ (two tailed). Thus, we rejected the null hypothesis that the total indirect effect is zero (Hayes, 2009). This implies that the bio-behavioural risk factors mediated the effect of SES on SBP and SES on DBP for males and females.

The specific indirect effects were then derived in SPSS AMOS by using the user-defined estimand function developed by Gaskin (2016). After examining the specific indirect effects for males in the model proposed in Figure 4.6, it was evident that BMI, alcohol consumption, smoker status and emotional well-being were significant mediators in the relationship between SES and SBP and SES and DBP for males. Furthermore, for females it is clear that BMI, heart rate, smoker status and exercise were the only mediators in the relationship between SES and SBP. However, in the model proposed in Figure 4.7, it was clear that BMI, heart rate, alcohol consumption, smoker status and exercise were mediators in the relationship between SES and DBP. In addition, after examining the size of the total standardised effects, it was clear that BMI and SES had the biggest statistically significant contribution to the prediction of SBP and DBP through all the presumed pathways for males and females (Byrne, 2010).

When the controls were included in the analysis, the direct effects from SES to DBP and from SES to SBP were not significantly different from zero for males thus the bio-behavioural risk factors completely mediate the relationship between SES and blood pressure (James and Brett, 1984). Since the direct effect from SES to SBP for females was not significantly different from zero after controlling for the use of antihypertensive medication and age, the bio-behavioural risk factors fully mediate the relationship between SES and SBP (James and Brett, 1984). However the direct effect from SES to DBP was significant and decreased by a nontrivial amount, but not to zero, thus the bio-behavioural risk factors partially mediate the relationship between SES and DBP for females (Preacher and Hayes, 2004). Furthermore, it was evident that the total indirect effect of SES on SBP and SES on DBP for males and females was significantly different from zero at $p < 0.05$ (two tailed) when controlling for the use of antihypertensive medication and age. Thus, we rejected the null hypothesis that the total indirect effect is zero (Hayes, 2009). This implies that the bio-behavioural risk factors mediated the effect of SES on SBP and SES on DBP for males and females when taking confounding factors such as the use of antihypertensive medication and age into account.

The specific indirect effects were of interest in the study in order to determine which of the behavioural risk factors mediate the relationship between SES and SBP and SES and DBP. After examining the specific indirect effects for males, it was evident that BMI, alcohol consumption and exercise were the only mediators in the relationship between SES and SBP when controlling for the use of antihypertensive medication and age. Furthermore, it was clear that BMI, alcohol consumption, heart rate and emotional well-being were the only mediators in the relationship between SES and DBP when controlling for the use of antihypertensive medication and age. In addition, for females it was evident that heart rate, BMI and exercise were the only mediators in the relationship between SES and SBP when controlling for the use of antihypertensive medication and age. Similarly for hypothesised model proposed in Figure 4.9, it was clear that BMI, alcohol consumption and heart rate were the only mediators in the relationship between SES and DBP for females when taking confounding factors such as the use of antihypertensive medication and age into account.

The size of the total standardised effects were then examined. This demonstrated that after controlling for the use of antihypertensive medication and age, BMI had the biggest statistically significant contribution to the prediction of SBP through all the presumed pathways for males (Byrne, 2010). In addition, BMI, SES and heart rate had the largest statistically significant contribution to the prediction of DBP through all the presumed pathways for males when controlling for the use of antihypertensive medication and age (Byrne, 2010). Furthermore, the total standardised effects for females demonstrated that BMI had the biggest statistically significant contribution to the prediction of SBP through all the presumed pathways after controlling for the use of antihypertensive medication and age. However BMI and SES had the largest statistically significant contribution to the prediction of DBP through all the presumed pathways for females after adjusting for the use of antihypertensive medication and age (Byrne, 2010).

Multiple linear regression was then used to estimate the relationship between demographic, behavioural, socio-economic variables and blood pressure. The data was first subdivided into a training set, validation set and test set for both males and females. Simple random sampling was applied to data in order to subdivide it into the training set, validation set and test set (Reitermanova, 2010). The training set comprised of 40% of the data while the test and validation sets each comprised of 30% of the data (Hair *et al.*, 1998; Harrell, 2015). This was done so that the results from the training, validation and testing

datasets could be compared in terms of the MSE in order to choose the best fitting model.

Separate linear regression models were analysed for each gender using SBP as a dependent variable and subsequently using DBP as the dependent variable (Wong *et al.*, 2015). Age and the use of antihypertensive medication are potentially confounding variables which may have an effect on the dependent variables (Ploubidis *et al.*, 2013; Miyaki *et al.*, 2013; Bardenheier *et al.*, 2013; Cois and Ehrlich, 2014; Keetile *et al.*, 2015). In regression analysis, confounding variables are treated as additional independent variables in the analysis (McNamee, 2005; Pourhoseingholi *et al.*, 2012). The comparison of the models including confounding variables and excluding confounding variables can give an indication of how much the confounders in the model distort the relationship between the independent and dependent variables (McNamee, 2005; Pourhoseingholi *et al.*, 2012). As a result, some of the models included the use of antihypertensive medication and age, which were considered to be controls, while others excluded them in order to assess the difference in the fit of these models. Since there were no variables that were moderators in the analysis, each regression model only included the main effects.

Before accepting a regression model, it was important to examine the influence and fit diagnostics to determine whether the model might be unduly influenced by a few observations and whether the data supports the assumptions that underlie the linear regression. The graphs describing the relationship between the residuals and the predicted value illustrated that the residuals for each model were distributed around zero which implies that the assumption of homoscedasticity was not violated for any of the linear regression models that were analysed (Tabachnick and Fidell, 2007). Furthermore the residuals did not appear to exhibit any particular pattern which implied independence of the residuals. The plot of the externally studentized residuals by leverage values revealed that there were a few observations with a high leverage however since less than 5% of the residuals exceeded the threshold value, the linear models were deemed appropriate (Stevens, 2002). The plot of Cook's distance (Cook, 1977) by observation was very useful for identifying influential points. Since none of the models had a Cook's distance greater than one, this implied that there were no highly influential points in any of the regression models analysed (Cook and Weisberg, 1982). The normal-probability Q-Q plot showed that there was a linear trend with a slight deviation at the tails for each model, which suggested that the assumption of normality was adhered to for each of the regression

models that were analysed (Hair *et al.*, 1998). The histogram of the residuals depicted that the normality assumption was only slightly violated thus the interpretation of the regression results were appropriate for each of the models that were analysed (Hair *et al.*, 1998).

When the use of antihypertensive medication and age were not included in the regression model, BMI, education, the ownership of a radio, income, alcohol consumption and emotional well-being were significant predictors of SBP for males. Furthermore 13.33% of the variation in SBP was explained by these independent variables. These results were fairly similar to the hypothesised SEM model for males excluding the controls which explained 13.1% of the variation in SBP. In the SEM analysis, SES, BMI and smoker status had a statistically significant total effect on SBP for males through all of their presumed pathways. However the SEM model allowed one to investigate not only the total effects on SBP, but the indirect and direct effects as well (Hair *et al.*, 1998; Preacher and Hayes, 2004, 2008; Hayes, 2009). This was more useful for the purposes of our analysis as we were more interested in investigating whether bio-behavioural risk factors mediate the relationship between SES and blood pressure (Preacher and Hayes, 2004, 2008; Hayes, 2009). Furthermore, the SEM model included SES as a latent variable whereas in the regression model, each of the indicators of SES were entered as independent predictors in the analysis.

Furthermore, when the use of antihypertensive medication and age were not included in the regression model, BMI, resting heart rate, the ownership of a radio, alcohol consumption, education, exercise and income were significant predictors of DBP for males. These independent variables explained 14.86% of the variation in DBP. In the SEM analysis, smoker status, heart rate, alcohol consumption, exercise, SES and BMI had a statistically significant total effect on DBP for males through all of their presumed pathways. The hypothesised SEM model for males explained 14.7% of the variation in DBP which is very similar to the variation explained by the regression model.

The regression analysis revealed that education, BMI, the ownership of a radio, income, emotional well-being, smoker status, employment status and heart rate were significant predictors of SBP for females when the use of antihypertensive medication and age were not included in the model. Furthermore 23.54% of the variation in SBP was explained by these independent variables. The SEM analysis revealed that smoker status, exercise,

heart rate, SES and BMI all had a statistically significant total effect on SBP for females through all of their presumed pathways. The hypothesised SEM model for females explained 10.9% of the variation in SBP which is much lower than the variation explained by the regression model.

Furthermore, when the use of antihypertensive medication and age were not included in the regression model, BMI, education, income, the ownership of a radio, smoker status and emotional well-being were significant predictors of DBP for females. These independent variables explained 18.12% of the variation in DBP. The SEM analysis demonstrated that smoker status, heart rate, exercise, SES and BMI all had a statistically significant total effect on DBP for females through all of their presumed pathways. The hypothesised SEM model for females explained 11.6% of the variation in DBP which is lower than the variation explained by the regression model.

When the use of antihypertensive medication and age were included in the regression model, age, BMI, the use of antihypertensive medication and alcohol consumption were significant predictors of SBP for males. Furthermore 22.55% of the variation in SBP was explained by these independent variables. When taking the use of antihypertensive medication and age into account, the SEM analysis revealed that SES, exercise, alcohol consumption, BMI, age and the use of antihypertensive medication all had a statistically significant total effect on SBP for males through all of their presumed pathways. The hypothesised SEM model for males explained 21% of the variation in SBP which is similar to the variation explained by the regression model.

Furthermore age, BMI, employment status, heart rate, the use of antihypertensive medication, emotional well-being and smoker status were significant predictors of DBP for males when including the use of antihypertensive medication and age in the regression model. These independent variables explained 25.85% of the variation in DBP. When taking the use of antihypertensive medication and age into account, the SEM analysis demonstrated that SES, heart rate, BMI, the use of antihypertensive medication and age all had a statistically significant total effect on DBP for males through all of their presumed pathways. The hypothesised SEM model for males explained 20.8% of the variation in DBP which is slightly lower than the variation explained by the regression model.

When age and the use of antihypertensive medication were included in the regression model, age, the use of antihypertensive medication, BMI, education, smoker status, emotional well-being, the ownership of a radio and the ownership of a hifi were significant predictors of SBP for females. These independent variables explained 36.48% of the variation in SBP. When taking the use of antihypertensive medication and age into account, the SEM analysis revealed that SES, heart rate, BMI, the use of antihypertensive medication and age all had a statistically significant total effect on SBP for females through all of their presumed pathways. The hypothesised SEM model for females explained 32.5% of the variation in SBP which is slightly lower than the variation explained by the regression model.

Furthermore, the regression analysis demonstrated that age, BMI, the use of antihypertensive medication, smoker status, education, the ownership of a radio, emotional well-being, the ownership of a hifi and employment status were significant predictors of DBP for females. These independent variables explained 24.98% of the variation in DBP. When taking the use of antihypertensive medication and age into account, the SEM analysis demonstrated that heart rate, SES, BMI, the use of antihypertensive medication and age all had a statistically significant total effect on DBP for females through all of their presumed pathways. The hypothesised SEM model for females explained 24.9% of the variation in DBP which is very similar to the variation explained by the regression model. The amount of variation explained in SBP and DBP was significantly better when including the controls in the analysis thus emphasizing the importance of adjusting for the use of antihypertensive medication and age. In most cases, the amount of variation in DBP and SBP that were explained was very similar when either SEM or regression analysis was performed.

The SEM solution however gives a better understanding of the variables and the factors that would help in explaining the SBP and DBP. These models include SES, a latent variable, which is absent in the regression model. In addition, SEM allows for the ease of interpretation and estimation when including latent variables (Musil, Jones and Warner, 1998; Gunzler, Chen, Wu and Zhang, 2013). The structural equation model provides strong evidence that income, education, employment status, the ownership a radio, the ownership of a hifi and the ownership of a cellphone measure the same concept. This is the benefit of SEM, it incorporates these factors into a latent variable, whereas the regression model includes the above indicators as predictor variables and

does not allow the inclusion of multiple factors (Hair *et al.*, 1998; Musil *et al.*, 1998). Even though summated factors representing theoretical constructs could be entered as variables in regression analysis, regression models treat these variables and constructs identically (Hair *et al.*, 1998). Thus multiple regression does not take into account any of the measurement properties that go along with forming a multi-item construct when estimating the relationship (Hair *et al.*, 1998). However, according to Preacher and Hayes (2004), SEM models involving latent variables with multiple measure indicators correct for the measurement error by estimating the unique and common variance separately.

Furthermore, SEM has the advantage over multiple linear regression since one can examine the theoretical model by including both the structural and measurement model in one analysis (Hair *et al.*, 1998). As a result, the interrelationship of all of the variables and the factors is taken into account in the SEM models, thus giving us a better understanding of the predictors of SBP and DBP. Moreover, one is able to understand the indirect, direct and total effects of each exogenous variable on the endogenous variables (Hair *et al.*, 1998; Preacher and Hayes, 2004, 2008; Hayes, 2009; Gunzler *et al.*, 2013). In addition, the SEM solution enables us to understand the effect of mediators in the relationship between SES and SBP and SES and DBP which is not possible in regression which is an additive model (Preacher and Hayes, 2004, 2008). As a result, mediation models are best estimated in a SEM context because of the greater flexibility SEM programs afford in model specification and estimation options (Preacher and Hayes, 2008).

One of the primary advantages of SEM is that it allows one to model many dependent (latent or observed) variables in the same research model and analyse all paths simultaneously rather than one at a time (Musil *et al.*, 1998; Alavifar *et al.*, 2012; Gunzler *et al.*, 2013). In regression analysis, separate models have to be built to analyse the association between independent variables and each dependent variable as was evident in this study (Hair *et al.*, 1998). Furthermore in SEM, a variable can be both predicted and explanatory for example, BMI in our study. This means a variable can appear on the left side of one equation and on the right side(s) of one or more equation(s) in the same model which implies that multiple interacting equations can be modelled simultaneously (Hair *et al.*, 1998). However in regression, a variable can only either be a response or an explanatory variable, thus if the study requires a variable to play both roles, it has to be in separate models and cannot be estimated at the same time (Hair *et al.*, 1998). An added benefit of using SEM for this study, is that the accepted model can be tested

over time to see if the same variables and factors are still important in explaining SBP or DBP or whether a new model needs to be tested. Furthermore, the simultaneous nature of the indirect and direct effects, and the dual role of the mediator as both a cause for the outcome variable and an effect of the intervention are more appropriately expressed using structural equations than using regression analysis (Gunzler *et al.*, 2013). Based on these advantages of using SEM, SEM proves to be more useful in analysing the mediating relationships between SES and blood pressure.

4.6 Summary

This chapter provided an overview of the data analysis and the results. The data analysis consisted of three parts, namely SEM, multiple linear regression and the evaluation and comparison of the results. Firstly the sample was described as the median and interquartile range for continuous variables and frequency or percentage for categorical variables. Before performing any analysis, the data were screened for missing data, outliers, normality, homoscedasticity, linearity and multicollinearity. The Cronbach alpha value was then analysed in order to assess the internal consistency of the items measured on a Likert scale to determine whether the scale was reliable. Exploratory factor analysis was then performed with varimax rotation on all variables that load onto latent variables. This was done to ensure that the variables loading onto the latent variables only load onto one factor.

SEM was then used to test the hypothesised conceptual framework. In this study, the relationships among several factors influencing SBP and DBP were analysed through two steps. CFA was first performed to assess the validity of the latent factors while the impact of the predictors influencing SBP and DBP was modelled in the second step. In the second step of the analysis, two different models were built for males and females to assess the research questions proposed in this study. The first model estimated the direct relationship between SES and SBP and SES and DBP while the second model investigated whether certain bio-behavioural risk factors mediate the relationship between SES and blood pressure. Since the data revealed evidence of multivariate non-normality, ADF estimation was used to build all the models, to perform model modification and to evaluate the model. After adjusting for the use of antihypertensive medication and age,

a higher SES was associated independently with a higher systolic and diastolic blood pressure in both males and females. Furthermore, it was evident that body mass index was a mediator of the indirect effect of socio-economic status on blood pressure for both genders. Smoker status, alcohol consumption, physical exercise, emotional well-being and resting heart rate were also mediators in the relationship between SES and blood pressure, however their role was modest in comparison to BMI.

Multiple linear regression was then used to examine the relationships between demographic, behavioural, socio-economic variables and blood pressure. The data was first subdivided into a training set, validation set and test set for both males and females. The training set comprised of 40% of the data while the test and validation sets each comprised of 30% of the data (Hair *et al.*, 1998; Harrell, 2015). This was done so that the results from the training, validation and testing datasets could be compared in terms of the MSE in order to choose the best fitting model. The best fitting models were then compared against the results from SEM. The comparison of the multiple linear regression and SEM results revealed that SEM was a more useful technique for analysing the mediating relationships between SES and blood pressure. The main findings of this study and recommendations are discussed in the next chapter.

Chapter 5

Conclusion

The implications of the current study and a discussion of these results are presented in this chapter. In section 5.1, the results of the study are discussed with regard to existing literature. In section 5.2, strengths and limitations of these results are presented while in section 5.3 suggestions for future research opportunities are provided.

5.1 Discussion of results

The purpose of this study was to investigate the relationship between SES and blood pressure. Multiple linear regression was applied to investigate the relationships between demographic, bio-behavioural risk factors and socio-economic status. SEM was applied to examine the direct effects of SES on SBP and SES on DBP without mediation. In addition, this study provided the opportunity to investigate whether bio-behavioural risk factors mediate the relationship between SES and blood pressure using SEM. The models proposed in this study are very complex, thus regression analysis did not have the same effect as SEM does. These models have five latent variables consisting of four unobserved endogenous variables and one unobserved exogenous variable. Furthermore these models include two observed exogenous variables and 23 observed endogenous variables. The benefit of SEM comes from including these observed variables as well as the latent variables in the model and therefore being able to analyse all the relationships at the same time (Hair *et al.*, 1998; Musil *et al.*, 1998; Alavifar *et al.*, 2012; Gunzler *et al.*, 2013). When simple research questions that require single linear relationships to be tested are proposed, then regression is a suitable technique however more complex causal models require a more complex technique. As a result, models that incorporate latent variables are best analysed using SEM (Musil *et al.*, 1998; Gunzler *et al.*, 2013).

Thus the type of statistical technique used, depends solely on the research question at hand.

The first element of the SEM analysis involved the investigation of the association between SES and blood pressure without adjusting for the effect of confounding variables such as the use of antihypertensive medication and age. These hypothesised models depicted an adequate model fit for both males and females. The chi-square statistic did not show a good model fit for any of the models proposed due to the sensitivity to sample size therefore other experimentally recommended model fit indices were considered. SES, BMI and smoker status had a statistically significant total effect on SBP for males. SES, exercise, alcohol consumption, smoker status, BMI and heart rate each had a statistically significant contribution to the prediction of DBP for males through all of their presumed pathways. SES, exercise frequency, smoker status, BMI and heart rate had a statistically significant total effect on SBP and DBP for females via all presumed pathways. Overall the model for males explained 13.1% of the variation in SBP and explained 14.7% of the variation in DBP while the model for females explained 10.9% of the variation in SBP and explained 11.6% of the variation in DBP.

The second element of the SEM analysis involved the investigation of the relationship between SES and blood pressure after adjusting for the use of antihypertensive medication and age. The hypothesised models depicted an adequate model fit for both males and females. After taking age and medication into account, SES, exercise, alcohol consumption and BMI all had a statistically significant total effect on SBP for males. These results revealed that males with a higher SES, a higher BMI, that consume more alcohol or exercise more frequently have a higher SBP. These findings are comparable to those of Brummett *et al.* (2011) where a higher BMI, smoking, a higher alcohol consumption and a greater waist circumference were each individually associated with a higher SBP. The size of the total standardised effects were then examined to determine which of the variables has the largest contribution to the prediction of SBP and DBP. The total standardised effects revealed that after including the controls such as age and medication, BMI had the largest statistically significant contribution to the prediction of SBP through all the presumed pathways for males. Similarly, after controlling for age and medication, SES, BMI and heart rate had a statistically significant total effect on DBP for males. These results demonstrated that males with a higher SES, a higher BMI and a higher resting heart rate tend to have a higher DBP. However, BMI had the largest statistically

significant contribution to the prediction of DBP through all the presumed pathways for males when controlling for the use of antihypertensive medication and age. Among males, BMI, alcohol consumption and exercise were mediators of the increasing SES on SBP after controlling for the use of antihypertensive medication and age. Furthermore, BMI, alcohol consumption, heart rate and emotional well-being were mediators in the relationship between SES and DBP for males after the controls were taken into account. Moreover, the hypothesised model for males explained 21% of the variation in SBP and explained 20.8% of the variation in DBP.

After taking age and the use of antihypertensive medication into account, SES, BMI and heart rate all had a statistically significant total effect on SBP and DBP for females via all their presumed pathways. These results demonstrated that females with a higher SES, a higher BMI or a lower heart rate have a higher SBP while females with a higher SES, higher BMI or higher heart rate have a higher DBP. After the examination of the size of the total standardised effects, it was clear that BMI had the largest statistically significant contribution to the prediction of SBP through all the presumed pathways for females when controlling for the use of antihypertensive medication and age. However BMI and SES had the largest statistically significant contribution to the prediction of DBP through all the presumed pathways for females after controlling for the use of antihypertensive medication and age. Among females, BMI, heart rate and exercise were mediators of the increasing SES on SBP after controlling for the use of antihypertensive medication and age. Furthermore, BMI, alcohol consumption and heart rate were mediators in the relationship between SES and DBP for females after the controls were taken into account. Moreover, the hypothesised model for females explained 32.5% of the variation in SBP and explained 24.9% of the variation in DBP.

Since the model including the adjustment of age and the use of antihypertensive medication explained more of the variation in SBP and DBP, these results will be compared to literature. Relationships between SES and blood pressure which are consistent with our results (that is a positive association between SES and blood pressure for both genders) have previously been found in other countries in Africa (Addo *et al.*, 2009; Ploubidis *et al.*, 2013; Fikadu and Lemma, 2016). These findings are in contrast with those of high income developed countries where a lower SES was often associated with higher mean blood pressure in both genders (Colhoun *et al.*, 1998; Brummett *et al.*, 2011; Lam, 2011). This is because in developed countries individuals with high socioeconomic status may be

early adopters of healthy lifestyles that help to lower the risk of hypertension. In middle income countries however, sections of the population that undergo more rapid social development, may also increase their risk factors for cardiovascular disease and thus be at a greater risk of disease compared to those with a lower SES. This could be due to the harmful lifestyle of individuals with higher SES in middle income countries (Fikadu and Lemma, 2016). Thus, the observed difference in the pattern of health inequalities in South Africa may be due to the fact that participants with a higher SES are more likely to afford processed foods that are high in salt and fats which could lead to obesity and ultimately to the development of high blood pressure (Addo *et al.*, 2009; Ploubidis *et al.*, 2013).

The overall patterns of association of physical activity, smoker status, alcohol consumption, emotional well-being and heart rate, although different in magnitude and significance, are consistent with the findings of higher income countries (Chaix *et al.*, 2010; Brummett *et al.*, 2011). However in contrast to those studies, but in agreement with other studies in Africa (Sodjinou, Agueh, Fayomi and Delisle, 2008; Ploubidis *et al.*, 2013; Cois and Ehrlich, 2014), BMI increases with an increasing SES. Thus BMI mediates the effect of an increasing SES on blood pressure accounting for a sizeable portion of the association in both genders. These contrasting results to higher income countries may be partly due to the unequal distribution of the income in South Africa where a large portion of the population lives near or below the poverty line (Leibbrandt, Woolard, Finn and Argent, 2010). It is thus likely that among people in South Africa, the increased knowledge of health risk and greater motivation to control body weight associated with increasing SES, which is often given as an explanation of the inverse SES/BMI relationship in high income countries, plays less of a role than the greater access to processed foods high in fat and salt among those with a higher SES. Furthermore, in our study, we found that SES affects SBP and DBP differently across gender which may explain the differences in the results of studies that analyse hypertension prevalence or use SBP as the only outcome variable.

As hypertension is becoming a serious public health concern in South Africa, it has to be considered as a top priority in the health agenda of the country. This study has shown that SES has a positive association with increased blood pressure in both males and females. Furthermore SES influences the practice of risky bio-behavioural factors that are responsible for hypertension. One of the implications of this study is that a

reduction in the BMI of an individual through a balanced diet will have a substantial impact on lowering hypertension. Furthermore, the promotion of healthy behaviours and interventions that target higher income groups need to be established so that these groups can make rational decisions in choosing their behaviours. The long term consequences that arise from poor lifestyle choices have to be stressed to the population in general and to groups with higher SES in particular.

5.2 Strengths and limitations of the study

The strengths of our study include the use of a large population based sample to simultaneously test multiple mediation pathways which avoids the bias that results from neglecting the correlation between mediators. Furthermore, our study allows for the explicit consideration of the measurement error in the bio-behavioural risk factors and uses an appropriate modelling strategy to estimate both the indirect and direct effects of SES on blood pressure. In addition, this study is one of the few studies that allows for mediation analysis of a large sample for both DBP and SBP.

One of the major restrictions of this study is the cross sectional nature of the study which limits the temporal sequence of the relationships. It is highly possible that diseases that were previously diagnosed may impact on behavioural patterns and to some extent affect the parameter estimates. Furthermore, although we have adjusted our results for age and the use of antihypertensive medication, our results depend on the correct specification of the model and may be impacted by unmeasured confounding variables such as salt intake or dietary patterns which have been shown in literature to affect blood pressure (Law, Frost and Wald, 1991; Weinberger, 1996). Despite the fact that we have controlled for the use of antihypertensive medication and age, we cannot exclude the fact that these variables may instead act as modifiers or mediators in the model. A further limitation is that participants may withhold information regarding their lifestyle (smoking, alcohol consumption or physical exercise) habits that may not be generally acceptable which may result in an underestimation of these behaviours. Therefore more precise measurements of these lifestyle behaviours may enhance the results of our analysis rather than invalidate them. Finally the highly unequal distribution of SES in South

Africa, make it essential to exercise caution when generalizing these results to other areas of Africa or other developing countries.

5.3 Recommendations for further work

Future longitudinal studies are required to take into consideration the temporal sequence of the associations between SES and blood pressure in South Africa. Furthermore, studies that include detailed dietary information are needed in order to further explore the positive relationship between SES and blood pressure. In addition, future work could be done to replicate our analysis in other settings with rapid but complex epidemiologic transitions.

References

Addo, J., Smeeth, L. and Leon, D. (2009). Socioeconomic position and hypertension: a study of urban civil servants in Ghana. *Journal of Epidemiology and Community Health*, Vol. 63, No. 8, pp. 646–650.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716–723.

Alavifar, A., Karimimalayer, M. and Anuar, M. (2012). Structural equation modelling vs. multiple regression. *Engineering science and technology: An International Journal*, Vol. 2, No. 2, pp. 326–329.

Altman, D., Gore, S., Gardner, M. and Pocock, S. (1983). Statistical guidelines for contributors to medical journals. *British Medical Journal*, Vol. 286, pp. 1489–1493.

Appel, L. (2003). Lifestyle modification as a means to prevent and treat high blood pressure. *Journal of the American Society of Nephrology*, Vol. 14, No. 2, pp. 99–102.

Arbuckle, J. L. (2014). *Amos 23 User's Guide*. Chicago: IBM SPSS.

Bardenheier, B., Bullard, K., Caspersen, C., Cheng, Y., Gregg, E. and Geiss, L. (2013). A novel use of structural equation models to examine factors associated with prediabetes among adults aged 50 Years and older. *Diabetes Care Journal*, Vol. 36, No. 9, pp. 2655–2662.

Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, Vol. 51, pp. 1173–1182.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, Vol. 42, No. 5, pp. 815–824.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, Vol. 3, No. 2, pp. 77–85.

Batel, P., Pessione, F., Maître, C. and Rueff, B. (1995). Relationship between alcohol and tobacco dependencies among alcoholics who smoke. *Addiction*, Vol. 90, No. 7, pp. 977–980.

Batista-Foguet, J. M., Coenders, G. and Ferragud, M. (2001). Using structural equation models to evaluate the magnitude of measurement error in blood pressure. *Journal of Statistics in Medicine*, Vol. 20, No. 15, pp. 2351–2368.

Becker, T. E. (2005). Potential problems in the statistical control of variables in organisational research: A qualitative analysis with recommendations. *Organisational Research Methods*, Vol. 8, No. 3, pp. 274–289.

Bentler, P. (1990). Comparative fit indexes in Structural Models. *Psychological Bulletin*, Vol. 107, No. 2, pp. 238–246.

Bentler, P. (2005). *EQS 6 Structral equations program manual*. CA: Multivariate software, Encino.

Bentler, P. and Bonnet, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, Vol. 88, No. 3, pp. 588–606.

Bentler, P. and Chou, C. (1987). Practical issues in structural modelling. *Sociological Methods and Research*, Vol. 16, No. 1, pp. 78–117.

Blunch, N. (2008). *Introduction to structural equation modelling using SPSS and AMOS*. Sage Publications, London, 270 pages.

Bollen, K. A. and Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, Vol. 20, pp. 115–140.

Bollen, K. (1989). *Structural equations with latent variables*. John Wiley & Sons Inc., New York, 514 pages.

Browne, M. (1984). Asymptotically distribution-free methods for analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, Vol. 37, pp. 62–83.

Brummett, B., Babyak, M., Siegler, I., Shanahan, M., Harris, K., Elder, G. and Williams, R. (2011). Systolic blood pressure, socioeconomic status, and biobehavioural risk

factors in a nationally representative US young adult sample. *Hypertension*, Vol. 58, No. 2, pp. 161–166.

Byrne, B. (2010). *Structural equation modelling with AMOS: Basic concepts, applications and programming*. Routledge, Taylor and Francis Group, New York, 396 pages.

Chaix, B., Bean, K., Leal, C., Thomas, F., Havard, S., Evans, D., Jégo, B. and Pannier, B. (2010). Individual/Neighbourhood social factors and blood pressure in the RECORD Cohort Study: which risk factors explain the associations? *Hypertension*, Vol. 55, No. 3, pp. 769–775.

Cois, A. and Ehrlich, R. (2014). Analysing the socioeconomic determinants of hypertension in South Africa: a structural equation modelling approach. *BMC Public Health*, Vol. 14, No. 1, pp. 414–424.

Colhoun, H., Hemingway, H. and Poulter, N. (1998). Socio-economic status and blood pressure: an overview analysis. *Journal of Human Hypertension*, Vol. 12, No. 2, pp. 91–110.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, Vol. 19, No. 1, pp. 15–18.

Cook, R. D. and Weisberg, S. (1982). *Residuals and influence in regression*. Chapman and Hall, New York, 230 pages.

Cooley, W. (1978). Explanatory observational studies. *Educational Researcher*, Vol. 7, No. 9, pp. 9–15.

Costello, A. B. and Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, Vol. 10, No. 7, pp. 1–9.

De Gregorio, J. and Lee, J. (2002). Education and income inequality: New evidence from cross-country data. *Review of Income and Wealth*, Vol. 48, No. 3, pp. 395–416.

De Stavalo, B., Nitsch, D., dos Santos Silva, I. and Leon, D. (2006). Statistical issues in life course epidemiology. *American Journal of Epidemiology*, Vol. 163, No. 1, pp. 84–96.

De Villiers, L., Brown, M., Woolard, I., Daniels, R. and Leibbrandt, M. (2013). *National Income Dynamics Study Wave 3 User Manual*. Technical report, Cape Town: Southern Africa Labour and Development Research Unit.

Diamantopoulos, A. and Siguaw, J. (2000). *Introducing LISREL*. Sage Publications, London, 192 pages.

Farrel, A. (1994). Structural equation modelling with longitudinal data: Strategies for examining group differences and reciprocal relationships. *Journal of consulting and clinical psychology*, Vol. 62, pp. 477–487.

Fernald, L. and Adler, N. (2008). Blood pressure and socioeconomic status in low-income women in Mexico: A reverse gradient. *Journal of Epidemiology and Community Health*, Vol. 62, No. 5, pp. e8–e8.

Fikadu, G. and Lemma, S. (2016). Socioeconomic Status and Hypertension among Teachers and Bankers in Addis Ababa, Ethiopia. *International Journal of Hypertension*, Vol. 11, pp. 1–7.

Fritz, M. S. and MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, Vol. 18, pp. 233–239.

Gaskin, J. (2016). *SEM series (2016) 8. Mediation, Gaskination's Statistics*, [online], available: http://youtube.com/Gaskination [23 December 2016].

Goldberger, A. (1973). *Structural equation models: An overview. In Structural equation models in the social sciences*. Seminar press, New York, pp. 1–18.

Green, M., Jucha, E., Luz, Y. and Israel, R. (1986). Blood pressure in smokers and non-smokers: Epidemiologic findings. *American Heart Journal*, Vol. 111, No. 5, pp. 932–950.

Green, S., Akey, T., Fleming, K., Hershberger, S. and Marquis, J. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. *Structural Equation Modelling*, Vol. 4, pp. 108–120.

Gunzler, D., Chen, T., Wu, P. and Zhang, H. (2013). Introduction to mediation analysis with structural equation modelling. *Shanghai Archives of Psychiatry*, Vol. 25, No. 6, pp. 390–394.

Hair, J., Anderson, R., Tatham, R. and Black, W. (1998). *Multivariate data analysis with readings*. Prentice Hall, Englewood Cliffs, NJ, 768 pages.

Harrell, F. E. J. (2015). *Regression modelling Strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer International Publishing, Heidelberg, 582 pages.

Hayes, A. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, Vol. 76, No. 4, pp. 408–420.

Hayes, A. and Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, Vol. 24, No. 10, pp. 1918–1927.

Ho, M., Stark, S. and Chernyshenko, O. (2012). *Graphical representation of structural equation models using path diagrams. In Handbook of structural equation modelling*. The Guilford Press, New York, pp. 43–55.

Hooper, D., Coughlan, J. and Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, Vol. 6, No. 1, pp. 53–60.

Hox, J. and Bechger, T. (2007). An introduction to structural equation modelling. *Family Science Review*, Vol. 11, pp. 354–373.

Hoyle, R. (1995). *The structural equation modelling approach: Basic concepts and fundamental issues. In Structural equation modelling: Concepts, issues, and applications*. SAGE Publications Inc., Thousand Oaks, California, pp. 1–15.

Hu, G., Barengo, N., Tuomilehto, J., Lakka, T., Nissinen, A. and Jousilahtio, P. (2004). Relationship of physical activity and body mass index to the risk of hypertension: A prospective study in Finland. *Hypertension*, Vol. 43, No. 1, pp. 25–30.

Hu, L. and Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation modelling*, Vol. 6, No. 1, pp. 1–55.

Jackson, D. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation modelling*, Vol. 10, pp. 128–141.

James, L. R. and Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, Vol. 69, pp. 307–321.

Jamison, R., Sbrocco, T. and Parris, W. (1989). The influence of problems with concentration and memory on emotional distress and daily activities in chronic pain patients. *The International Journal of Psychiatry in Medicine*, Vol. 18, No. 2, pp. 183–191.

Jorde, L., Williams, R. and Kuida, H. (1986). Factor analysis suggesting contrasting determinants for different blood pressure measurements. *Hypertension*, Vol. 8, No. 3, pp. 243–251.

Jöreskog, K. (1973). *A general method for estimating a linear structural equation system. In Structural equation models in the social sciences.* Seminar Press, New York, pp. 85–112.

Jöreskog, K. and Sörbom, D. (1981). *LISREL 5: Analysis of linear structural relationships by maximum likelihood and least squares methods.* University of Uppsala, Department of Statistics, Chicago.

Jöreskog, K. and Sörbom, D. (1989). *LISREL 7: A guide to the program and its applications.* SPSS, Chicago.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, Vol. 39, No. 1, pp. 31–36.

Kaplan, G. and Keil, J. (1993). Socioeconomic factors and cardiovascular disease: A review of the literature. *Circulation*, Vol. 88, No. 4, pp. 1973–1998.

Kastorini, C., Milionis, H., Georgousopoulouy, E., Kalantzi, K., Nikolaouz, V., Vemmosx, K., Goudevenos, J. and Panagiotakos, D. (2015). Defining the path between social and economic factors, clinical and lifestyle determinants, and cardiovascular disease. *Global Heart*, Vol. 10, No. 4, pp. 255–263.

Keesling, J. (1972). *Maximum likelihood approaches to causal flow analysis.* PhD thesis, University of Chicago.

Keetile, M., Navaneetham, K. and Letamo, G. (2015). Patterns and determinants of hypertension in Botswana. *Journal of Public Health*, Vol. 23, No. 5, pp. 311–318.

Kleinbaum, D., Kupper, L. and Muller, K. (1988). *Applied regression analysis and other multivariable methods*. PWS - Kent publishing company, Boston, 798 pages.

Kline, R. (2011). *Principles and practice of structural equation modelling*. The Guilford Press, New York, London, 427 pages.

Kvalseth, T. O. (1985). Cautionary Note about R^2. *The American Statistician*, Vol. 39, No. 4, pp. 279–285.

Lam, C. (2011). The socioeconomics of hypertension: How $50 000 may buy a drop in blood pressure. *Hypertension*, Vol. 58, No. 2, pp. 140–141.

Lang, T., Pariente, P., Salem, G. and Tap, D. (1988). Social, professional conditions and arterial hypertension: an epidemiological study in Dakar, Senegal. *Journal of Hypertension*, Vol. 6, No. 4, pp. 271–276.

Law, M., Frost, C. and Wald, N. (1991). By how much does dietary salt reduction lower blood pressure. Analysis of observational data among populations. *British Medical Journal*, Vol. 302, No. 6780, pp. 819–824.

Lei, P. and Wu, Q. (2012). *Estimation in structural equation modelling. In Handbook of structural equation modelling*. The Guilford Press, New York, pp. 164–180.

Leibbrandt, M., Woolard, I. and De Villiers, L. (2009). *Methodology: Report on NIDS Wave 1 Technical Paper*. Technical Report No. 1, Cape Town: Southern Africa Labour and Development Research Unit, 31 July.

Leibbrandt, M., Woolard, I., Finn, A. and Argent, J. (2010). *Trends in South African income distribution and poverty since the fall of apartheid. In OECD Social, Employment and Migration Working Papers*. Technical Report No. 101, Paris: OECD Publishing, 28 May.

Leth-Steensen, C. and Gallito, E. (2016). Testing mediation in structural equation modelling: The effectiveness of the Test of Joint Significance. *Educational and Psychological Measurement*, Vol. 76, pp. 339–351.

MacCallum, R. (1986). Specification searches in covariance structure modelling. *Psychological bulletin*, Vol. 100, pp. 107–120.

MacCallum, R., Browne, M. and Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modelling. *Psychological Methods*, Vol. 1, No. 2, pp. 130–149.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. and Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, Vol. 7, pp. 83–104.

MacKinnon, D. P., Lockwood, C. M. and Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, Vol. 39, pp. 99–128.

MacKinnon, D. P., Fairchild, A. J. and Fritz, M. S. (2007)a. Mediation Analysis. *Annual Review of Psychology*, Vol. 58, pp. 593–614.

MacKinnon, D. P., Fritz, M. S., Williams, J. and Lockwood, C. M. (2007)b. Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, Vol. 39, pp. 384–389.

Mallinckrodt, B., Abraham, W. T., Wei, M. and Russell, D. W. (2006). Advances in testing the statistical significance of mediation effects. *Journal of Counseling Psychology*, Vol. 53, No. 3, pp. 372–378.

Mardia, K. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, Vol. 57, No. 3, pp. 519–530.

McDonald, R. and Ho, M. (2002). Principles and practice in reporting statistical equation analyses. *Psychological Methods*, Vol. 7, No. 1, pp. 64–82.

McIntosh, C. (2006). Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett. *Personality and Individual Differences*, Vol. 42, No. 5, pp. 859–867.

McNamee, R. (2005). Regression modelling and other methods to control confounding. *Occupational & Environmental Medicine*, Vol. 62, No. 7, pp. 500–506.

Mishra, V., Arnold, F., Semenov, G., Hong, R. and Mukuria, A. (2005). *DHS working paper: Epidemiology of obesity and hypertension in Uzbekistan*. Technical Report No. 25, United States Agency for International Development, 31 August.

Miyaki, K., Song, Y., Taneichi, S., Tsutsumi, A., Hashimoto, H., Kawakami, N., Takahashi, M., Shimazu, A., Inoue, A., Kurioka, S. and Shimbo, T. (2013). Socioeconomic status is significantly associated with the dietary intakes of folate and depression scales in Japanese workers (J-HOPE Study). *Nutrients*, Vol. 5, No. 2, pp. 565–578.

Mohanlal, P. (1998). *Structural equation modelling.* Master's thesis, University of South Africa.

Montgomery, D., Peck, E. and Vining, G. (2015). *Introduction to linear regression analysis.* John Wiley & Sons Inc., New York, 672 pages.

Mulaik, S., James, L., Van Alstine, J., Bennet, N., Lind, S. and Stilwell, C. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, Vol. 105, No. 3, pp. 430–435.

Musil, C. M., Jones, S. L. and Warner, C. D. (1998). Structural equation modelling and its relationship to multiple regression and factor analysis. *Research in Nursing & Health*, Vol. 21, pp. 271–281.

Muthén, B. and Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, Vol. 38, pp. 171–189.

Nachtigal, C., Kroehne, U., Funke, F., Steyer, R. and Schiller, F. (2003). (Why) should we use SEM? Pros and cons of structural equation modelling. *Methods of Psychological Research Online*, Vol. 8, No. 2, pp. 1–22.

NICE (2011). *Hypertension:Clinical management of primary hypertension in adults. NICE Clinical Guideline 127*, [online], available: http://www.nice.org.uk/guidance/cg127 [16 March 2016].

Norman, R., Bradshaw, D. and Steyn, K. (2001). *Chronic diseases, risk factors and lifestyle based on the, South Africa Adult Demographic and Health Survey. In poverty and chronic diseases in South Africa.* Technical Report No. 53, Tygerberg: Medical Research Council of South Africa - Burden of Disease Research Unit,.

Papathanasiou, G., Zerva, E., Zacharis, I., Papandreou, M., Papageorgiou, E., Tzima, C., Georgakopoulos, D. and Evangelou, A. (2015). Association of High Blood Pressure

with Body Mass Index, Smoking and Physical Activity in Healthy Young Adults. *The Open Cardiovascular Medicine Journal*, Vol. 9, pp. 5–17.

Pikhart, H., Bobak, M., Rose, R. and Marmot, M. (2003). Household item ownership and self-rated health: material and psychosocial explanations. *BMC Public Health*, Vol. 3, No. 1, pp. 38–44.

Ploubidis, G., Mathenge, W., De Stavola, B., Grundy, E., Foster, A. and Kuper, H. (2013). Socioeconomic position and later life prevalence of hypertension, diabetes and visual impairment in Nakuru, Kenya. *International Journal of Public Health*, Vol. 58, pp. 133–141.

Pourhoseingholi, M. A., Baghestani, A. R. and Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology Bed Bench*, Vol. 5, No. 2, pp. 79–83.

Preacher, K. and Hayes, A. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, Vol. 36, No. 4, pp. 717–731.

Preacher, K. and Hayes, A. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behaviour Research Methods*, Vol. 40, No. 3, pp. 879–891.

Preacher, K. and Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Journal of Psychological Methods*, Vol. 16, No. 2, pp. 93–115.

Primatesta, P., Falaschetti, E., Gupta, S., Marmot, M. and Poulter, N. (2001). Association between smoking and blood pressure evidence from the Health Survey for England. *Hypertension*, Vol. 37, No. 2, pp. 187–193.

Raykov, T. and Marcoulides, G. (2000). *A first course in structural equation modelling.* Lawrence Erlbaum Associates, Mahwah, New Jersey, 248 pages.

Reitermanova, Z. (2010). Data splitting. In *WDS'10 Proceedings of Contributed Papers: Paper I*, Prague. Matfyzpress, pp. 31–36.

Rigdon, E. E. (1998). *Structural equation modelling. In Modern methods for business research.* Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 251–294.

Robinson, J. P., Shaver, P. R. and Wrightsman, L. S. (1991). *Criteria for scale selection and evaluation. In Measures of personality and social psychological attitudes.* Academic Press, San Diego, pp. 1–16.

SAS Institute Inc. (2008). *SAS/STAT 9.3 User's Guide.* Cary, NC.

Satorra, A. and Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, Vol. 66, pp. 507–514.

Schumacker, R. and Lomax, R. (2004). *A beginner's guide to structural equation modelling.* Lawrence Erlbaum Associates, Mahwah, New Jersey, 498 pages.

Shisana, O., Labadarios, D., Rehle, T., Simbayi, L., Zuma, K., Dhansay, A., Reddy, P., Parker, W., Hoosain, E., Naidoo, P., Hongoro, C., Mchiza, Z., Steyn, N., Dwane, N., Makoae, M., Maluleke, T., Ramlagan, S., Zungu, N., Evans, M., Jacobs, L., Faber, M. and SANHANES-1 Team (2014). *South African National Health and Nutrition Examination Survey*. Technical report, Human Sciences Research Council.

Shrout, P. E. and Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, Vol. 7, pp. 422–445.

Silva, R., Scheines, R., Glymour, C. and Spirtes, P. (2006). Learning the structure of linear latent variable models. *Journal of Machine Learning Research*, Vol. 7, pp. 191–246.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In *Sociological Methodology*, San Franciso. Jossey-Bass, pp. 290–312.

Sodjinou, R., Agueh, V., Fayomi, B. and Delisle, H. (2008). Obesity and cardio-metabolic risk factors in urban adults of Benin: relationship with socio-economic status, urbanisation, and lifestyle patterns. *BMC Publlic Health*, Vol. 8, No. 1, pp. 84–96.

Southern Africa Labour and Development Research Unit (2013). *National Income Dynamics Study 2012, Wave 3 [dataset]. Version 1.2. Cape Town*, [online], available: http://www.nids.uct.ac.za [28 February 2016].

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, Vol. 15, No. 2, pp. 201–292.

Steiger, J. (1990). Structural model evaluation and modification. *Multivariate behavioural research*, Vol. 25, No. 2, pp. 173–180.

Steiger, J. (2007). Understanding the limitations of global fit assessment in structural equation modelling. *Personality and Individual Differences*, Vol. 42, No. 5, pp. 893–898.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 720 pages.

Steyn, K., Bradshaw, D., Norman, R. and Laubscher, R. (2008). Determinants and treatment of hypertension in South Africans: The first Demographic and Health Survey. *South African Medical Journal*, Vol. 98, No. 5, pp. 376–380.

Suhr, D. (2006). *The basics of structural equation modelling*, [online], available: http://www.lexjansen.com/wuss/2006/tutorials/TUT-suhr [3 March 2016].

Tabachnick, B. and Fidell, L. (2007). *Using Multivariate Statistics*. Allyn and Bacon, New York, 1008 pages.

Tsuruta, M., Adachi, H., Hirai, Y., Fujiura, Y. and Imaizumi, T. (2000). Association between alcohol intake and development of hypertension in Japanese normotensive men: 12-Year follow-up study. *American Journal of Hypertension*, Vol. 13, No. 5, pp. 482–487.

Tucker, L. and Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, Vol. 38, pp. 1–10.

Ullman, J. (2006). Structural equation modelling: Reviewing the basics and moving forward. *Journal of Personality Assessment*, Vol. 87, No. 1, pp. 35–50.

Van Rooyen, J., Kruger, H., Huisman, H., Wissing, M., Margetts, B., Venter, C., and Vorster, H. (2000). An epidemiological study of hypertension and its determinants in a population in transition: the THUSA study. *Journal of Human Hypertension*, Vol. 14, No. 12, pp. 779–787.

Weinberger, M. (1996). Salt sensitivity of blood pressure in humans. *Journal of Hypertension*, Vol. 27, No. 3, pp. 481–490.

Wheaton, B., Muthen, B., Alwin, D. and Summers, G. (1977). Assessing Reliability and Stability in Panel Models. *Sociological Methodology*, Vol. 8, No. 1, pp. 84–136.

WHO (2012). *World Health Statistics 2012 Global Health Observatory report*, [online], available: http://www.who.int/gho/publications/world-health-statistics/2012/en/ [24 March 2016].

WHO (2014). *Global status report on non-communicable diseases*, [online], available: http://www.who.int/nmh/publications/ncd-status-report-2014/en/ [24 March 2016].

Wiley, D. (1973). *The identification problem for structural equation models with unmeasured variables. In Structural equation models in the social sciences* . Seminar Press, New York, pp. 69–83.

Wong, M., Wang, H., Leung, M., Tsang, C., Lo, S. and Griffiths, S. (2015). The rising prevalence of self-reported hypertension among Chinese subjects: a population-based study from 121 895 household interviews. *QJM: An international Journal of Medicine*, Vol. 108, No. 1, pp. 9–17.

Wright, S. (1918). On the nature of size factors. *Genetics*, Vol. 3, pp. 367–374.

Wu, A. and Zumbo, B. (2008). Understanding and using mediators and moderators. *Social Indicators Research*, Vol. 87, pp. 367–392.

Yung, Y. and Bentler, P. (1994). Bootstrap-corrected ADF test statistics in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, Vol. 47, No. 1, pp. 63–84.

Zhang, J. and Kesteloot, H. (1999). Anthropometric, lifestyle and metabolic determinants of resting heart rate. A population study. *European Heart Journal*, Vol. 20, No. 2, pp. 103–110.

Appendix A

Description of variables

Table A.1: Description of variables used in analysis

Variable	Description
Age	Age of an individual
Gen	Gender
Race	Race
Empl_stat	Employment Status
Ownrad	Own radio
Ownhif	Own Hi-fi, mp3 player or cd player
Ownsew	Own sewing machine
Ownvehpri	Own private vehicle
Ownvehcom	Own bakkie or truck for commercial use
Ownmot	Own motorcycle or scooter
Ownbic	Own bicycle
Owncom	Own computer
Owncam	Own camera
Owncel	Own cellphone
Income	Income derived from various questions related to income
Hlbp_med	Use of antihypertensive medication
Emobth	Felt bothered
Emomnd	Trouble focusing
Emodep	Felt depressed
Emoeff	Felt everything was an effort
Emohope	Felt hopeful about future
Emofear	Felt fearful
Emoslp	Sleep was restless
Emohap	Was happy

Variable	Description
Emolone	Felt lonely
Emogo	Could not get going
Alcohol	Alcohol consumption derived from questions related to alcohol
Educ	Education derived from years of schooling
Smoker_Stat	Smoker Status derived from questions related to smoking
Exercise	How regularly respondent exercises
SBP_1	Systolic blood pressure reading 1
SBP_2	Systolic blood pressure reading 2
DBP_1	Diastolic blood pressure reading 1
DBP_2	Diastolic blood pressure reading 2
HR_1	Resting heart rate reading 1
HR_2	resting heart rate reading 2
Height_1	Height reading 1
Height_2	Height reading 2
Weight_1	Weight reading 1
Weight_2	Weight reading 2
BMI_1	Body mass index reading 1
BMI_2	Body mass index reading 2
Avg_BMI	Average body mass index reading

Appendix B

Sample characteristics

Table B.1: Demographic and SES statistics of population

Variable	N	Median or Percentage	IQR or Freq	Range
Age	18 707			
15-19		17.2%	3 218	
20-24		14.6%	2 723	
25-29		11.9%	2 217	
30-34		8.9%	1 659	
35-39		8.2%	1 529	
40-44		7.0%	1 317	
45-49		7.0%	1 303	
50-54		6.5%	1 213	
55-59		5.4%	1 011	
60-64		4.4%	829	
65-69		3.1%	576	
70-74		2.7%	499	
75-79		1.5%	287	
80-84		1.1%	205	
85+		0.6%	121	
Gender	18 710			
Male		40.5%	7 571	
Female		59.5%	11 139	

Variable	N	Median or Percentage	IQR or Freq	Range
Race	18 710			
African		82.4%	15 408	
Coloured		13.8%	2 576	
Asian/Indian		1.0%	195	
White		2.8%	531	
Employment	18 655			
Not Economically Active		49.8%	9 290	
Discouraged Unemployed		2.3%	438	
Strict Unemployed		14.1%	2 633	
Employed		33.7%	6 294	
Individual income (ZAR)	18 710	800	[0,1740]	[0,956000]
Education	18 695			
None		10.4%	1 944	
Primary		21.7%	4 059	
Secondary		56.7%	10 601	
Tertiary		11.2%	2 091	
Ownership of durable goods				
Radio	18 707	38.5%	7 204	
Hi-Fi, cd player or mp3 player	18 697	20.9%	3 905	
Sewing or Knitting machine	18 699	3.1%	581	
Motor vehicle (private) in running condition	18 694	6.9%	1 284	
Bakkie or truck in running condition	18 700	1.9%	360	
Motorcycle or scooter	18 695	0.6%	104	
Bicycle	18 701	2.5%	470	
Computer	18 700	6.4%	1 202	

Variable	N	Median or Percentage	IQR or Freq	Range
Camera	18 690	4.7%	877	
Cellphone	18 704	76.4%	14 287	
N = number of non-missing cases, IQR = interquartile range and Freq = Frequency				

Table B.2: Bio-behavioural factors statistics for population

Variable	N	Percentage	Freq
Physical exercise	18 697		
Never		71.7%	13 411
Less than once a week		7.0%	1 305
Once a week		4.1%	770
Twice a week		5.1%	949
Three or more times a week		12.1%	2 262
Alcohol consumption	18 582		
Non-drinker		77.0%	14 300
1 or 2 standard drinks		6.8%	1 270
3 or 4 standard drinks		6.4%	1 192
5 or 6 standard drinks		4.7%	877
7 or more standard drinks		5.1%	943
Smoker Status	18 595		
Non-smoker		80.8%	15 030
Smoke 5 or less cigarettes per day		10.3%	1 923
Smoke between 6 and 10 cigarettes per day		5.8%	1 072
Smoke more than 10 cigarettes per day		3.1%	570
Emotional well-being of individual			
Respondent bothered in past week	18 657		
Rarely or none of the time (less than 1 day)		69.1%	12 887
Some or little of the time (1-2 days)		23.3%	4 351
Occasionally or a moderate amount of time (3-4 days)		5.8%	1 073
All of the time (5-7 days)		1.9%	346

Variable	N	Percentage	Freq
Respondent had trouble focusing in past week	18 661		
Rarely or none of the time (less than 1 day)		61.6%	11 504
Some or little of the time (1-2 days)		28.3%	5 290
Occasionally or a moderate amount of time (3-4 days)		8.0%	1 496
All of the time (5-7 days)		2.0%	371
Respondent felt depressed in past week	18 630		
Rarely or none of the time (less than 1 day)		55.6%	10 358
Some or little of the time (1-2 days)		30.4%	5 664
Occasionally or a moderate amount of time (3-4 days)		11.0%	2 056
All of the time (5-7 days)		3.0%	552
Respondent felt that everything was an effort in past week	18 649		
Rarely or none of the time (less than 1 day)		50.0%	9 331
Some or little of the time (1-2 days)		25.5%	4 754
Occasionally or a moderate amount of time (3-4 days)		13.4%	2 506
All of the time (5-7 days)		11.0%	2 058
Respondent felt hopeful about the future in past week	18 651		
Rarely or none of the time (less than 1 day)		25.6%	4 771
Some or little of the time (1-2 days)		21.1%	3 929
Occasionally or a moderate amount of time (3-4 days)		20.7%	3 866
All of the time (5-7 days)		32.6%	6 085
Respondent felt fearful in past week	18 659		
Rarely or none of the time (less than 1 day)		63.7%	11 888
Some or little of the time (1-2 days)		27.8%	5 192
Occasionally or a moderate amount of time (3-4 days)		6.7%	1 257
All of the time (5-7 days)		1.7%	322

Variable	N	Percentage	Freq
Respondent's sleep was restless in past week	18 669		
Rarely or none of the time (less than 1 day)		56.1%	10 474
Some or little of the time (1-2 days)		31.0%	5 795
Occasionally or a moderate amount of time (3-4 days)		9.7%	1 809
All of the time (5-7 days)		3.2%	591
Respondent was happy in past week	18 671		
Rarely or none of the time (less than 1 day)		14.2%	2 644
Some or little of the time (1-2 days)		17.7%	3 303
Occasionally or a moderate amount of time (3-4 days)		28.9%	5 396
All of the time (5-7 days)		39.2%	7 328
Respondent felt lonely in past week	18 655		
Rarely or none of the time (less than 1 day)		60.9%	11 357
Some or little of the time (1-2 days)		28.7%	5 349
Occasionally or a moderate amount of time (3-4 days)		7.6%	1 427
All of the time (5-7 days)		2.8%	522
Respondent could not get going in past week	18 627		
Rarely or none of the time (less than 1 day)		68.7%	12 806
Some or little of the time (1-2 days)		22.7%	4 228
Occasionally or a moderate amount of time (3-4 days)		6.6%	1 237
All of the time (5-7 days)		1.9%	356
N = number of non-missing cases and Freq = Frequency			

Table B.3: Blood pressure, resting heart rate and medication use

Variable	N	Median or Percentage	IQR or Freq	Range
BMI reading 1	18 328	25	[21.5,29.9]	[11.8,78.5]
BMI reading 2	18 328	25	[21.5,29.9]	[11.8,78.2]
SBP reading 1	18 317	121	[110,136]	[80,239]
SBP reading 2	18 302	119	[108,133]	[80,235]
DBP reading 1	18 347	80	[72,90]	[40,140]
DBP reading 2	18 336	79	[71,89]	[40,140]
Heart rate reading 1	18 333	74	[65,83]	[32,150]
Heart rate reading 2	18 320	73	[65,82]	[30,142]
Use of Antihypertensive medication	18 705	13.5%	2 529	
Male	7 570	7.6%	577	
Female	11 135	17.5%	1 952	
Diagnosed with high blood pressure	18 666	17.3%	3 224	
Male	7 548	10.5%	790	
Female	11 118	21.9%	2 434	
Hypertension prevalence (based on SBP and DBP reading 1)	18 317	30.9%	5 658	
Male	7 420	30.1%	2 234	
Female	10 897	31.4%	3 424	
N = number of non-missing cases, IQR = interquartile range and Freq = Frequency				

Appendix C

Normality

Table C.1: Skewness and kurtosis for variables exceeding the limits for males dataset

	Income	Ownsew	Ownvehcom	Ownmot	Ownbic	Owncom	Owncam
Valid	7088	7088	7088	7088	7088	7088	7088
Skew	32.34	-8.24	-5.12	-9.99	-4.34	-3.01	-3.92
Kurt	1473.48	65.94	24.22	97.80	16.80	7.05	13.34

Table C.2: Skewness and kurtosis for variables exceeding the limits for females dataset

	Income	Ownsew	Ownvehpri	Ownvehcom	Ownmot	Ownbic	Owncom	Owncam
Valid	10508	10508	10508	10508	10508	10508	10508	10508
Skew	85.05	-4.58	-4.27	-10.61	-17.50	-9.28	-4.14	-4.70
Kurt	8128.40	18.99	16.24	110.53	304.21	84.10	15.11	20.07

Table C.3: Univariate and Multivariate normality for males

Variable	**min**	**max**	**skew**	**c.r.**	**kurtosis**	**c.r.**
Exercise	1.00	5.00	0.77	26.57	-1.17	-20.04
Alcohol	3.00	99.00	-0.50	-17.01	-1.75	-30.15
Smoker_Stat	1.00	4.00	1.44	49.54	0.97	16.66
Avg_BMI	12.83	60.34	1.24	42.52	2.78	47.83
HR_2	30.00	138.00	0.63	21.73	0.62	10.63
HR_1	35.00	136.00	0.64	21.98	0.64	11.02
Emogo	1.00	4.00	1.82	62.71	2.92	50.20
Emolone	1.00	4.00	1.56	53.56	1.99	34.14
Emoslp	1.00	4.00	1.35	46.26	1.23	21.11

Variable	min	max	skew	c.r.	kurtosis	c.r.
Emofear	1.00	4.00	1.61	55.41	2.29	39.37
Emodep	1.00	4.00	1.24	42.71	0.84	14.40
Emomnd	1.00	4.00	1.46	50.20	1.63	28.02
Emobth	1.00	4.00	1.84	63.17	3.11	53.36
SBP_2	80.00	227.00	1.21	41.60	2.43	41.72
SBP_1	80.00	227.00	1.12	38.40	2.06	35.41
DBP_2	41.00	140.00	0.57	19.41	0.75	12.96
DBP_1	41.00	138.00	0.58	19.79	0.65	11.23
Owncel_r	1.00	2.00	-1.16	-39.87	-0.66	-11.25
Ownhif_r	1.00	2.00	1.17	40.29	-0.63	-10.76
Ownrad_r	1.00	2.00	0.31	10.76	-1.90	-32.69
Educ	1.00	4.00	-0.60	-20.60	0.26	4.45
Ln_inc	0.00	13.11	-0.24	-8.32	-1.73	-29.68
Empl_stat	0.00	3.00	-0.08	-2.76	-1.86	-31.87
Multivariate					**112.52**	**139.67**

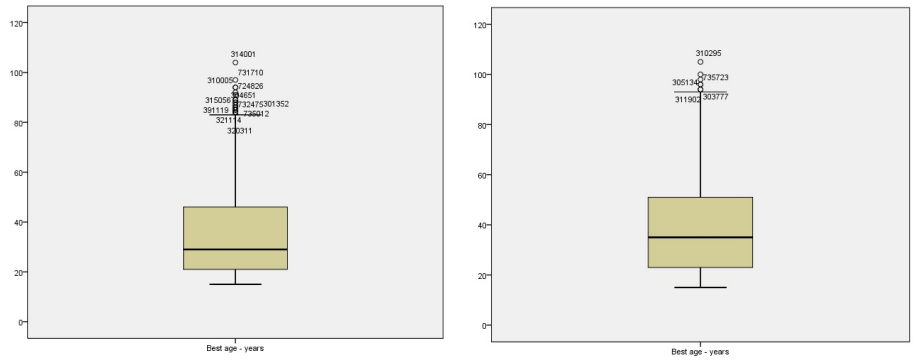
Table C.4: Univariate and Multivariate normality for females

Variable	min	max	skew	c.r.	kurtosis	c.r.
Exercise	1.00	5.00	2.36	98.65	4.14	86.68
Alcohol	3.00	99.00	-2.24	-93.92	3.04	63.58
Smoker_Stat	1.00	4.00	4.14	173.44	17.57	367.55
Avg_BMI	11.85	60.85	0.75	31.52	0.55	11.40
HR_2	39.00	142.00	0.41	16.96	0.33	6.92
HR_1	32.00	150.00	0.41	17.00	0.49	10.19
Emogo	1.00	4.00	1.64	68.66	2.21	46.34
Emolone	1.00	4.00	1.34	56.07	1.30	27.29
Emoslp	1.00	4.00	1.15	48.09	0.69	14.37
Emofear	1.00	4.00	1.38	57.62	1.47	30.77
Emodep	1.00	4.00	1.10	46.11	0.48	10.00
Emomnd	1.00	4.00	1.34	55.86	1.27	26.66
Emobth	1.00	4.00	1.73	72.34	2.72	56.89
SBP_2	80.00	235.00	1.28	53.60	2.10	43.93

Variable	**min**	**max**	**skew**	**c.r.**	**kurtosis**	**c.r.**
SBP_1	80.00	239.00	1.22	50.96	1.89	39.62
DBP_2	40.00	140.00	0.62	25.87	0.64	13.31
DBP_1	40.00	140.00	0.64	26.67	0.67	14.04
Owncel_r	1.00	2.00	-1.38	-57.90	-0.09	-1.80
Ownhif_r	1.00	2.00	1.66	69.38	0.75	15.66
Ownrad_r	1.00	2.00	0.59	24.70	-1.65	-34.56
Educ	1.00	4.00	-0.57	-23.67	-0.18	-3.84
Ln_inc	0.00	13.77	-0.80	-33.63	-1.00	-20.99
Empl_stat	0.00	3.00	0.37	15.50	-1.70	-35.52
Multivariate					**129.13**	**195.17**

Appendix D

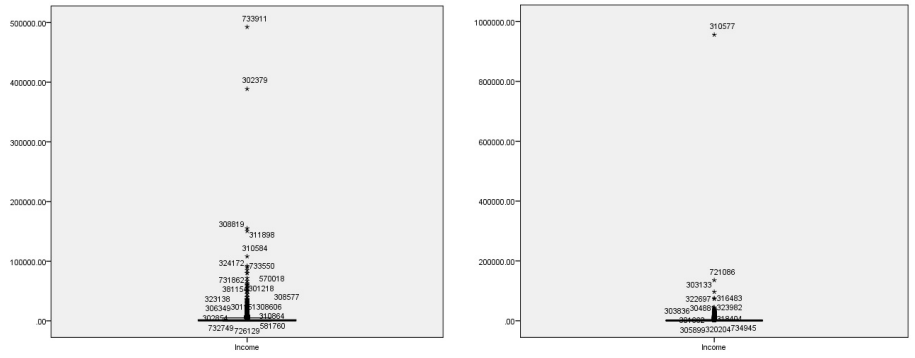
Outliers



(a) Age boxplot for males

(b) Age boxplot for females

Figure D.1: Age boxplots for both genders



(a) Income boxplot for males

(b) Income boxplot for females

Figure D.2: Income boxplots for both genders

Table D.1: First 10 Mahalanobis d-squared values for males

Observation number	**Mahalanobis d-squared**	**p1**	**p2**
3249	270.52	0	0
6962	199.34	0	0
6712	183.62	0	0
3985	173.41	0	0
6439	169.90	0	0
4497	139.26	0	0
3778	118.42	0	0
3474	108.18	0	0
6041	100.46	0	0
888	98.16	0	0

Table D.2: First 10 Mahalanobis d-squared values for females

Observation number	**Mahalanobis d-squared**	**p1**	**p2**
9948	237.18	0	0
2462	158.15	0	0
4584	143.42	0	0
2210	134.86	0	0
3030	134.15	0	0
267	123.14	0	0
9567	120.41	0	0
9466	106.44	0	0
208	104.74	0	0
6344	100.75	0	0

Appendix E

Cronbach's alpha

Table E.1: Reliability Statistics for 7 variables loading onto emotional well-being for females

Cronbach's Alpha	**Cronbach's Alpha Based on Standardised Items**	**Items**
0.80	0.80	7

Table E.2: Inter-Item Correlation Matrix for 7 variables loading onto emotional well-being for females

	Emobth	**Emomnd**	**Emodep**	**Emofear**	**Emoslp**	**Emolone**	**Emogo**
Emobth	1.00	0.51	0.42	0.31	0.33	0.32	0.32
Emomnd	0.51	1.00	0.43	0.37	0.35	0.34	0.37
Emodep	0.42	0.43	1.00	0.38	0.38	0.40	0.36
Emofear	0.31	0.37	0.38	1.00	0.31	0.37	0.35
Emoslp	0.33	0.35	0.38	0.31	1.00	0.33	0.33
Emolone	0.32	0.34	0.40	0.37	0.33	1.00	0.36
Emogo	0.32	0.37	0.36	0.35	0.33	0.36	1.00

Table E.3: Item-Total Statistics for 7 variables loading onto emotional well-being for females

	Scale Mean if Item Deleted	**Scale Variance if Item Deleted**	**Corrected Item-Total Correlation**	**Squared Multiple Correlation**	**Cronbach's Alpha if Item Deleted**
Emobth	9.25	9.46	0.54	0.33	0.77
Emomnd	9.15	9.14	0.58	0.37	0.77
Emodep	9.03	8.79	0.58	0.34	0.76
Emofear	9.17	9.48	0.51	0.26	0.78

	Scale Mean if Item Deleted	**Scale Variance if Item Deleted**	**Corrected Item-Total Correlation**	**Squared Multiple Correlation**	**Cronbach's Alpha if Item Deleted**
Emoslp	9.04	9.21	0.49	0.24	0.78
Emolone	9.12	9.28	0.52	0.27	0.78
Emogo	9.23	9.50	0.51	0.26	0.78

Appendix F

Exploratory Factor Analysis

Table F.1: Rotated Component matrix for males iteration 1

	Component				
	1	2	3	4	5
SBP_1	0.92				
SBP_2	0.92				
HR_1				0.97	
HR_2				0.97	
DBP_1	0.89				
DBP_2	0.89				
Emobth		0.68			
Emomnd		0.74			
Emodep		0.70			
Emofear		0.66			
Emoslp		0.65			
Emolone		0.63			
Emogo		0.64			
Empl_stat			0.70		
Educ					0.84
Ln_inc			0.78		
Ownrad			0.71		
Ownhif			0.64		
Owncel					0.61
Ownvehpri			0.39		0.42
Rotation Method: Varimax with Kaiser Normalization Rotation converged in 5 iterations					

Table F.2: KMO and Bartlett's Test for females

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.73
Bartlett's Test of Sphericity	Approx. Chi-Square	93768.05
	df	171
	Sig.	0.000

Table F.3: Variance explained when eigenvalues exceed 1 for females

Component	**Initial Eigenvalues**			**Rotation Sums Squared Loadings**		
	Total	**% Variance**	**Cum %**	**Total**	**% Variance**	**Cum %**
1	3.83	20.18	20.18	3.61	19.01	19.01
2	3.07	16.14	36.32	3.22	16.92	35.93
3	1.97	10.38	46.70	1.91	10.04	45.97
4	1.84	9.69	56.40	1.65	8.67	54.64
5	1.13	5.93	62.33	1.46	7.69	62.33
Extraction Method: Principal Component Analysis and Cum = Cumulative						

Table F.4: Rotated Component matrix for females

	Component				
	1	2	3	4	5
SBP_1	0.93				
SBP_2	0.93				
HR_1			0.97		
HR_2			0.97		
DBP_1	0.91				
DBP_2	0.91				
Emobth		0.69			
Emomnd		0.73			
Emodep		0.72			
Emofear		0.65			
Emoslp		0.63			
Emolone		0.66			

	Component				
	1	2	3	4	5
Emogo		0.65			
Empl_stat				0.84	
Educ				0.53	
Ln_inc				0.59	
Ownrad_r					0.84
Ownhif_r					0.83
Owncel_r				0.51	
Rotation Method: Varimax with Kaiser Normalization Rotation converged in 5 iterations					

Appendix G

Confirmatory Factor Analysis

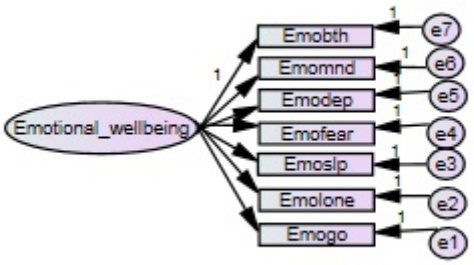


Figure G.1: Baseline model: Emotional well-being CFA model for males and females

Table G.1: Baseline model: Covariance modification indices for emotional well-being CFA model for males

			M.I.	Par Change
e7	⟷	e6	49.959	0.036
e6	⟷	e5	17.054	-0.023
e7	⟷	e4	10.31	-0.015
e7	⟷	e3	9.016	-0.016
e7	⟷	e1	8.102	-0.013
e5	⟷	e3	7.012	0.017
e7	⟷	e5	6.836	0.014
e4	⟷	e3	6.024	-0.014
e6	⟷	e2	4.897	-0.012
e2	⟷	e1	4.738	0.013
e4	⟷	e2	4.684	0.012
e7	⟷	e2	4.443	-0.011
e3	⟷	e2	4.292	0.013

Table G.2: Baseline model: Model fit indices for emotional well-being CFA model for males

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
140.271	10.019	0.925	0.016	0.0257	0.036	0.612	0.888

Table G.3: Modified model: Covariance modification indices for emotional well-being CFA model for males

			M.I.	Par Change
e7	$\longleftrightarrow$	e5	21.099	0.025
e4	$\longleftrightarrow$	e3	8.958	-0.017
e5	$\longleftrightarrow$	e2	5.814	-0.015
e5	$\longleftrightarrow$	e1	5.302	-0.013
e2	$\longleftrightarrow$	e1	4.19	0.013

Table G.4: Modified model: Model fit indices for emotional well-being CFA model for males

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
56.65	4.358	0.974	0.009	0.015	0.022	0.599	0.958

Table G.5: Modified model: Unstandardised regression coefficients for emotional well-being CFA model for males

			Estimate	S.E.	C.R.	P
Emobth	$\longleftarrow$	Emotional_well-being	1			
Emomnd	$\longleftarrow$	Emotional_well-being	1.256	0.034	36.476	***
Emodep	$\longleftarrow$	Emotional_well-being	1.329	0.042	31.275	***
Emofear	$\longleftarrow$	Emotional_well-being	1.064	0.039	27.564	***
Emoslp	$\longleftarrow$	Emotional_well-being	1.194	0.041	29.24	***
Emolone	$\longleftarrow$	Emotional_well-being	1.118	0.042	26.677	***
Emogo	$\longleftarrow$	Emotional_well-being	1.036	0.039	26.855	***

Table G.6: Baseline model: Covariance modification indices for emotional well-being CFA model for females

			M.I.	Par Change
e7	$\longleftrightarrow$	e6	84.154	0.04
e7	$\longleftrightarrow$	e4	20.702	-0.019
e6	$\longleftrightarrow$	e2	20.381	-0.02
e4	$\longleftrightarrow$	e2	10.132	0.016
e5	$\longleftrightarrow$	e3	8.032	0.016
e7	$\longleftrightarrow$	e1	7.967	-0.013
e2	$\longleftrightarrow$	e1	6.763	0.013
e6	$\longleftrightarrow$	e5	6.563	-0.012
e6	$\longleftrightarrow$	e3	6.05	-0.012
e7	$\longleftrightarrow$	e2	5.826	-0.011

Table G.7: Baseline model: Model fit indices for emotional well-being CFA model for females

χ^2	$\chi^2\backslash$DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
222.8	15.914	0.922	0.018	0.0282	0.038	0.611	0.883

Table G.8: Modified model: Covariance modification indices for emotional well-being CFA model for females

			M.I.	Par Change
e5	$\longleftrightarrow$	e1	15.049	-0.019
e7	$\longleftrightarrow$	e5	15.017	0.018
e6	$\longleftrightarrow$	e2	13.208	-0.016
e4	$\longleftrightarrow$	e2	8.435	0.014
e7	$\longleftrightarrow$	e4	6.955	-0.011
e6	$\longleftrightarrow$	e4	4.901	0.01
e2	$\longleftrightarrow$	e1	4.758	0.011

Table G.9: Modified model: Model fit indices for emotional well-being CFA model for females

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
75.717	5.824	0.976	0.009	0.0145	0.021	0.602	0.962

Table G.10: Modified model: Unstandardised regression coefficients for emotional well-being CFA model for females

			Estimate	S.E.	C.R.	P
Emobth	⟵	Emotional_well-being	1			
Emomnd	⟵	Emotional_well-being	1.145	0.025	46.435	***
Emodep	⟵	Emotional_well-being	1.371	0.034	40.601	***
Emofear	⟵	Emotional_well-being	1.052	0.032	33.315	***
Emoslp	⟵	Emotional_well-being	1.126	0.032	35.223	***
Emolone	⟵	Emotional_well-being	1.131	0.033	34.339	***
Emogo	⟵	Emotional_well-being	1.044	0.031	33.374	***

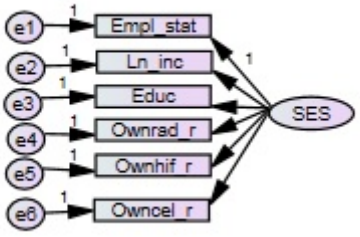


Figure G.2: Baseline model: SES CFA model for males and females

Table G.11: Baseline model: Covariance modification indices for SES CFA model for males

			M.I.	Par Change
e4	⟷	e5	247.549	0.03
e3	⟷	e6	161.532	0.043
e2	⟷	e3	131.883	-0.299
e3	⟷	e5	128.121	0.036
e3	⟷	e4	116.463	-0.039
e1	⟷	e2	77.949	0.308

			M.I.	Par Change
e1	$\longleftrightarrow$	e5	68.587	-0.04
e1	$\longleftrightarrow$	e3	63.767	0.081
e1	$\longleftrightarrow$	e4	52.057	-0.041
e4	$\longleftrightarrow$	e6	48.12	0.013
e2	$\longleftrightarrow$	e4	16.529	-0.062
e2	$\longleftrightarrow$	e5	16.145	-0.053
e2	$\longleftrightarrow$	e6	15.866	-0.058
e5	$\longleftrightarrow$	e6	14.864	-0.005

Table G.12: Baseline model: Model fit indices for SES CFA model for males

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
1258.992	139.888	0.821	0.132	0.0905	0.14	0.492	0.702

Table G.13: Modified model: Covariance modification indices for SES CFA model for males

			M.I.	Par Change
e1	$\longleftrightarrow$	e6	19.03	-0.025
e4	$\longleftrightarrow$	e6	13.007	0.007
e2	$\longleftrightarrow$	e5	8.265	-0.038
e5	$\longleftrightarrow$	e6	5.397	0.003

Table G.14: Modified model: Model fit indices for SES CFA model for males

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
76.286	19.071	0.99	0.024	0.0209	0.051	0.264	0.961

Table G.15: Modified model: Unstandardised regression coefficients for SES CFA model for males

			Estimate	S.E.	C.R.	P
Empl_stat	⟵	SES	1			
Ln_inc	⟵	SES	3.27	0.095	34.398	***
Educ	⟵	SES	0.271	0.015	18.111	***
Ownrad_r	⟵	SES	0.243	0.007	35.506	***
Ownhif_r	⟵	SES	0.182	0.006	30.5	***
Owncel_r	⟵	SES	0.187	0.008	22.811	***

Table G.16: Baseline model: Covariance modification indices for SES CFA model for females

			M.I.	Par Change
e2	⟷	e3	488.042	-0.446
e4	⟷	e5	425.953	0.029
e1	⟷	e3	355.394	0.165
e3	⟷	e6	225.287	0.043
e3	⟷	e4	157.4	-0.04
e1	⟷	e4	154.11	-0.06
e3	⟷	e5	108.496	0.027
e1	⟷	e2	78.77	0.21
e2	⟷	e5	51.554	-0.062
e1	⟷	e5	50.106	-0.027
e4	⟷	e6	42.444	0.01
e5	⟷	e6	42.332	-0.007
e1	⟷	e6	10.718	-0.014

Table G.17: Baseline model: Model fit indices for SES CFA model for females

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
2634.737	292.749	0.632	0.13	0.1145	0.167	0.379	0.386

Table G.18: Modified model: Covariance modification indices for SES CFA model for females

			M.I.	Par Change
e1	$\longleftrightarrow$	e5	22.21	0.018
e1	$\longleftrightarrow$	e6	13.728	-0.016
e3	$\longleftrightarrow$	e5	8.063	-0.007
e3	$\longleftrightarrow$	e6	7.63	0.008

Table G.19: Modified model: Model fit indices for SES CFA model for females

χ^2	$\chi^2 \backslash$DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
54.564	18.188	0.993	0.016	0.0131	0.04	0.198	0.964

Table G.20: Modified model: Unstandardised regression coefficients for SES CFA model for females

			Estimate	S.E.	C.R.	P
Empl_stat	$\longleftarrow$	SES	1			
Ln_inc	$\longleftarrow$	SES	3.4	0.171	19.836	***
Educ	$\longleftarrow$	SES	0.612	0.028	22.047	***
Ownrad_r	$\longleftarrow$	SES	0.213	0.008	28.228	***
Ownhif_r	$\longleftarrow$	SES	0.118	0.006	20.15	***
Owncel_r	$\longleftarrow$	SES	0.247	0.011	22.998	***

Appendix H

SEM model 1 results

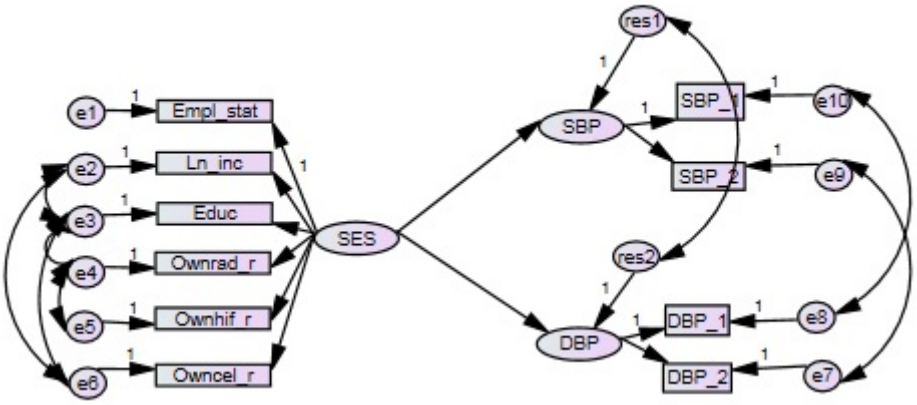


Figure H.1: Baseline model: Model investigating direct relationship of SES and blood pressure for males

Table H.1: Baseline model: Model fit indices for SEM model 1 for males

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
441.152	17.646	0.963	17.075	0.0556	0.048	0.534	0.933

Table H.2: Baseline model: Unstandardised regression coefficients for SEM model 1 for males

			Estimate	S.E.	C.R.	P
DBP	⟵	SES	3.173	0.171	18.569	***
SBP	⟵	SES	4.388	0.25	17.562	***
Empl_stat	⟵	SES	1			
Ln_inc	⟵	SES	3.117	0.078	39.931	***
Educ	⟵	SES	0.205	0.012	16.867	***
Ownrad_r	⟵	SES	0.241	0.006	37.581	***
Ownhif_r	⟵	SES	0.179	0.006	31.677	***

			Estimate	S.E.	C.R.	P
Owncel_r	⟵	SES	0.17	0.007	24.113	***
DBP_1	⟵	DBP	1			
DBP_2	⟵	DBP	1.042	0.022	48.231	***
SBP_1	⟵	SBP	1			
SBP_2	⟵	SBP	1.033	0.021	49.431	***

Table H.3: Baseline model: Covariances for SEM model 1 for males

			Estimate	S.E.	C.R.	P
res1	⟷	res2	147.465	5.158	28.592	***
e5	⟷	e4	0.06	0.002	24.092	***
e6	⟷	e3	0.04	0.004	10.335	***
e3	⟷	e2	-0.463	0.034	-13.517	***
e6	⟷	e2	-0.227	0.019	-11.884	***
e3	⟷	e4	-0.033	0.004	-8.421	***
e7	⟷	e9	-3.02	3.404	-0.887	0.375
e8	⟷	e10	16.729	3.214	5.206	***

Table H.4: Modified model: Model fit indices for SEM model 1 for males

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
441.96	16.998	0.963	17.123	0.0557	0.048	0.555	0.935

Table H.5: Modified model: Unstandardised regression coefficients for SEM model 1 for males

			Estimate	S.E.	C.R.	P
DBP	⟵	SES	3.215	0.165	19.512	***
SBP	⟵	SES	4.452	0.241	18.458	***
Empl_stat	⟵	SES	1			
Ln_inc	⟵	SES	3.117	0.078	39.89	***
Educ	⟵	SES	0.205	0.012	16.834	***
Ownrad_r	⟵	SES	0.24	0.006	37.547	***
Ownhif_r	⟵	SES	0.179	0.006	31.649	***

			Estimate	S.E.	C.R.	P
Owncel_r	⟵	SES	0.17	0.007	24.101	***
DBP_1	⟵	DBP	1			
DBP_2	⟵	DBP	1.025	0.01	103.317	***
SBP_1	⟵	SBP	1			
SBP_2	⟵	SBP	1.016	0.008	125.483	***

Table H.6: Modified model: Covariances for SEM model 1 for males

			Estimate	S.E.	C.R.	P
res1	⟷	res2	149.608	4.653	32.154	***
e5	⟷	e4	0.06	0.002	24.1	***
e6	⟷	e3	0.041	0.004	10.337	***
e3	⟷	e2	-0.463	0.034	-13.518	***
e6	⟷	e2	-0.227	0.019	-11.887	***
e3	⟷	e4	-0.033	0.004	-8.438	***
e8	⟷	e10	13.972	0.985	14.183	***

Table H.7: Modified model: Variances for SEM model 1 for males

	Estimate	S.E.	C.R.	P
SES	0.973	0.03	32.765	***
res1	285.167	8.833	32.284	***
res2	131.644	3.4	38.717	***
e8	28.406	1.445	19.653	***
e7	16.975	1.163	14.593	***
e9	16.779	1.888	8.885	***
e10	44.907	2.123	21.149	***
e6	0.155	0.003	58.847	***
e3	0.488	0.009	52.116	***
e2	4.994	0.25	19.984	***
e5	0.154	0.002	63.913	***
e1	0.956	0.028	33.826	***
e4	0.186	0.003	69.569	***

Table H.8: Modified model: Squared multiple correlations for SEM model 1 for males

	Estimate
SBP	0.063
DBP	0.071
SBP_2	0.949
SBP_1	0.871
DBP_2	0.898
DBP_1	0.833
Owncel_r	0.154
Ownhif_r	0.168
Ownrad_r	0.232
Educ	0.077
Ln_inc	0.654
Empl_stat	0.504

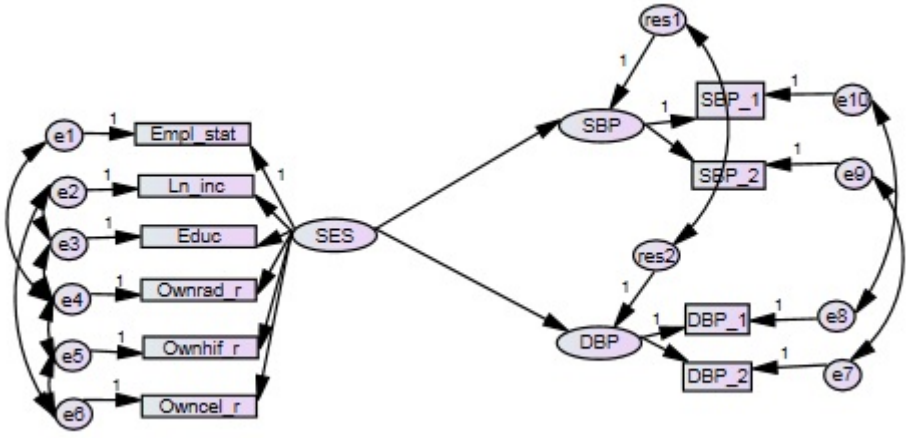


Figure H.2: Baseline model: Model investigating direct relationship of SES and blood pressure for females

Table H.9: Baseline model: Model fit indices for SEM model 1 for females

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
1565.06	65.211	0.945	49.204	0.0846	0.058	0.466	0.916

Table H.10: Baseline model: Unstandardised regression coefficients for SEM model 1 for females

			Estimate	S.E.	C.R.	P
DBP	⟵	SES	2.129	0.183	11.642	***
SBP	⟵	SES	2.389	0.277	8.623	***
Empl_stat	⟵	SES	1			
Ln_inc	⟵	SES	2.848	0.115	24.741	***
Educ	⟵	SES	0.385	0.018	21.83	***
Ownrad_r	⟵	SES	0.201	0.007	29.815	***
Ownhif_r	⟵	SES	0.129	0.006	23.334	***
Owncel_r	⟵	SES	0.193	0.008	22.783	***
DBP_1	⟵	DBP	1			
DBP_2	⟵	DBP	1.025	0.035	28.88	***
SBP_1	⟵	SBP	1			
SBP_2	⟵	SBP	1.04	0.036	28.571	***

Table H.11: Baseline model: Covariances for SEM model 1 for females

			Estimate	S.E.	C.R.	P
res1	⟷	res2	159.861	6.827	23.417	***
e3	⟷	e2	-0.685	0.036	-19.281	***
e4	⟷	e5	0.07	0.002	34.214	***
e4	⟷	e3	-0.043	0.004	-11.642	***
e4	⟷	e1	-0.03	0.006	-4.765	***
e6	⟷	e2	-0.17	0.019	-9.088	***
e6	⟷	e5	0.004	0.001	3.067	0.002
e7	⟷	e9	-3.045	6.033	-0.505	0.614
e8	⟷	e10	20.708	5.743	3.606	***

Table H.12: Modified model: Model fit indices for SEM model 1 for females

χ^2	χ^2\DF	CFI	RMR	SRMR	RMSEA	PNFI	TLI
1565.318	62.613	0.945	49.235	0.0846	0.057	0.485	0.925

Table H.13: Modified model: Unstandardised regression coefficients for SEM model 1 for females

			Estimate	S.E.	C.R.	P
DBP	⟵	SES	2.163	0.174	12.443	***
SBP	⟵	SES	2.44	0.267	9.143	***
Empl_stat	⟵	SES	1			
Ln_inc	⟵	SES	2.847	0.115	24.743	***
Educ	⟵	SES	0.384	0.018	21.803	***
Ownrad_r	⟵	SES	0.201	0.007	29.839	***
Ownhif_r	⟵	SES	0.129	0.006	23.322	***
Owncel_r	⟵	SES	0.193	0.008	22.775	***
DBP_1	⟵	DBP	1			
DBP_2	⟵	DBP	1.008	0.009	108.923	***
SBP_1	⟵	SBP	1			
SBP_2	⟵	SBP	1.022	0.007	144.443	***

Table H.14: Modified model: Covariances for SEM model 1 for females

			Estimate	S.E.	C.R.	P
res1	⟷	res2	162.531	4.487	36.224	***
e3	⟷	e2	-0.684	0.035	-19.283	***
e4	⟷	e5	0.07	0.002	34.198	***
e4	⟷	e3	-0.043	0.004	-11.642	***
e4	⟷	e1	-0.03	0.006	-4.77	***
e6	⟷	e2	-0.17	0.019	-9.083	***
e6	⟷	e5	0.004	0.001	3.072	0.002
e8	⟷	e10	17.799	0.935	19.041	***

Table H.15: Modified model: Variances for SEM model 1 for females

	Estimate	S.E.	C.R.	P
SES	0.681	0.029	23.266	***
res1	288.5	8.714	33.106	***
res2	134.756	3.085	43.687	***
e8	30.256	1.23	24.597	***
e7	16.918	0.987	17.139	***
e9	15.702	1.515	10.363	***
e10	48.642	1.889	25.749	***
e6	0.132	0.002	61.279	***
e4	0.2	0.002	95.435	***
e3	0.491	0.009	57.663	***
e2	5.089	0.237	21.438	***
e5	0.139	0.002	62.058	***
e1	1.151	0.029	40.352	***

Table H.16: Modified model: Squared multiple correlations for SEM model 1 for females

	Estimate
SBP	0.014
DBP	0.023
SBP_2	0.951
SBP_1	0.857
DBP_2	0.892
DBP_1	0.82
Owncel_r	0.161
Ownhif_r	0.075
Ownrad_r	0.121
Educ	0.17
Ln_inc	0.52
Empl_stat	0.372

Appendix I

SEM model 2 results

Table I.1: Unstandardised regression coefficients for SEM model 2.1 for males

			Estimate	S.E.	C.R.	P
Exercise	⟵	SES	0.007	0.02	0.369	0.712
Emotional_well-being	⟵	SES	-0.011	0.005	-2.276	0.023
Smoker_Stat	⟵	SES	0.168	0.01	16.47	***
Alcohol	⟵	SES	0.281	0.016	17.448	***
Heart_Rate	⟵	SES	0.202	0.145	1.392	0.164
Heart_Rate	⟵	Smoker_Stat	0.814	0.182	4.484	***
Heart_Rate	⟵	Exercise	-0.714	0.084	-8.521	***
Avg_BMI	⟵	SES	1.629	0.058	27.995	***
Avg_BMI	⟵	Exercise	-0.298	0.03	-10.04	***
Avg_BMI	⟵	Smoker_Stat	-0.284	0.074	-3.832	***
Avg_BMI	⟵	Alcohol	-0.254	0.041	-6.182	***
Avg_BMI	⟵	Emotional_well-being	-0.395	0.169	-2.334	0.02
SBP	⟵	Smoker_Stat	1.803	0.299	6.03	***
DBP	⟵	Alcohol	0.437	0.11	3.96	***
SBP	⟵	Exercise	0.382	0.122	3.142	0.002
DBP	⟵	Exercise	-0.084	0.083	-1.011	0.312
DBP	⟵	Emotional_well-being	1.163	0.452	2.573	0.01
SBP	⟵	Heart_Rate	0.027	0.02	1.364	0.173
DBP	⟵	Heart_Rate	0.142	0.013	10.651	***
DBP	⟵	Smoker_Stat	0.545	0.186	2.926	0.003
DBP	⟵	SES	1.931	0.157	12.273	***
SBP	⟵	Alcohol	0.636	0.171	3.726	***
SBP	⟵	Emotional_well-being	1.51	0.73	2.069	0.039
SBP	⟵	SES	2.209	0.239	9.26	***

			Estimate	S.E.	C.R.	P
DBP	⟵	Avg_BMI	0.573	0.034	16.871	***
SBP	⟵	Avg_BMI	0.91	0.058	15.787	***
Empl_stat	⟵	SES	1			
Ln_inc	⟵	SES	2.777	0.057	49.091	***
Educ	⟵	SES	0.162	0.009	17.239	***
Ownrad_r	⟵	SES	0.23	0.006	40.9	***
Ownhif_r	⟵	SES	0.167	0.005	33.305	***
Owncel_r	⟵	SES	0.157	0.006	28.01	***
DBP_1	⟵	DBP	1			
DBP_2	⟵	DBP	1.035	0.01	107.624	***
SBP_1	⟵	SBP	1			
SBP_2	⟵	SBP	1.018	0.008	126.831	***
Emobth	⟵	Emotional_well-being	1			
Emomnd	⟵	Emotional_well-being	1.266	0.036	35.125	***
Emodep	⟵	Emotional_well-being	1.36	0.045	30.123	***
Emofear	⟵	Emotional_well-being	1.088	0.041	26.463	***
Emoslp	⟵	Emotional_well-being	1.2	0.043	27.822	***
Emolone	⟵	Emotional_well-being	1.131	0.044	25.833	***
Emogo	⟵	Emotional_well-being	1.041	0.04	25.726	***
HR_1	⟵	Heart_Rate	1.04	0.026	40.024	***
HR_2	⟵	Heart_Rate	1			

Table I.2: Standardised regression coefficients for SEM model 2.1 for males

			Estimate
Exercise	⟵	SES	0.005
Emotional_well-being	⟵	SES	-0.033
Smoker_Stat	⟵	SES	0.214
Alcohol	⟵	SES	0.222
Heart_Rate	⟵	SES	0.019
Heart_Rate	⟵	Smoker_Stat	0.059
Heart_Rate	⟵	Exercise	-0.099
Avg_BMI	⟵	SES	0.372

			Estimate
Avg_BMI	⟵	Exercise	-0.103
Avg_BMI	⟵	Smoker_Stat	-0.051
Avg_BMI	⟵	Alcohol	-0.073
Avg_BMI	⟵	Emotional_well-being	-0.03
SBP	⟵	Smoker_Stat	0.09
DBP	⟵	Alcohol	0.05
SBP	⟵	Exercise	0.037
DBP	⟵	Exercise	-0.012
DBP	⟵	Emotional_well-being	0.035
SBP	⟵	Heart_Rate	0.019
DBP	⟵	Heart_Rate	0.141
DBP	⟵	Smoker_Stat	0.039
DBP	⟵	SES	0.175
SBP	⟵	Alcohol	0.051
SBP	⟵	Emotional_well-being	0.032
SBP	⟵	SES	0.14
DBP	⟵	Avg_BMI	0.228
SBP	⟵	Avg_BMI	0.253
Empl_stat	⟵	SES	0.773
Ln_inc	⟵	SES	0.781
Educ	⟵	SES	0.258
Ownrad_r	⟵	SES	0.501
Ownhif_r	⟵	SES	0.416
Owncel_r	⟵	SES	0.401
DBP_1	⟵	DBP	0.915
DBP_2	⟵	DBP	0.953
SBP_1	⟵	SBP	0.932
SBP_2	⟵	SBP	0.973
Emobth	⟵	Emotional_well-being	0.562
Emomnd	⟵	Emotional_well-being	0.653
Emodep	⟵	Emotional_well-being	0.638
Emofear	⟵	Emotional_well-being	0.594

			Estimate
Emoslp	⟵	Emotional_well-being	0.576
Emolone	⟵	Emotional_well-being	0.558
Emogo	⟵	Emotional_well-being	0.567
HR_1	⟵	Heart_Rate	0.97
HR_2	⟵	Heart_Rate	0.938

Table I.3: Squared multiple correlations for SEM model 2.1 for males

	Estimate
Exercise	0
Alcohol	0.049
Smoker_Stat	0.046
Emotional_well-being	0.001
Avg_BMI	0.141
Heart_Rate	0.014
SBP	0.131
DBP	0.147
HR_2	0.88
HR_1	0.941
Emogo	0.322
Emolone	0.311
Emoslp	0.331
Emofear	0.352
Emodep	0.407
Emomnd	0.427
Emobth	0.316
SBP_2	0.947
SBP_1	0.869
DBP_2	0.908
DBP_1	0.837
Owncel_r	0.161
Ownhif_r	0.173
Ownrad_r	0.251

	Estimate
Educ	0.067
Ln_inc	0.61
Empl_stat	0.598

Table I.4: Standardised indirect effects for SEM model 2.1 for males

	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate
Avg_BMI	-0.027	0	0	0	0	0	0
Heart_Rate	0.012	0	0	0	0	0	0
SBP	0.118	-0.028	-0.018	-0.012	-0.007	0	0
DBP	0.101	-0.037	-0.017	-0.003	-0.007	0	0
HR_2	0.029	-0.093	0	0.055	0	0	0
HR_1	0.03	-0.096	0	0.057	0	0	0
Emogo	-0.019	0	0	0	0	0	0
Emolone	-0.018	0	0	0	0	0	0
Emoslp	-0.019	0	0	0	0	0	0
Emofear	-0.019	0	0	0	0	0	0
Emodep	-0.021	0	0	0	0	0	0
Emomnd	-0.021	0	0	0	0	0	0
Emobth	-0.018	0	0	0	0	0	0
SBP_2	0.251	0.009	0.032	0.076	0.023	0.246	0.018
SBP_1	0.24	0.008	0.03	0.073	0.022	0.236	0.017
DBP_2	0.264	-0.047	0.032	0.034	0.027	0.217	0.134
DBP_1	0.253	-0.045	0.031	0.032	0.026	0.208	0.129

Table I.5: Significance of standardised indirect effects for SEM model 2.1 for males

	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate
Avg_BMI	0.005	...	...	...	...	...	...
Heart_Rate	0.009	...	...	...	...	...	...
SBP	0.003	0.002	0.003	0.011	0.115	...	...
DBP	0.005	0.002	0.004	0.505	0.125	...	...
HR_2	0.052	0.003	...	0.007	...	...	...
HR_1	0.048	0.005	...	0.006	...	...	...
Emogo	0.074	...	...	...	...	...	...
Emolone	0.082	...	...	...	...	...	...
Emoslp	0.093	...	...	...	...	...	...
Emofear	0.072	...	...	...	...	...	...
Emodep	0.074	...	...	...	...	...	...
Emomnd	0.079	...	...	...	...	...	...
Emobth	0.082	...	...	...	...	...	...
SBP_2	0.011	0.635	0.093	0.002	0.16	0.001	0.256
SBP_1	0.011	0.628	0.095	0.002	0.163	0.002	0.261
DBP_2	0.007	0.003	0.052	0.049	0.07	0.003	0.001
DBP_1	0.011	0.003	0.05	0.052	0.073	0.003	0.001

Table I.6: Specific indirect effects for SEM model 2.1 for males

Parameter	Standardised estimate	Estimate	Lower	Upper	P	Mediation
SES - BMI - SBP	0.0941	1.482	1.297	1.728	0.001	Yes
SES - BMI - DBP	0.0848	0.933	0.824	1.082	0.002	Yes
SES - Alcohol - SBP	0.0113	0.179	0.085	0.261	0.009	Yes
SES - Alcohol - DBP	0.0111	0.123	0.059	0.179	0.008	Yes
SES - Smoker Stat - SBP	0.0193	0.303	0.215	0.428	0.001	Yes
SES - Smoker Stat - DBP	0.0083	0.092	0.036	0.162	0.007	Yes
SES - Exercise - SBP	0.0002	0.003	-0.015	0.026	0.635	No
SES - Exercise - DBP	-0.0001	-0.001	-0.009	0.002	0.532	No
SES - HR - SBP	0.0004	0.005	-0.002	0.025	0.197	No
SES - HR - DBP	0.0027	0.029	-0.01	0.071	0.184	No
SES - Emotional well-being - SBP	-0.0011	-0.016	-0.05	-0.001	0.049	Yes
SES - Emotional well-being - DBP	-0.0012	-0.013	-0.031	-0.001	0.049	Yes

Table I.7: Standardised total effects for SEM model 2.1 for males

	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate	SBP	DBP
Exercise	0.005	0	0	0	0	0	0	0	0
Alcohol	0.222	0	0	0	0	0	0	0	0
Smoker_Stat	0.214	0	0	0	0	0	0	0	0
Emotional_well-being	-0.033	0	0	0	0	0	0	0	0
Avg_BMI	0.345	-0.103	-0.073	-0.051	-0.03	0	0	0	0
Heart_Rate	0.031	-0.099	0	0.059	0	0	0	0	0
SBP	0.258	0.009	0.032	0.078	0.024	0.253	0.019	0	0
DBP	0.277	-0.049	0.034	0.035	0.028	0.228	0.141	0	0
HR_2	0.029	-0.093	0	0.055	0	0	0.938	0	0
HR_1	0.03	-0.096	0	0.057	0	0	0.97	0	0
Emogo	-0.019	0	0	0	0.567	0	0	0	0
Emolone	-0.018	0	0	0	0.558	0	0	0	0
Emoslp	-0.019	0	0	0	0.576	0	0	0	0
Emofear	-0.019	0	0	0	0.594	0	0	0	0
Emodep	-0.021	0	0	0	0.638	0	0	0	0
Emomnd	-0.021	0	0	0	0.653	0	0	0	0
Emobth	-0.018	0	0	0	0.562	0	0	0	0
SBP_2	0.251	0.009	0.032	0.076	0.023	0.246	0.018	0.973	0

	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate	SBP	DBP
SBP_1	0.24	0.008	0.03	0.073	0.022	0.236	0.017	0.932	0
DBP_2	0.264	-0.047	0.032	0.034	0.027	0.217	0.134	0	0.953
DBP_1	0.253	-0.045	0.031	0.032	0.026	0.208	0.129	0	0.915
Owncel_r	0.401	0	0	0	0	0	0	0	0
Ownhif_r	0.416	0	0	0	0	0	0	0	0
Ownrad_r	0.501	0	0	0	0	0	0	0	0
Educ	0.258	0	0	0	0	0	0	0	0
Ln_inc	0.781	0	0	0	0	0	0	0	0
Empl_stat	0.773	0	0	0	0	0	0	0	0

Table I.8: Unstandardised regression coefficients for SEM model 2.1 for females

			Estimate	S.E.	C.R.	P
Exercise	⟵	SES	0.069	0.015	4.667	***
Emotional_well-being	⟵	SES	-0.01	0.005	-1.956	0.05
Smoker_Stat	⟵	SES	0.018	0.006	3.059	0.002
Alcohol	⟵	SES	0.047	0.01	4.942	***
Heart_Rate	⟵	SES	-0.701	0.149	-4.706	***
Heart_Rate	⟵	Smoker_Stat	1.117	0.336	3.322	***
Heart_Rate	⟵	Exercise	-0.349	0.126	-2.769	0.006
Avg_BMI	⟵	SES	2.247	0.084	26.662	***
Avg_BMI	⟵	Exercise	-0.241	0.066	-3.667	***
Avg_BMI	⟵	Smoker_Stat	0.044	0.185	0.235	0.814
Avg_BMI	⟵	Alcohol	-0.551	0.097	-5.696	***
Avg_BMI	⟵	Emotional_well-being	0.479	0.197	2.426	0.015
SBP	⟵	Smoker_Stat	3.15	0.603	5.222	***
DBP	⟵	Alcohol	0.627	0.185	3.389	***
SBP	⟵	Exercise	1.208	0.202	5.98	***
DBP	⟵	Exercise	0.547	0.125	4.39	***
DBP	⟵	Emotional_well-being	0.636	0.378	1.682	0.093
SBP	⟵	Heart_Rate	-0.09	0.016	-5.516	***
DBP	⟵	Heart_Rate	0.04	0.01	3.92	***
DBP	⟵	Smoker_Stat	1.312	0.343	3.827	***
DBP	⟵	SES	1.524	0.165	9.235	***
SBP	⟵	Alcohol	0.136	0.271	0.502	0.616
SBP	⟵	Emotional_well-being	0.336	0.62	0.543	0.587
SBP	⟵	SES	1.406	0.254	5.529	***
DBP	⟵	Avg_BMI	0.493	0.02	24.139	***
SBP	⟵	Avg_BMI	0.713	0.033	21.437	***
Empl_stat	⟵	SES	1			
Ln_inc	⟵	SES	2.448	0.074	33.299	***
Educ	⟵	SES	0.274	0.012	22.308	***
Ownrad_r	⟵	SES	0.206	0.006	32.873	***
Ownhif_r	⟵	SES	0.131	0.005	25.459	***

			Estimate	S.E.	C.R.	P
Owncel_r	⟵	SES	0.168	0.006	25.994	***
DBP_1	⟵	DBP	1			
DBP_2	⟵	DBP	1.007	0.008	121.335	***
SBP_1	⟵	SBP	1			
SBP_2	⟵	SBP	1.021	0.007	154.572	***
Emobth	⟵	Emotional_well-being	1			
Emomnd	⟵	Emotional_well-being	1.176	0.026	45.014	***
Emodep	⟵	Emotional_well-being	1.425	0.036	39.48	***
Emofear	⟵	Emotional_well-being	1.095	0.033	32.89	***
Emoslp	⟵	Emotional_well-being	1.168	0.034	34.313	***
Emolone	⟵	Emotional_well-being	1.188	0.035	33.873	***
Emogo	⟵	Emotional_well-being	1.088	0.033	32.981	***
HR_1	⟵	Heart_Rate	0.938	0.028	33.208	***
HR_2	⟵	Heart_Rate	1			

Table I.9: Standardised regression coefficients for SEM model 2.1 for females

			Estimate
Exercise	⟵	SES	0.062
Emotional_well-being	⟵	SES	-0.024
Smoker_Stat	⟵	SES	0.039
Alcohol	⟵	SES	0.06
Heart_Rate	⟵	SES	-0.052
Heart_Rate	⟵	Smoker_Stat	0.039
Heart_Rate	⟵	Exercise	-0.029
Avg_BMI	⟵	SES	0.309
Avg_BMI	⟵	Exercise	-0.037
Avg_BMI	⟵	Smoker_Stat	0.003
Avg_BMI	⟵	Alcohol	-0.059
Avg_BMI	⟵	Emotional_well-being	0.027
SBP	⟵	Smoker_Stat	0.078
DBP	⟵	Alcohol	0.038
SBP	⟵	Exercise	0.071

			Estimate
DBP	⟵	Exercise	0.047
DBP	⟵	Emotional_well-being	0.02
SBP	⟵	Heart_Rate	-0.064
DBP	⟵	Heart_Rate	0.041
DBP	⟵	Smoker_Stat	0.047
DBP	⟵	SES	0.116
SBP	⟵	Alcohol	0.006
SBP	⟵	Emotional_well-being	0.007
SBP	⟵	SES	0.074
DBP	⟵	Avg_BMI	0.273
SBP	⟵	Avg_BMI	0.273
Empl_stat	⟵	SES	0.667
Ln_inc	⟵	SES	0.677
Educ	⟵	SES	0.336
Ownrad_r	⟵	SES	0.391
Ownhif_r	⟵	SES	0.312
Owncel_r	⟵	SES	0.387
DBP_1	⟵	DBP	0.914
DBP_2	⟵	DBP	0.95
SBP_1	⟵	SBP	0.929
SBP_2	⟵	SBP	0.974
Emobth	⟵	Emotional_well-being	0.568
Emomnd	⟵	Emotional_well-being	0.621
Emodep	⟵	Emotional_well-being	0.67
Emofear	⟵	Emotional_well-being	0.578
Emoslp	⟵	Emotional_well-being	0.558
Emolone	⟵	Emotional_well-being	0.597
Emogo	⟵	Emotional_well-being	0.596
HR_1	⟵	Heart_Rate	0.919
HR_2	⟵	Heart_Rate	0.982

Table I.10: Squared multiple correlations for SEM model 2.1 for females

	Estimate
Exercise	0.004
Alcohol	0.004
Smoker_Stat	0.002
Emotional_well-being	0.001
Avg_BMI	0.097
Heart_Rate	0.005
SBP	0.109
DBP	0.116
HR_2	0.964
HR_1	0.845
Emogo	0.355
Emolone	0.356
Emoslp	0.311
Emofear	0.334
Emodep	0.448
Emomnd	0.386
Emobth	0.323
SBP_2	0.949
SBP_1	0.863
DBP_2	0.903
DBP_1	0.836
Owncel_r	0.15
Ownhif_r	0.097
Ownrad_r	0.153
Educ	0.113
Ln_inc	0.458
Empl_stat	0.445

Table I.11: Standardised indirect effects for SEM model 2.1 for females

	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate
Avg_BMI	-0.006	0	0	0	0	0	0
Heart_Rate	0	0	0	0	0	0	0
SBP	0.094	-0.008	-0.016	-0.002	0.007	0	0
DBP	0.087	-0.011	-0.016	0.002	0.007	0	0
HR_2	-0.051	-0.028	0	0.038	0	0	0
HR_1	-0.048	-0.027	0	0.036	0	0	0
Emogo	-0.015	0	0	0	0	0	0
Emolone	-0.015	0	0	0	0	0	0
Emoslp	-0.014	0	0	0	0	0	0
Emofear	-0.014	0	0	0	0	0	0
Emodep	-0.016	0	0	0	0	0	0
Emomnd	-0.015	0	0	0	0	0	0
Emobth	-0.014	0	0	0	0	0	0
SBP_2	0.163	0.061	-0.01	0.074	0.014	0.266	-0.062
SBP_1	0.156	0.058	-0.01	0.071	0.014	0.254	-0.059
DBP_2	0.193	0.034	0.02	0.047	0.026	0.26	0.039
DBP_1	0.186	0.032	0.019	0.045	0.025	0.25	0.037

Table I.12: Significance of standardised indirect effects for SEM model 2.1 for females

	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate
Avg_BMI	0.002	...	...	...	...	...	...
Heart_Rate	0.787	...	...	...	...	...	...
SBP	0.004	0.014	0.003	0.815	0.04	...	...
DBP	0.005	0.002	0.003	0.602	0.039	...	...
HR_2	0.004	0.003	...	0.005	...	...	...
HR_1	0.004	0.003	...	0.004	...	...	...
Emogo	0.22	...	...	...	...	...	...
Emolone	0.213	...	...	...	...	...	...
Emoslp	0.22	...	...	...	...	...	...
Emofear	0.22	...	...	...	...	...	...
Emodep	0.22	...	...	...	...	...	...
Emomnd	0.22	...	...	...	...	...	...
Emobth	0.209	...	...	...	...	...	...
SBP_2	0.005	0.005	0.429	0.009	0.43	0.007	0.008
SBP_1	0.005	0.005	0.435	0.009	0.43	0.008	0.008
DBP_2	0.005	0.011	0.098	0.006	0.068	0.009	0.005
DBP_1	0.006	0.01	0.096	0.006	0.069	0.008	0.005

Table I.13: Specific indirect effects for SEM model 2.1 for females

Parameter	Standardised estimate	Estimate	Lower	Upper	P	Mediation
SES - BMI - SBP	0.0844	1.603	1.404	1.811	0.005	Yes
SES - BMI - DBP	0.0844	1.108	0.974	1.236	0.007	Yes
SES - Alcohol - SBP	0.0004	0.006	-0.018	0.034	0.634	No
SES - Alcohol - DBP	0.0023	0.030	0.014	0.053	0.005	Yes
SES - Smoker Stat - SBP	0.0030	0.058	0.027	0.101	0.008	Yes
SES - Smoker Stat - DBP	0.0018	0.024	0.01	0.044	0.008	Yes
SES - Exercise - SBP	0.0044	0.083	0.048	0.159	0.001	Yes
SES - Exercise - DBP	0.0029	0.038	0.02	0.071	0.001	Yes
SES - HR - SBP	0.0033	0.063	0.031	0.1	0.005	Yes
SES - HR - DBP	-0.0021	-0.028	-0.052	-0.011	0.006	Yes
SES - Emotional well-being - SBP	-0.0002	-0.003	-0.03	0.009	0.528	No
SES - Emotional well-being - DBP	-0.0005	-0.006	-0.028	0.001	0.195	No

Table I.14: Standardised total effects for SEM model 2.1 for females

	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate	SBP	DBP
Exercise	0.062	0	0	0	0	0	0	0	0
Alcohol	0.06	0	0	0	0	0	0	0	0
Smoker_Stat	0.039	0	0	0	0	0	0	0	0
Emotional_well-being	-0.024	0	0	0	0	0	0	0	0
Avg_BMI	0.302	-0.037	-0.059	0.003	0.027	0	0	0	0
Heart_Rate	-0.052	-0.029	0	0.039	0	0	0	0	0
SBP	0.168	0.063	-0.011	0.076	0.015	0.273	-0.064	0	0
DBP	0.203	0.035	0.021	0.049	0.027	0.273	0.041	0	0
HR_2	-0.051	-0.028	0	0.038	0	0	0.982	0	0
HR_1	-0.048	-0.027	0	0.036	0	0	0.919	0	0
Emogo	-0.015	0	0	0	0.596	0	0	0	0
Emolone	-0.015	0	0	0	0.597	0	0	0	0
Emoslp	-0.014	0	0	0	0.558	0	0	0	0
Emofear	-0.014	0	0	0	0.578	0	0	0	0
Emodep	-0.016	0	0	0	0.67	0	0	0	0
Emomnd	-0.015	0	0	0	0.621	0	0	0	0
Emobth	-0.014	0	0	0	0.568	0	0	0	0
SBP_2	0.163	0.061	-0.01	0.074	0.014	0.266	-0.062	0.974	0

	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate	SBP	DBP
SBP_1	0.156	0.058	-0.01	0.071	0.014	0.254	-0.059	0.929	0
DBP_2	0.193	0.034	0.02	0.047	0.026	0.26	0.039	0	0.95
DBP_1	0.186	0.032	0.019	0.045	0.025	0.25	0.037	0	0.914
Owncel_r	0.387	0	0	0	0	0	0	0	0
Ownhif_r	0.312	0	0	0	0	0	0	0	0
Ownrad_r	0.391	0	0	0	0	0	0	0	0
Educ	0.336	0	0	0	0	0	0	0	0
Ln_inc	0.677	0	0	0	0	0	0	0	0
Empl_stat	0.667	0	0	0	0	0	0	0	0

Table I.15: Unstandardised regression coefficients for SEM model 2.2 for males

			Estimate	S.E.	C.R.	P
Exercise	⟵	SES	-0.167	0.017	-9.737	***
Emotional_well-being	⟵	SES	0.011	0.004	2.664	0.008
Smoker_Stat	⟵	SES	0.236	0.008	27.945	***
Alcohol	⟵	SES	0.304	0.013	23.502	***
Heart_Rate	⟵	SES	0.668	0.131	5.079	***
Heart_Rate	⟵	Smoker_Stat	0.386	0.195	1.979	0.048
Heart_Rate	⟵	Exercise	-0.609	0.079	-7.671	***
Avg_BMI	⟵	SES	1.738	0.054	32.231	***
Avg_BMI	⟵	Exercise	-0.132	0.027	-4.825	***
Avg_BMI	⟵	Smoker_Stat	-0.704	0.08	-8.84	***
Avg_BMI	⟵	Alcohol	-0.201	0.041	-4.95	***
Avg_BMI	⟵	Emotional_well-being	-0.653	0.166	-3.945	***
SBP	⟵	Smoker_Stat	0.387	0.321	1.208	0.227
DBP	⟵	Alcohol	0.388	0.108	3.585	***
SBP	⟵	Exercise	0.653	0.109	5.972	***
DBP	⟵	Exercise	0.015	0.077	0.19	0.849
DBP	⟵	Emotional_well-being	0.896	0.445	2.013	0.044
SBP	⟵	Heart_Rate	0.011	0.019	0.594	0.553
DBP	⟵	Heart_Rate	0.132	0.013	9.972	***
DBP	⟵	Smoker_Stat	-0.177	0.206	-0.857	0.391
DBP	⟵	SES	0.453	0.28	1.619	0.105
SBP	⟵	Alcohol	0.657	0.165	3.991	***
SBP	⟵	Emotional_well-being	1.103	0.7	1.576	0.115
SBP	⟵	SES	0.092	0.437	0.211	0.833
DBP	⟵	Avg_BMI	0.445	0.04	11.102	***
SBP	⟵	Avg_BMI	0.508	0.064	7.894	***
DBP	⟵	Hlbp_med	-5.399	1.253	-4.309	***
SBP	⟵	Hlbp_med	-9.667	2.129	-4.541	***
SBP	⟵	Age	0.434	0.036	11.952	***
DBP	⟵	Age	0.254	0.022	11.357	***
Empl_stat	⟵	SES	1			

			Estimate	S.E.	C.R.	P
Ln_inc	⟵	SES	2.652	0.044	60.812	***
Educ	⟵	SES	0.031	0.007	4.428	***
Ownrad_r	⟵	SES	0.241	0.005	49.593	***
Ownhif_r	⟵	SES	0.149	0.004	35.602	***
Owncel_r	⟵	SES	0.119	0.005	26.355	***
DBP_1	⟵	DBP	1			
DBP_2	⟵	DBP	1.041	0.01	108.483	***
SBP_1	⟵	SBP	1			
SBP_2	⟵	SBP	1.011	0.008	128.725	***
Emobth	⟵	Emotional_well-being	1			
Emomnd	⟵	Emotional_well-being	1.25	0.036	34.466	***
Emodep	⟵	Emotional_well-being	1.348	0.046	29.484	***
Emofear	⟵	Emotional_well-being	1.088	0.042	26.078	***
Emoslp	⟵	Emotional_well-being	1.192	0.044	27.244	***
Emolone	⟵	Emotional_well-being	1.114	0.044	25.319	***
Emogo	⟵	Emotional_well-being	1.012	0.04	25.127	***
HR_1	⟵	Heart_Rate	1.062	0.027	39.894	***
HR_2	⟵	Heart_Rate	1			

Table I.16: Standardised regression coefficients for SEM model 2.2 for males

			Estimate
Exercise	⟵	SES	-0.116
Emotional_well-being	⟵	SES	0.036
Smoker_Stat	⟵	SES	0.33
Alcohol	⟵	SES	0.258
Heart_Rate	⟵	SES	0.067
Heart_Rate	⟵	Smoker_Stat	0.028
Heart_Rate	⟵	Exercise	-0.088
Avg_BMI	⟵	SES	0.459
Avg_BMI	⟵	Exercise	-0.05
Avg_BMI	⟵	Smoker_Stat	-0.133
Avg_BMI	⟵	Alcohol	-0.062

			Estimate
Avg_BMI	⟵	Emotional_well-being	-0.053
SBP	⟵	Smoker_Stat	0.019
DBP	⟵	Alcohol	0.044
SBP	⟵	Exercise	0.065
DBP	⟵	Exercise	0.002
DBP	⟵	Emotional_well-being	0.027
SBP	⟵	Heart_Rate	0.008
DBP	⟵	Heart_Rate	0.127
DBP	⟵	Smoker_Stat	-0.012
DBP	⟵	SES	0.044
SBP	⟵	Alcohol	0.054
SBP	⟵	Emotional_well-being	0.023
SBP	⟵	SES	0.006
DBP	⟵	Avg_BMI	0.164
SBP	⟵	Avg_BMI	0.134
DBP	⟵	Hlbp_med	-0.083
SBP	⟵	Hlbp_med	-0.105
SBP	⟵	Age	0.338
DBP	⟵	Age	0.277
Empl_stat	⟵	SES	0.837
Ln_inc	⟵	SES	0.787
Educ	⟵	SES	0.062
Ownrad_r	⟵	SES	0.568
Ownhif_r	⟵	SES	0.41
Owncel_r	⟵	SES	0.33
DBP_1	⟵	DBP	0.913
DBP_2	⟵	DBP	0.953
SBP_1	⟵	SBP	0.931
SBP_2	⟵	SBP	0.971
Emobth	⟵	Emotional_well-being	0.564
Emomnd	⟵	Emotional_well-being	0.652
Emodep	⟵	Emotional_well-being	0.634

			Estimate
Emofear	⟵	Emotional_well-being	0.601
Emoslp	⟵	Emotional_well-being	0.578
Emolone	⟵	Emotional_well-being	0.553
Emogo	⟵	Emotional_well-being	0.566
HR_1	⟵	Heart_Rate	0.979
HR_2	⟵	Heart_Rate	0.927

Table I.17: Squared multiple correlations for SEM model 2.2 for males

	Estimate
Exercise	0.014
Alcohol	0.067
Smoker_Stat	0.109
Emotional_well-being	0.001
Avg_BMI	0.192
Heart_Rate	0.016
SBP	0.21
DBP	0.208
HR_2	0.859
HR_1	0.959
Emogo	0.321
Emolone	0.306
Emoslp	0.334
Emofear	0.361
Emodep	0.402
Emomnd	0.425
Emobth	0.318
SBP_2	0.943
SBP_1	0.868
DBP_2	0.908
DBP_1	0.833
Owncel_r	0.109
Ownhif_r	0.168

	Estimate
Ownrad_r	0.323
Educ	0.004
Ln_inc	0.619
Empl_stat	0.701

Table I.18: Standardised indirect effects for SEM model 2.2 for males

	Age	Hlbp_med	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate
Avg_BMI	0	0	-0.056	0	0	0	0	0	0
Heart_Rate	0	0	0.019	0	0	0	0	0	0
SBP	0	0	0.068	-0.007	-0.008	-0.018	-0.007	0	0
DBP	0	0	0.085	-0.019	-0.01	-0.018	-0.009	0	0
HR_2	0	0	0.08	-0.081	0	0.026	0	0	0
HR_1	0	0	0.085	-0.086	0	0.027	0	0	0
Emogo	0	0	0.02	0	0	0	0	0	0
Emolone	0	0	0.02	0	0	0	0	0	0
Emoslp	0	0	0.021	0	0	0	0	0	0
Emofear	0	0	0.022	0	0	0	0	0	0
Emodep	0	0	0.023	0	0	0	0	0	0
Emomnd	0	0	0.024	0	0	0	0	0	0
Emobth	0	0	0.02	0	0	0	0	0	0
SBP_2	0.328	-0.102	0.072	0.056	0.044	0.002	0.016	0.13	0.008
SBP_1	0.315	-0.098	0.069	0.054	0.042	0.002	0.015	0.124	0.007
DBP_2	0.264	-0.079	0.123	-0.017	0.033	-0.029	0.017	0.156	0.121
DBP_1	0.253	-0.075	0.118	-0.016	0.031	-0.028	0.016	0.15	0.116

Table I.19: Significance of standardised indirect effects for SEM model 2.2 for males

	Age	Hlbp_med	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate
Avg_BMI	...	...	0.011	...	...	...	...	...	...
Heart_Rate	...	...	0.019	...	...	...	...	...	...
SBP	...	...	0.006	0.003	0.004	0.005	0.005	...	...
DBP	...	...	0.005	0.001	0.005	0.002	0.005	...	...
HR_2	...	...	0.005	0.002	...	0.198	...	...	...
HR_1	...	...	0.005	0.003	...	0.195	...	...	...
Emogo	...	...	0.093	...	...	...	...	...	...
Emolone	...	...	0.107	...	...	...	...	...	...
Emoslp	...	...	0.112	...	...	...	...	...	...
Emofear	...	...	0.109	...	...	...	...	...	...
Emodep	...	...	0.099	...	...	...	...	...	...
Emomnd	...	...	0.101	...	...	...	...	...	...
Emobth	...	...	0.103	...	...	...	...	...	...
SBP_2	0.006	0.003	0.023	0.005	0.012	0.927	0.267	0.008	0.513
SBP_1	0.007	0.003	0.024	0.005	0.011	0.927	0.285	0.008	0.519
DBP_2	0.004	0.003	0.006	0.166	0.063	0.079	0.15	0.003	0.002
DBP_1	0.006	0.003	0.006	0.16	0.06	0.072	0.153	0.004	0.001

Table I.20: Specific indirect effects for SEM model 2.2 for males

Parameter	Standardised estimate	Estimate	Lower	Upper	P	Mediation
SES - BMI - SBP	0.0615	0.883	0.633	1.086	0.007	Yes
SES - BMI - DBP	0.0753	0.773	0.646	0.933	0.003	Yes
SES - Alcohol - SBP	0.0139	0.2	0.087	0.293	0.006	Yes
SES - Alcohol - DBP	0.0114	0.118	0.05	0.183	0.008	Yes
SES - Smoker Stat - SBP	0.0063	0.091	-0.052	0.232	0.279	No
SES - Smoker Stat - DBP	-0.0040	-0.042	-0.133	0.056	0.462	No
SES - Exercise - SBP	-0.0075	-0.109	-0.159	-0.065	0.004	Yes
SES - Exercise - DBP	-0.0002	-0.002	-0.03	0.026	0.873	No
SES - HR - SBP	0.0005	0.008	-0.014	0.039	0.426	No
SES - HR - DBP	0.0085	0.088	0.053	0.136	0.001	Yes
SES - Emotional well-being - SBP	0.0008	0.012	-0.001	0.043	0.157	No
SES - Emotional well-being - DBP	0.0010	0.01	0.001	0.03	0.049	Yes

Table I.21: Standardised total effects for SEM model 2.2 for males

	Age	Hlbp_med	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate	SBP	DBP
Exercise	0	0	-0.116	0	0	0	0	0	0	0	0
Alcohol	0	0	0.258	0	0	0	0	0	0	0	0
Smoker_Stat	0	0	0.33	0	0	0	0	0	0	0	0
Emotional_well-being	0	0	0.036	0	0	0	0	0	0	0	0
Avg_BMI	0	0	0.403	-0.05	-0.062	-0.133	-0.053	0	0	0	0
Heart_Rate	0	0	0.086	-0.088	0	0.028	0	0	0	0	0
SBP	0.338	-0.105	0.074	0.058	0.045	0.002	0.016	0.134	0.008	0	0
DBP	0.277	-0.083	0.129	-0.017	0.034	-0.031	0.018	0.164	0.127	0	0
HR_2	0	0	0.08	-0.081	0	0.026	0	0	0.927	0	0
HR_1	0	0	0.085	-0.086	0	0.027	0	0	0.979	0	0
Emogo	0	0	0.02	0	0	0	0.566	0	0	0	0
Emolone	0	0	0.02	0	0	0	0.553	0	0	0	0
Emoslp	0	0	0.021	0	0	0	0.578	0	0	0	0
Emofear	0	0	0.022	0	0	0	0.601	0	0	0	0
Emodep	0	0	0.023	0	0	0	0.634	0	0	0	0
Emomnd	0	0	0.024	0	0	0	0.652	0	0	0	0
Emobth	0	0	0.02	0	0	0	0.564	0	0	0	0
SBP_2	0.328	-0.102	0.072	0.056	0.044	0.002	0.016	0.13	0.008	0.971	0

	Age	Hlbp_med	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate	SBP	DBP
SBP_1	0.315	-0.098	0.069	0.054	0.042	0.002	0.015	0.124	0.007	0.931	0
DBP_2	0.264	-0.079	0.123	-0.017	0.033	-0.029	0.017	0.156	0.121	0	0.953
DBP_1	0.253	-0.075	0.118	-0.016	0.031	-0.028	0.016	0.15	0.116	0	0.913
Owncel_r	0	0	0.33	0	0	0	0	0	0	0	0
Ownhif_r	0	0	0.41	0	0	0	0	0	0	0	0
Ownrad_r	0	0	0.568	0	0	0	0	0	0	0	0
Educ	0	0	0.062	0	0	0	0	0	0	0	0
Ln_inc	0	0	0.787	0	0	0	0	0	0	0	0
Empl_stat	0	0	0.837	0	0	0	0	0	0	0	0

Table I.22: Unstandardised regression coefficients for SEM model 2.2 for females

			Estimate	S.E.	C.R.	P
Exercise	⟵	SES	-0.065	0.016	-4.018	***
Emotional_well-being	⟵	SES	0.044	0.006	7.128	***
Smoker_Stat	⟵	SES	0.087	0.006	14.641	***
Alcohol	⟵	SES	0.047	0.009	5.027	***
Heart_Rate	⟵	SES	-1.295	0.177	-7.32	***
Heart_Rate	⟵	Smoker_Stat	0.636	0.377	1.689	0.091
Heart_Rate	⟵	Exercise	-0.422	0.115	-3.676	***
Avg_BMI	⟵	SES	4.25	0.109	39.038	***
Avg_BMI	⟵	Exercise	-0.433	0.058	-7.484	***
Avg_BMI	⟵	Smoker_Stat	-0.834	0.209	-3.981	***
Avg_BMI	⟵	Alcohol	-0.447	0.101	-4.44	***
Avg_BMI	⟵	Emotional_well-being	0.009	0.19	0.049	0.961
SBP	⟵	Smoker_Stat	1.251	0.61	2.05	0.04
DBP	⟵	Alcohol	0.423	0.184	2.292	0.022
SBP	⟵	Exercise	0.474	0.161	2.949	0.003
DBP	⟵	Exercise	0.156	0.109	1.427	0.154
DBP	⟵	Emotional_well-being	0.222	0.362	0.614	0.54
SBP	⟵	Heart_Rate	-0.069	0.015	-4.682	***
DBP	⟵	Heart_Rate	0.043	0.01	4.361	***
DBP	⟵	Smoker_Stat	0.192	0.367	0.522	0.602
DBP	⟵	SES	1.967	0.434	4.536	***
SBP	⟵	Alcohol	0.079	0.252	0.315	0.753
SBP	⟵	Emotional_well-being	-0.146	0.558	-0.261	0.794
SBP	⟵	SES	-0.417	0.658	-0.633	0.527
DBP	⟵	Avg_BMI	0.338	0.025	13.664	***
SBP	⟵	Avg_BMI	0.437	0.038	11.543	***
SBP	⟵	Hlbp_med	-12.275	0.949	-12.938	***
DBP	⟵	Hlbp_med	-6.446	0.563	-11.456	***
DBP	⟵	Age	0.186	0.02	9.46	***
SBP	⟵	Age	0.509	0.031	16.307	***
Empl_stat	⟵	SES	1			

			Estimate	S.E.	C.R.	P
Ln_inc	⟵	SES	3.295	0.069	47.762	***
Educ	⟵	SES	-0.165	0.012	-13.994	***
Ownrad_r	⟵	SES	0.319	0.007	42.865	***
Ownhif_r	⟵	SES	0.176	0.005	32.284	***
Owncel_r	⟵	SES	0.146	0.006	23.095	***
DBP_1	⟵	DBP	1			
DBP_2	⟵	DBP	1.007	0.007	140.034	***
SBP_1	⟵	SBP	1			
SBP_2	⟵	SBP	1.011	0.005	190.107	***
Emobth	⟵	Emotional_well-being	1			
Emomnd	⟵	Emotional_well-being	1.176	0.026	45.548	***
Emodep	⟵	Emotional_well-being	1.396	0.035	39.916	***
Emofear	⟵	Emotional_well-being	1.087	0.033	32.984	***
Emoslp	⟵	Emotional_well-being	1.153	0.033	34.619	***
Emolone	⟵	Emotional_well-being	1.178	0.034	34.232	***
Emogo	⟵	Emotional_well-being	1.048	0.032	32.99	***
HR_1	⟵	Heart_Rate	0.94	0.03	31.56	***
HR_2	⟵	Heart_Rate	1			

Table I.23: Standardised regression coefficients for SEM model 2.2 for females

			Estimate
Exercise	⟵	SES	-0.045
Emotional_well-being	⟵	SES	0.085
Smoker_Stat	⟵	SES	0.157
Alcohol	⟵	SES	0.05
Heart_Rate	⟵	SES	-0.077
Heart_Rate	⟵	Smoker_Stat	0.021
Heart_Rate	⟵	Exercise	-0.037
Avg_BMI	⟵	SES	0.481
Avg_BMI	⟵	Exercise	-0.071
Avg_BMI	⟵	Smoker_Stat	-0.052
Avg_BMI	⟵	Alcohol	-0.048

			Estimate
Avg_BMI	⟵	Emotional_well-being	0.001
SBP	⟵	Smoker_Stat	0.027
DBP	⟵	Alcohol	0.023
SBP	⟵	Exercise	0.026
DBP	⟵	Exercise	0.013
DBP	⟵	Emotional_well-being	0.007
SBP	⟵	Heart_Rate	-0.045
DBP	⟵	Heart_Rate	0.042
DBP	⟵	Smoker_Stat	0.006
DBP	⟵	SES	0.114
SBP	⟵	Alcohol	0.003
SBP	⟵	Emotional_well-being	-0.003
SBP	⟵	SES	-0.016
DBP	⟵	Avg_BMI	0.174
SBP	⟵	Avg_BMI	0.149
SBP	⟵	Hlbp_med	-0.201
DBP	⟵	Hlbp_med	-0.16
DBP	⟵	Age	0.215
SBP	⟵	Age	0.387
Empl_stat	⟵	SES	0.554
Ln_inc	⟵	SES	0.72
Educ	⟵	SES	-0.199
Ownrad_r	⟵	SES	0.499
Ownhif_r	⟵	SES	0.353
Owncel_r	⟵	SES	0.282
DBP_1	⟵	DBP	0.923
DBP_2	⟵	DBP	0.957
SBP_1	⟵	SBP	0.945
SBP_2	⟵	SBP	0.978
Emobth	⟵	Emotional_well-being	0.576
Emomnd	⟵	Emotional_well-being	0.629
Emodep	⟵	Emotional_well-being	0.668

			Estimate
Emofear	⟵	Emotional_well-being	0.583
Emoslp	⟵	Emotional_well-being	0.563
Emolone	⟵	Emotional_well-being	0.598
Emogo	⟵	Emotional_well-being	0.59
HR_1	⟵	Heart_Rate	0.919
HR_2	⟵	Heart_Rate	0.979

Table I.24: Squared multiple correlations for SEM model 2.2 for females

	Estimate
Exercise	0.002
Alcohol	0.002
Smoker_Stat	0.025
Emotional_well-being	0.007
Avg_BMI	0.235
Heart_Rate	0.007
SBP	0.325
DBP	0.249
HR_2	0.959
HR_1	0.845
Emogo	0.348
Emolone	0.357
Emoslp	0.317
Emofear	0.34
Emodep	0.446
Emomnd	0.396
Emobth	0.332
SBP_2	0.957
SBP_1	0.893
DBP_2	0.915
DBP_1	0.852
Owncel_r	0.08
Ownhif_r	0.125

	Estimate
Ownrad_r	0.249
Educ	0.04
Ln_inc	0.518
Empl_stat	0.307

Table I.25: Standardised indirect effects for SEM model 2.2 for females

	Age	Hlbp_med	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate
Avg_BMI	0	0	-0.007	0	0	0	0	0	0
Heart_Rate	0	0	0.005	0	0	0	0	0	0
SBP	0	0	0.076	-0.009	-0.007	-0.009	0	0	0
DBP	0	0	0.081	-0.014	-0.008	-0.008	0	0	0
HR_2	0	0	-0.071	-0.036	0	0.021	0	0	0
HR_1	0	0	-0.067	-0.034	0	0.019	0	0	0
Emogo	0	0	0.05	0	0	0	0	0	0
Emolone	0	0	0.051	0	0	0	0	0	0
Emoslp	0	0	0.048	0	0	0	0	0	0
Emofear	0	0	0.05	0	0	0	0	0	0
Emodep	0	0	0.057	0	0	0	0	0	0
Emomnd	0	0	0.054	0	0	0	0	0	0
Emobth	0	0	0.049	0	0	0	0	0	0
SBP_2	0.379	-0.197	0.059	0.017	-0.004	0.018	-0.003	0.145	-0.044
SBP_1	0.366	-0.19	0.057	0.017	-0.004	0.017	-0.003	0.14	-0.042
DBP_2	0.205	-0.153	0.187	-0.001	0.014	-0.002	0.006	0.166	0.04
DBP_1	0.198	-0.148	0.181	-0.001	0.014	-0.002	0.006	0.16	0.038

Table I.26: Significance of standardised indirect effects for SEM model 2.2 for females

	Age	Hlbp_med	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate
Avg_BMI	...	...	0.042	...	...	...	...	...	...
Heart_Rate	...	...	0.102	...	...	...	...	...	...
SBP	...	...	0.006	0.006	0.004	0.004	0.85	...	...
DBP	...	...	0.008	0.004	0.004	0.01	0.85	...	...
HR_2	...	...	0.004	0.006	...	0.27	...	...	...
HR_1	...	...	0.004	0.006	...	0.249	...	...	...
Emogo	...	...	0.009	...	...	...	...	...	...
Emolone	...	...	0.01	...	...	...	...	...	...
Emoslp	...	...	0.01	...	...	...	...	...	...
Emofear	...	...	0.01	...	...	...	...	...	...
Emodep	...	...	0.01	...	...	...	...	...	...
Emomnd	...	...	0.009	...	...	...	...	...	...
Emobth	...	...	0.008	...	...	...	...	...	...
SBP_2	0.004	0.002	0.081	0.122	0.606	0.283	0.842	0.007	0.008
SBP_1	0.005	0.002	0.083	0.124	0.606	0.283	0.842	0.007	0.007
DBP_2	0.002	0.001	0.013	0.964	0.157	0.849	0.519	0.009	0.002
DBP_1	0.002	0.001	0.013	0.964	0.157	0.849	0.519	0.009	0.002

Table I.27: Specific indirect effects for SEM model 2.2 for females

Parameter	Standardised estimate	Estimate	Lower	Upper	P	Mediation
SES - BMI - SBP	0.0717	1.859	1.533	2.216	0.006	Yes
SES - BMI - DBP	0.0837	1.435	1.157	1.644	0.011	Yes
SES - Alcohol - SBP	0.0002	0.004	-0.020	0.030	0.754	No
SES - Alcohol - DBP	0.0012	0.020	0.004	0.045	0.038	Yes
SES - Smoker Stat - SBP	0.0042	0.109	-0.002	0.217	0.108	No
SES - Smoker Stat - DBP	0.0009	0.017	-0.050	0.078	0.767	No
SES - Exercise - SBP	-0.0012	-0.031	-0.062	-0.004	0.018	Yes
SES - Exercise - DBP	-0.0006	-0.010	-0.030	0.003	0.214	No
SES - HR - SBP	0.0035	0.090	0.041	0.139	0.007	Yes
SES - HR - DBP	-0.0032	-0.055	-0.093	-0.025	0.003	Yes
SES - Emotional well-being - SBP	-0.0003	-0.006	-0.063	0.038	0.838	No
SES - Emotional well-being - DBP	0.0006	0.010	-0.024	0.046	0.608	No

Table I.28: Standardised total effects for SEM model 2.2 for females

	Age	Hlbp_med	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate	SBP	DBP
Exercise	0	0	-0.045	0	0	0	0	0	0	0	0
Alcohol	0	0	0.05	0	0	0	0	0	0	0	0
Smoker_Stat	0	0	0.157	0	0	0	0	0	0	0	0
Emotional_well-being	0	0	0.085	0	0	0	0	0	0	0	0
Avg_BMI	0	0	0.473	-0.071	-0.048	-0.052	0.001	0	0	0	0
Heart_Rate	0	0	-0.072	-0.037	0	0.021	0	0	0	0	0
SBP	0.387	-0.201	0.06	0.017	-0.004	0.018	-0.003	0.149	-0.045	0	0
DBP	0.215	-0.16	0.196	-0.001	0.015	-0.002	0.007	0.174	0.042	0	0
HR_2	0	0	-0.071	-0.036	0	0.021	0	0	0.979	0	0
HR_1	0	0	-0.067	-0.034	0	0.019	0	0	0.919	0	0
Emogo	0	0	0.05	0	0	0	0.59	0	0	0	0
Emolone	0	0	0.051	0	0	0	0.598	0	0	0	0
Emoslp	0	0	0.048	0	0	0	0.563	0	0	0	0
Emofear	0	0	0.05	0	0	0	0.583	0	0	0	0
Emodep	0	0	0.057	0	0	0	0.668	0	0	0	0
Emomnd	0	0	0.054	0	0	0	0.629	0	0	0	0
Emobth	0	0	0.049	0	0	0	0.576	0	0	0	0
SBP_2	0.379	-0.197	0.059	0.017	-0.004	0.018	-0.003	0.145	-0.044	0.978	0

	Age	Hlbp_med	SES	Exercise	Alcohol	Smoker_Stat	Emotional_well-being	Avg_BMI	Heart_Rate	SBP	DBP
SBP_1	0.366	-0.19	0.057	0.017	-0.004	0.017	-0.003	0.14	-0.042	0.945	0
DBP_2	0.205	-0.153	0.187	-0.001	0.014	-0.002	0.006	0.166	0.04	0	0.957
DBP_1	0.198	-0.148	0.181	-0.001	0.014	-0.002	0.006	0.16	0.038	0	0.923
Owncel_r	0	0	0.282	0	0	0	0	0	0	0	0
Ownhif_r	0	0	0.353	0	0	0	0	0	0	0	0
Ownrad_r	0	0	0.499	0	0	0	0	0	0	0	0
Educ	0	0	-0.199	0	0	0	0	0	0	0	0
Ln_inc	0	0	0.72	0	0	0	0	0	0	0	0
Empl_stat	0	0	0.554	0	0	0	0	0	0	0	0

Appendix J

Multiple linear regression results

Table J.1: Dummy variables used in regression analysis

Dummy variables	**Label**	**Value**
Empl_stat_d1	Dummy variable indicating that Empl_stat = 0 ("Not Economically Active")	0 = False, 1 = True
Empl_stat_d2	Dummy variable indicating that Empl_stat = 1 ("Unemployed_Discouraged")	0 = False, 1 = True
Empl_stat_d3	Dummy variable indicating that Empl_stat = 2 ("Unemployed_Strict")	0 = False, 1 = True
Empl_stat_d4	Dummy variable indicating that Empl_stat = 3 ("Employed")	0 = False, 1 = True
Exercise_d1	Dummy variable indicating that Exercise = 1 ("Never")	0 = False, 1 = True
Exercise_d2	Dummy variable indicating that Exercise = 2 ("Less than once a week")	0 = False, 1 = True
Exercise_d3	Dummy variable indicating that Exercise = 3 ("Once a week")	0 = False, 1 = True
Exercise_d4	Dummy variable indicating that Exercise = 4 ("Twice a week")	0 = False, 1 = True
Exercise_d5	Dummy variable indicating that Exercise = 5 ("Three or more times a week")	0 = False, 1 = True
Alcohol_d1	Dummy variable indicating that Alcohol = 4 ("7 or more standard drinks")	0 = False, 1 = True
Alcohol_d2	Dummy variable indicating that Alcohol = 3 ("5 to 6 standard drinks")	0 = False, 1 = True

Dummy variables	Label	Value
Alcohol_d3	Dummy variable indicating that Alcohol = 2 ("3 or 4 standard drinks")	0 = False, 1 = True
Alcohol_d4	Dummy variable indicating that Alcohol = 1 ("1 or 2 standard drinks")	0 = False, 1 = True
Alcohol_d5	Dummy variable indicating that Alcohol = 0 ("Non-drinker")	0 = False, 1 = True
Educ_d1	Dummy variable indicating that Educ = 1 ("No schooling")	0 = False, 1 = True
Educ_d2	Dummy variable indicating that Educ = 2 ("Primary")	0 = False, 1 = True
Educ_d3	Dummy variable indicating that Educ = 3 ("Secondary")	0 = False, 1 = True
Educ_d4	Dummy variable indicating that Educ = 4 ("Tertiary")	0 = False, 1 = True
Smoker_Stat_d1	Dummy variable indicating that Smoker_Stat = 1 ("Non-smoker")	0 = False, 1 = True
Smoker_Stat_d2	Dummy variable indicating that Smoker_Stat = 2 ("Smoke 5 or less")	0 = False, 1 = True
Smoker_Stat_d3	Dummy variable indicating that Smoker_Stat = 3 ("Smoke between 6 and 10")	0 = False, 1 = True
Smoker_Stat_d4	Dummy variable indicating that Smoker_Stat = 4 ("Smoke more than 10")	0 = False, 1 = True
Ownrad_r_d2	Dummy variable indicating that Ownrad_r = 2 ("Yes")	0 = False, 1 = True
Ownhif_r_d2	Dummy variable indicating that Ownhif_r = 2 ("Yes")	0 = False, 1 = True
Owncel_r_d2	Dummy variable indicating that Owncel_r = 2 ("Yes")	0 = False, 1 = True
Hlbp_med_d1	Dummy variable indicating that Hlbp_med = 1 ("Yes")	0 = False, 1 = True

Table J.2: Independent variables used after stepwise regression

Model	**Independent variables**
1	Avg_BMI, Education, Ln_inc, Employment_Status, Ownrad_r_d2, Alcohol
2	Avg_BMI, Ln_inc, Smoker_Status, Education, Ownrad_r_d2, Emotional_well-being
3	Avg_BMI, Education, Ownrad_r_d2, Ln_inc, Alcohol, Emotional_well-being
4	Age, Avg_BMI, Hlbp_med_d1, Alcohol, Exercise, Employment_Status, Owncel_r_d2
5	Age, Avg_BMI, Hlbp_med_d1, Smoker_Status, Employment_Status
6	Age, Avg_BMI, Hlbp_med_d1, Alcohol
7	Avg_BMI, Heart_Rate, Education, Ln_inc, Employment_Status, Ownrad_r_d2, Alcohol
8	Avg_BMI, Ln_inc, Smoker_Status, Heart_Rate, Ownrad_r_d2, Education, Employment_Status, Emotional_well-being
9	Avg_BMI, Heart_Rate, Ownrad_r_d2, Alcohol, Education, Exercise, Ln_inc
10	Age, Avg_BMI, Heart_Rate, Education, Alcohol, Hlbp_med_d1
11	Age, Avg_BMI, Employment_Status, Heart_Rate, Hlbp_med_d1, Emotional_well-being, Smoker_Status
12	Age, Avg_BMI, Heart_Rate, Alcohol, Ownrad_r_d2, Exercise
13	Education, Avg_BMI, Ownrad_r_d2, Employment_Status, Ln_inc, Owncel_r_d2, Heart_Rate, Smoker_Status, Alcohol
14	Education, Avg_BMI, Ownrad_r_d2, Smoker_Status, Ln_inc, Employment_Status, Owncel_r_d2, Heart_Rate
15	Education, Avg_BMI, Ownrad_r_d2, Ln_inc, Emotional_well-being, Smoker_Status, Employment_Status, Heart_Rate
16	Age, Hlbp_med_d1, Avg_BMI, Education, Alcohol, Heart_Rate
17	Age, Hlbp_med_d1, Avg_BMI, Smoker_Status, Ownhif_r_d2, Education, Heart_Rate, Owncel_r_d2, Ln_inc
18	Age, Hlbp_med_d1, Avg_BMI, Education, Smoker_Status, Emotional_well-being, Ownrad_r_d2, Ownhif_r_d2
19	Avg_BMI, Education, Ownrad_r_d2, Heart_Rate, Ln_inc, Owncel_r_d2, Alcohol
20	Avg_BMI, Education, Smoker_Status, Ln_inc, Ownrad_r_d2, Heart_Rate
21	Avg_BMI, Education, Ln_inc, Ownrad_r_d2, Smoker_Status, Emotional_well-being
22	Age, Avg_BMI, Hlbp_med_d1, Heart_Rate, Education, Alcohol, Ownrad_r_d2

Model	Independent variables
23	Age, Avg_BMI, Hlbp_med_d1, Smoker_Status, Employment_Status, Heart_Rate, Education, Alcohol
24	Age, Avg_BMI, Hlbp_med_d1, Smoker_Status, Education, Ownrad_r_d2, Emotional_well-being, Ownhif_r_d2, Employment_Status

Appendix K

Fit diagnostics

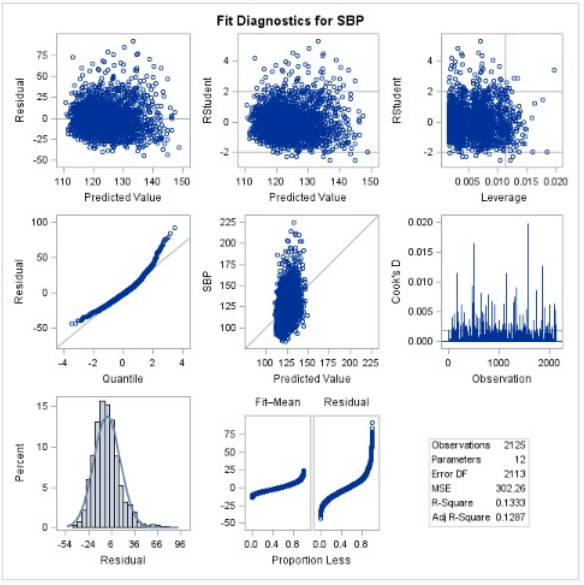


Figure K.1: Test dataset: Fit diagnostics for model 3

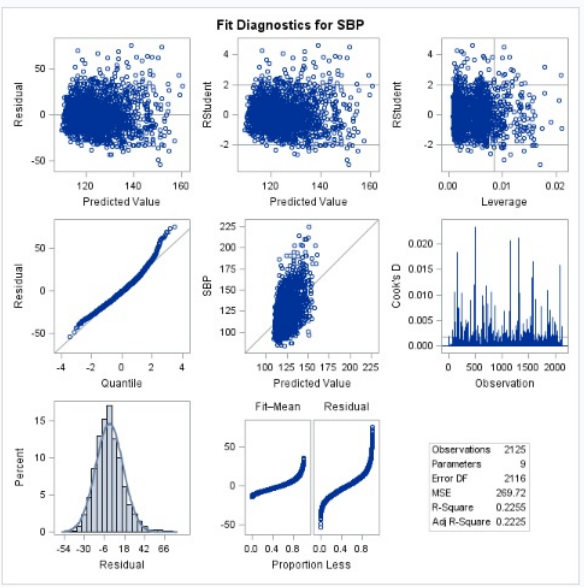


Figure K.2: Test dataset: Fit diagnostics for model 6

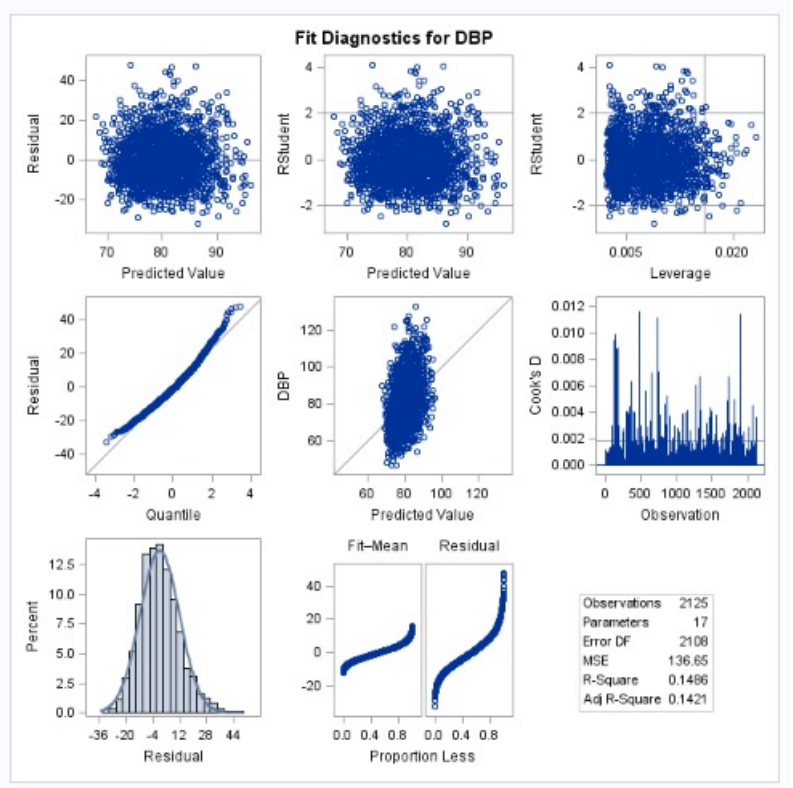


Figure K.3: Test dataset: Fit diagnostics for model 9

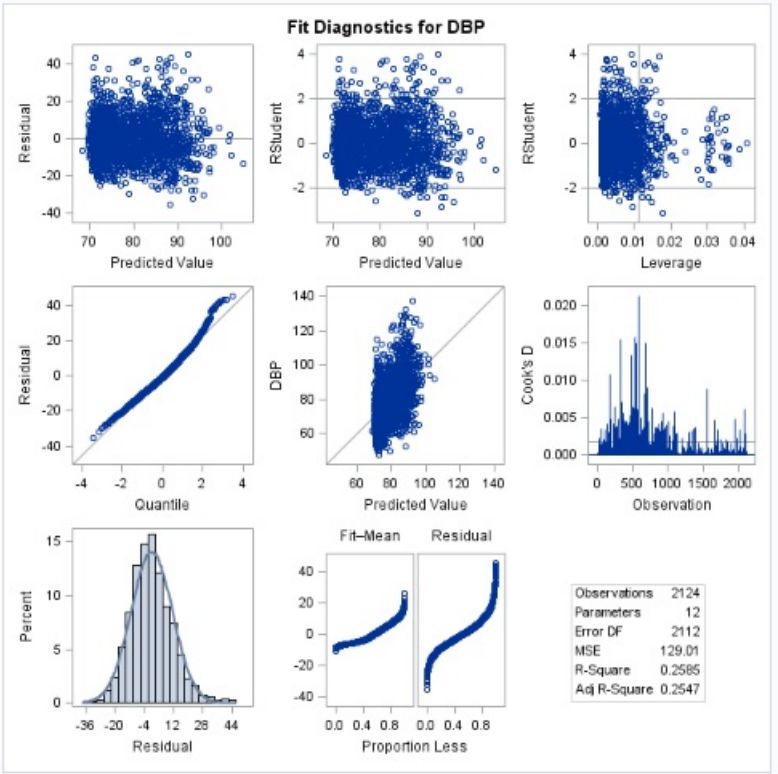


Figure K.4: Validation dataset: Fit diagnostics for model 11

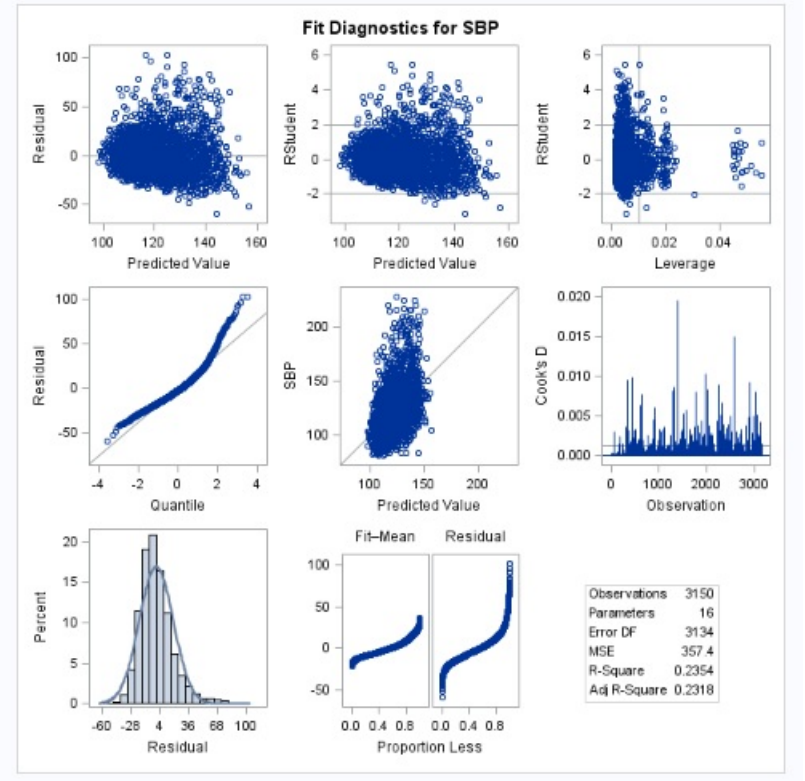


Figure K.5: Test dataset: Fit diagnostics for model 15

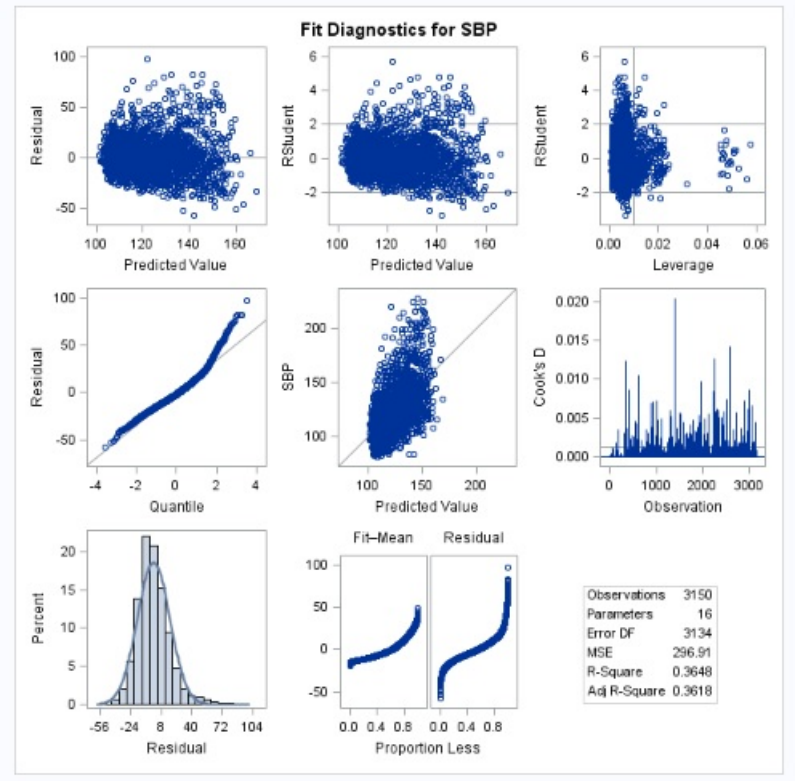


Figure K.6: Test dataset: Fit diagnostics for model 18

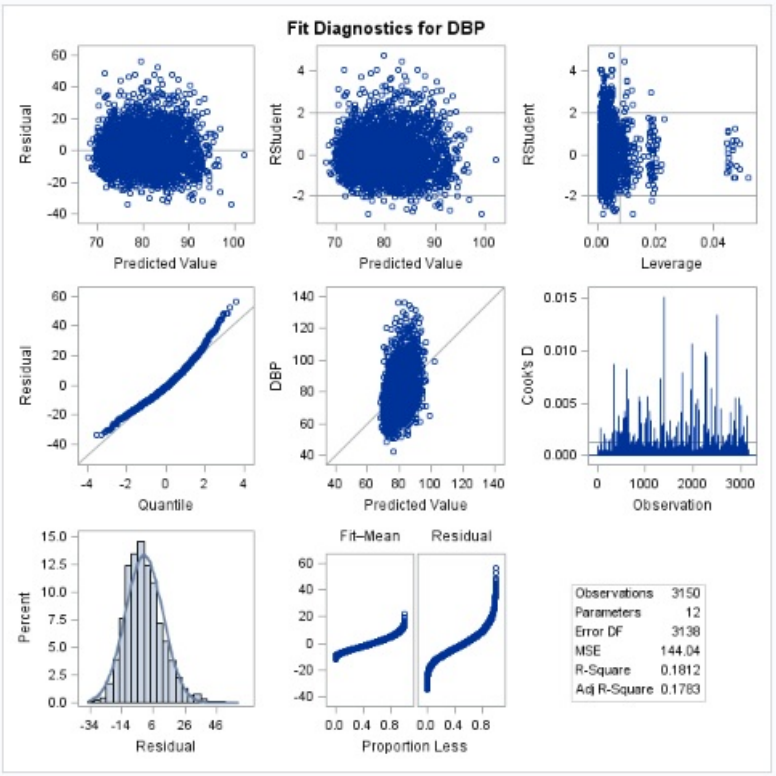


Figure K.7: Test dataset: Fit diagnostics for model 21

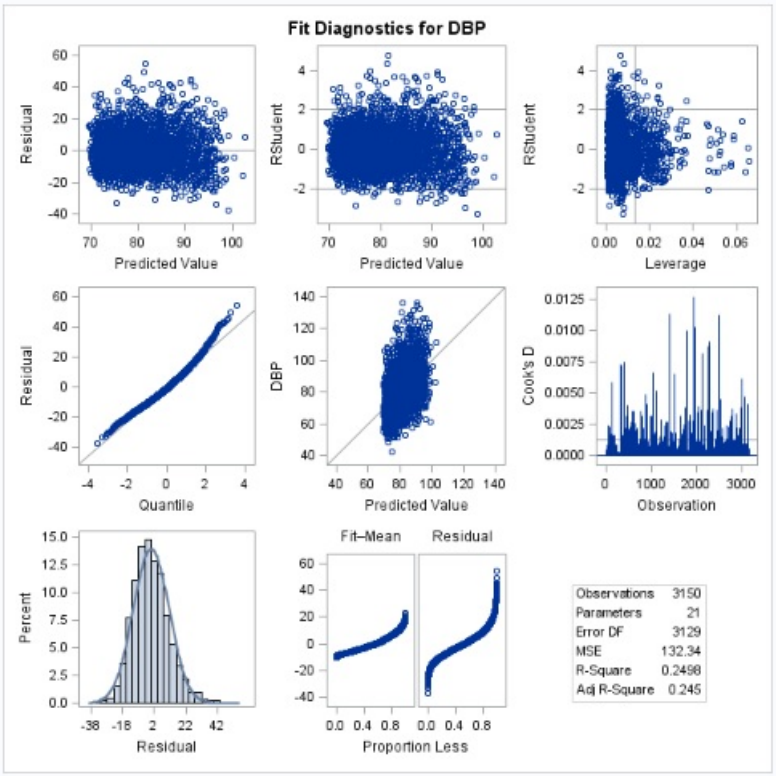


Figure K.8: Test dataset: Fit diagnostics for model 24

Appendix L

SAS code for multiple linear regression models

SAS code for model excluding age and the use of antihypertensive medication where SBP is the dependent variable:

```
proc import datafile="C:\Users\Luke Raquel\Desktop\Studies\Research\2016
\Structural equation modelling\Background to structural equation modelling\
Application\SAS\Males_MLR_v1.csv"
dbms=csv out=work.Males_MLR_v1 replace;
getnames=yes;
guessingrows=10000;
run;
ods graphics on;

proc reg data=Males_MLR_v1;
model SBP = Heart_Rate Emotional_well-being Ln_inc Avg_BMI Ownhif_r_d2
Owncel_r_d2 Ownrad_r_d2 {Smoker_Stat_d1 Smoker_Stat_d2 Smoker_Stat_d3
Smoker_Stat_d4} {Educ_d1 Educ_d2 Educ_d3 Educ_d4} {Alcohol_d1
Alcohol_d2 Alcohol_d3 Alcohol_d4 Alcohol_d5} {Exercise_d1
Exercise_d2 Exercise_d3 Exercise_d4 Exercise_d5} {Empl_stat_d1
Empl_stat_d2 Empl_stat_d3 Empl_stat_d4}
/SPEC WHITE R VIF TOL selection=stepwise SLE=0.05 SLS=0.05
groupnames='Heart_Rate' 'Emotional_well-being' 'Ln_inc' 'Avg_BMI'
'Ownhif_r_d2' 'Owncel_r_d2' 'Ownrad_r_d2' 'Smoker_Status' 'Education'
'Alcohol' 'Exercise' 'Employment_Status';
```

```
OUTPUT OUT=RESID
PREDICTED=pred
RESIDUAL=RESID
RSTUDENT=r1
DFFITS=dffits
COOKD=cookd
H=hatvalue
PRESS=res_del ;;
PLOT RESIDUAL.*PREDICTED. ;
PLOT RESIDUAL.*NQQ.;
RUN;

PROC UNIVARIATE
DATA=RESID
PLOT NORMAL;
VAR RESID;
RUN;
```

For each model, the same principle was applied however the dataset was changed in each iteration.

SAS code for model including age and the use of antihypertensive medication where SBP is the dependent variable:

```
proc import datafile="C:\Users\Luke Raquel\Desktop\Studies\Research\2016
\Structural equation modelling\Background to structural equation modelling\
Application\SAS\Males_MLR_v1.csv"
dbms=csv out=work.Males_MLR_v1 replace;
getnames=yes;
guessingrows=10000;
run;
```

```
ods graphics on;

proc reg data=Males_MLR_v1;
model SBP = Heart_Rate Emotional_well-being Ln_inc Avg_BMI Ownhif_r_d2
Owncel_r_d2 Ownrad_r_d2 {Smoker_Stat_d1 Smoker_Stat_d2 Smoker_Stat_d3
Smoker_Stat_d4} {Educ_d1 Educ_d2 Educ_d3 Educ_d4} {Alcohol_d1
Alcohol_d2 Alcohol_d3 Alcohol_d4 Alcohol_d5} {Exercise_d1
Exercise_d2 Exercise_d3 Exercise_d4 Exercise_d5} {Empl_stat_d1
Empl_stat_d2 Empl_stat_d3 Empl_stat_d4} Age Hlbp_med_d1
/SPEC WHITE R VIF TOL selection=stepwise SLE=0.05 SLS=0.05
groupnames='Heart_Rate' 'Emotional_well-being' 'Ln_inc' 'Avg_BMI'
'Ownhif_r_d2' 'Owncel_r_d2' 'Ownrad_r_d2' 'Smoker_Status' 'Education'
'Alcohol' 'Exercise' 'Employment_Status' 'Age' 'Hlbp_med_d1';
OUTPUT OUT=RESID
PREDICTED=pred
RESIDUAL=RESID
RSTUDENT=r1
DFFITS=dffits
COOKD=cookd
H=hatvalue
PRESS=res_del ;;
PLOT RESIDUAL.*PREDICTED. ;
PLOT RESIDUAL.*NQQ.;
RUN;

PROC UNIVARIATE
DATA=RESID
PLOT NORMAL;
VAR RESID;
RUN;
```

In the cases where DBP was the dependent variable, SBP was changed to DBP in the SAS code.