FOUND IN TRANSLATION:
Methods to Increase Meaning and Interpretability of Confound Variables

by

Ryan Geoffrey Nathaniel Seltzer

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2013
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Ryan Seltzer, titled Found in Translation: Methods to Increase Meaning and Interpretability of Confound Variables and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

______________________________ Date: 11/22/2013
Richard Bootzin

______________________________ Date: 11/22/2013
Lee Sechrest

______________________________ Date: 11/22/2013
W. Jake Jacobs

______________________________ Date: 11/22/2013
Aurelio José Figueredo

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

______________________________ Date: 11/22/2013
Dissertation Director: Richard Bootzin
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Ryan Seltzer
ACKNOWLEDGEMENTS

I wish to thank a number of people who helped me complete this dissertation and graduate degree.

Lee Sechrest provided invaluable and incomparable guidance when demonstrating how to think, write, and celebrate like a great methodologist. Watching him blend creative and scientific thinking was the pinnacle of my graduate education.

I also want to thank Richard Bootzin for graciously welcoming me into his lab and providing essential support as the chair of my many Ph.D. milestones.

Aurelio José Figueredo spent hours working with me in the trenches on various statistical analyses and always provided unique analytical and theoretical perspectives to scientific problems.

Thanks also to W. Jake Jacobs for giving me detailed feedback and frequent, encouraging nods of the head that intimated, “You are saying the smartest thing ever,” even if I knew better.

I want to thank my wife, Heather York, for her constant support and encouragement through this process and for staying up late, well after we put the kids to bed, to keep me company as we worked into the night.

I wish to thank my mother, Madeline Seltzer, my father, Alvin Seltzer, my sister, Lara Seltzer, and my brother, Jordan Seltzer, for their support and interest in my progress and goals.

Finally, graduate school is notorious for striking fear, anxiety, and intimidation into many a student. I, however, never fell victim to these emotions because I never once lost perspective of what was important in life. I owe this perspective to my three children, Jordan, Remy, and Shae Seltzer. All I ever had to do was look into their faces or hear their cries of joy when I arrived home to know where my heart will always truly lie. Thank you.
TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................................................. 7

LIST OF FIGURES ..................................................................................................................................................... 8

ABSTRACT .................................................................................................................................................................. 9

CHAPTER 1: INTRODUCTION ................................................................................................................................... 11

Confounds Translated ............................................................................................................................................. 13

Confounds in Use ...................................................................................................................................................... 15

Income in Health Research ........................................................................................................................................ 20

Income in Criminology Research ............................................................................................................................ 22

Income in Education Research .................................................................................................................................. 23

Education as a Proxy Variable ................................................................................................................................... 24

When Income Measures Income ........................................................................................................................... 27

Effects of Imperfect Proxy Variables .................................................................................................................... 29

CHAPTER 2: METHOD ............................................................................................................................................. 33

Data ........................................................................................................................................................................... 33

Measures ................................................................................................................................................................... 33

Analysis ...................................................................................................................................................................... 35

Confound Specification Analysis ........................................................................................................................... 36

Proxy Variable Analysis ........................................................................................................................................... 37

CHAPTER 3: RESULTS AND DISCUSSION .............................................................................................................. 39

Method 1: Confound Specification ........................................................................................................................ 39

Spurious Correlation ............................................................................................................................................... 39
# TABLE OF CONTENTS – Continued

Mediation ........................................................................................................................................ 41

Multiple Confounds .................................................................................................................... 42

Method 2: Proxy Variables ........................................................................................................ 46

Moderate Representation of Proxy Variable: Age .................................................................. 47

Poor Representation of Proxy Variable: Wealth ................................................................. 49

Good Representation of Proxy Variable: General Health .................................................. 51

Chapter 4: GENERAL DISCUSSION .............................................................................................. 55

Limitation ....................................................................................................................................... 58

APPENDIX A: TABLES AND FIGURES .......................................................................................... 61

APPENDIX B: PHYSICAL HEALTH QUESTIONS FROM SHARE SURVEY USED IN STRUCTURAL MODELS .................................................................................................................................................. 78

APPENDIX C: PHYSICAL HEALTH QUESTIONS FROM SHARE SURVEY USED IN STRUCTURAL MODELS .................................................................................................................................................. 81

APPENDIX D: ACTIVITIES OF DAILY LIVING QUESTIONS FROM SHARE SURVEY USED IN STRUCTURAL MODELS ................................................................................................................................. 82

APPENDIX E: QUESTIONS USED TO CONSTRUCT THE EURO-D SCALE ........................................ 84

APPENDIX F: PHYSICAL ACTIVITY QUESTIONS FROM SHARE SURVEY USED IN STRUCTURAL MODELS .................................................................................................................................................. 88

APPENDIX G: GLOSSARY OF TERMS.............................................................................................. 89
LIST OF TABLES

Table 1. Alphas of items for Constructs ................................................................. 61

Table 2. Correlations of proxy variables with constructs they are intended to represent .... 62

Table 3. Fit indices for the Age Reference Model and the Age Proxy Model .................. 63

Table 4. Fit indices for the Wealth Reference Model and the Wealth Proxy Model ............. 64

Table 5. Fit indices for the General Health Reference Model and the General Health Proxy Model ........................................................................................................ 65

Table 6. Summary of path coefficients ($\beta$) across the four different models ................. 66
LIST OF FIGURES

Figure 1. Design Model estimating the effect of physical inactivity on depression ..................... 67
Figure 2. Model specifying spurious relationship between physical inactivity and depression .. 68
Figure 3. Model specifying physical disability as a mediator of physical inactivity and depression ..................................................................................................................................... 69
Figure 4. Model specifying multiple confounds ............................................................................ 70
Figure 5. Reference Model ............................................................................................................ 71
Figure 6. Age Reference Model .................................................................................................... 72
Figure 7. Age Proxy Model ............................................................................................................ 73
Figure 8. Wealth Reference Model ............................................................................................... 74
Figure 9. Wealth Proxy Model ...................................................................................................... 75
Figure 10. General Health Reference Model ................................................................................ 76
Figure 11. General Health Proxy Model ........................................................................................ 77
ABSTRACT

The process of research is fraught with rote terminology that, when used blindly, can bend our methodological actions away from our theoretical intentions. This investigation is aimed at developing two methods for bringing meaning and interpretability to research when we work with confounds. I argue, with the first method, that granting confounds substantive influence in a network of related variables (rather than viewing confounds as nuisance variables) enhances the conceptual dimension with which phenomena can be explained. I evaluated models differing in how confounds were specified using data from the Survey of Health, Ageing and Retirement in Europe (SHARE). Generally, minor alterations to model specifications, such as direction of causal pathways, did not change model parameter estimates; however, the conceptual meaning of how the confounds interacted with other variables in the model changed drastically. Another frequent misconceptualization of confounds, detailed by the second method, occurs when confounds are used as proxy variables to control for variance that is not directly measureable, and no explicit attempt is made to ensure that the proxy variable adequately represents the underlying, intended construct. For this second demonstration, I used SHARE data to estimate models varying in the degree to which proxy variables represent intended variables. Results showed that parameter estimates can differ substantially across different levels of proxy variable representation. When imperfect proxy variables are used, an insufficient amount of variance is removed from the observed spurious relationship between design variables. The findings from this methodological demonstration underscore the importance of precisely imbuing confounds with conceptual
meaning and selecting proxy variables that accurately represent the underlying construct for which control is intended.
1. INTRODUCTION

The scientific method was developed to help minimize the inherent biases of the human mind as it makes inferences about the world. A hallmark of this method is the empirical analysis of relations among variables that is made possible when we convert a latent or observed phenomenon into a number. An inherent consequence of this conversion is that subsequent discussions of identified relationships may be difficult to interpret. This loss of meaning happens, at least in part, because our analysis and the interpretation of the results are based on and are in the format of numbers. We are, therefore, working and thinking on a scale that is not entirely conducive to making rapid transitions between numbers and the conceptual phenomena that they represent. These numbers are often, if not always, arbitrary assignments (scores on a depression inventory to characterize depression), and it is, therefore, difficult, especially without extensive experience with such measurements, to interpret how the new metric maps, both numerically and conceptually, onto the phenomenon in question.

Furthermore, we become familiar with the language of research and often speak in and apply terms without much consideration of the nuanced circumstances under which these concepts were developed. For instance, the term confound is often used as a synonym for nuisance variable, when in fact this variable, by definition, has substantial explanatory power in a theoretical model. A blind reliance on routine, empty empirical language further obscures our theoretical intentions with our methodological actions. This difficulty in aligning numerical representation with conceptual meaning not only affects interpretation, but it can, more alarmingly, affect the research method. Many researchers are quick to incorporate proxy variables in their research, but if these researchers are not cognizant of the underlying
construct they seek to represent, they may select variables that do capture the variance they plan to remove. It is up to the investigator, as well as critics and readers, to accurately work with quantitative representations and translate results back into the original meaning of the research.

An integral component of research is estimating the effect that one variable has on another. Statistical analysis yields quantitative metrics that allow investigators to infer whether or not these variables are significantly (e.g., causally) related. A common problem that can arise when a significant association is revealed is that the observed relationship is, in fact, spurious. That is, the empirical association exists for reasons other than the one variable directly causing the other. This fabrication can occur when an unmeasured third variable influences these two design variables, such that they appear to be directly related. If this third variable is not accounted for, the obtained results of the spurious correlation (parameter estimates, tests of statistical significance), and the resultant conclusions, may be biased or altogether incorrect.

The idea of the third variable, or confound, is not at all new, and to address this issue, investigators often collect data on this offending meddler and incorporate it into the statistical model. By estimating the effect of this confound, they can remove its influence mathematically from the comparison of interest. If investigators show that the primary variables are statistically related when confounds are controlled (i.e., their effects held constant), the investigators can infer, with some confidence, that the effect of the predictor on the outcome is present above and beyond the influence of these other measured variables. If the correlation of these primary variables drops to zero when the third variable is controlled, investigators can infer, with equal confidence, that the relationship between the two primary variables is spurious and any
Correlation is caused by the influence of this confound (Lynam, Moffitt, & Stouthamer-Loeber, 1993; Moffitt, Gabrielli, Mednick, & Schulsinger, 1981).

Confounds Translated

In this volume, I examine research processes that have stripped confounds of conceptual meaning, and I evaluate methods to help investigators see how important confounds are in accurately describing causal dynamics.

The first step in understanding confounds is to explicitly define the functional process that allows them to cause two variables to be spuriously correlated. We also need to have a sound understanding of how confounds arise and what effect they have in research.

In order for a confound to exist, this third variable must conform to two rules. First, the confound must be correlated with the original predictor, and second, it must be correlated with the outcome (Elmes, Kantowitz, & III, 2011). If the first condition is not met, but the second is, then this third variable is simply another predictor variable and does not obscure the direct effect that the primary predictor has on the outcome. If the first condition is met, but the second is not, then this third variable will again have no simultaneous causal influence on both the predictor and outcome, since changes in the third variable will not correspond to changes in the outcome.

Confounds can have many forms in research, but two common circumstances that are particularly worthy of discussion are those that arise as a function of the investigation process and those that are relevant predictors of an outcome that are not conceptualized as part of the research design. Confounds deriving from the research process can result from features of the study design and methodology such as subject selection bias, ineffective group randomization,
survey response bias, non-random subject drop-out, and features of equipment that impact outcomes. For instance, Shapiro et al. (2002) conducted a randomized controlled trial to evaluate a program to help breast cancer patients manage stress. During the research process, however, the method turned into a quasi-experimental design. The authors presented the premature disclosure effect to account for the observed nonequivalence of groups in terms of quality of life and psychological distress after successful randomization was obtained on demographic and health status variables. This effect was likely a result of the requirement, based on intervention logistics, to notify participants of their treatment group prior to baseline assessments. Upon learning of group assignment, participants may have been more likely to exhibit psychological characteristics such as stress or drop out of the study completely.

Confounds emerging as relevant variables that are not included in a model cause spurious correlations between the primary predictor and outcome because these confounds are simultaneously related to both design variables. Shapiro et al. (2002) used the terms “a priori relevant predictors” and “hidden predictors” to reference this class of confound.

Donahue, Lichtenstein, Långström, and D’Onofrio (2013) used the Study of Twin Adults: Genes and Environment (STAGE) data to estimate the relationship between age at first intercourse and psychosocial health outcomes and high risk behaviors exhibited as young adults. The authors were primarily interested in whether familial factors confounded the observed relationship between age at first intercourse and health-related behaviors. They found that twins who had shared familial backgrounds but who differed in age at first intercourse, on average, had similar psychosocial health outcomes such as drug use, criminal offenses, and depression. This finding supported the hypothesis that familial influences are likely causes of
psychosocial health outcomes and, therefore, need to be included in models evaluating factors related to such psychosocial health outcomes.

Confounds in Use

There are two common problems that arise when we work with confounds that are likely caused by the inadequate conceptualization of what confounds are and how they relate to the research at hand. The first problem deals with the perception of the role that confounds play in research. As the method of estimating and removing confounding effects produces a cleaner assessment of the relationship between the two variables of interest, many investigators have historically interpreted this third variable as a nuisance. The problem with viewing confounds strictly as noise or inert obstacles that need to be extracted and discarded is that they become bereft of scientific viability. These variables, however, are often quite meaningful (especially if they completely account for a perceived relationship between two other variables) and may ultimately play an integral role in generating the substantive knowledge required for continued epistemological progression. Thus, blindly throwing confounds into a model without considering their place within a particular phenomenon can greatly inhibit our comprehensive understanding.

As an example of this common practice in research, Brandt, Deindl, and Hank (2012) investigated a suite of historical and current demographic variables to evaluate influences on successful aging in older Europeans. Analyzing data from the Survey of Health, Ageing and Retirement in Europe (SHARE), Brandt et al. (2012) controlled for a number of demographic, socioeconomic status (SES), and health risk behaviors when estimating predictors on a derived variable measuring successful aging. These authors found significant associations between
successful aging and variables such as education, income, smoking and alcohol use, and childhood SES indicators. While these variables were discussed as influencing successful aging, there were no models that tested hypotheses of the dynamics regarding how these variables interact to allow someone to age successfully. These variables were simply entered stepwise into statistical models to isolate which variables were related to successful aging. It would seem likely that childhood SES indicators would be predictive of education and tobacco and alcohol use, which would in turn predict income. Incorporating plausible hypotheses into a structural equation model (SEM) would permit a more detailed explanation of how historic and current demographic variables interact to produce elderly individuals leading healthy and productive lives.

Another example of confounds not being conceptually integrated into a model is Buber and Engelhardt's (2011) research on the association of age on depression. Using SHARE data, these authors included a number of socio-demographic (education, economic strain) and health (chronic disease, physical disability) confound variables to accurately estimate the relationship between age and depression. The authors found that depression was significantly associated with older age groups (> 70 years old) while controlling for confounds. Furthermore, many of these confounds, such as low education, economic strain, and physical disability were also related to depression. Despite the observed statistical and conceptual significance of the confounds, no attempt was made to explain what part these confounds played in causing a statistically significant relationship between age and depression. Rather, the discussion consisted mostly of how the relationship between age and depression changed with the addition of different covariates in the model. While this perspective is certainly informative, it
would have been more so if the authors had discussed how these covariates relate to age, depression, and to each other. Causal theories could be tested to explain relationships among education, economic strain, and various health outcomes that were identified in the study and that have been repeatedly found in the literature. Despite lumping these variables into a univariate general linear model (GLM) for the majority of the analysis, Buber et al. did make a sound attempt at explaining the relationships among the variables by running a Sobel-Goodman test for mediation. The authors found that physical disability and education could be effects of age, and that these variables could, in turn, have a direct effect on depression. It is clear from the conclusions based on this test for mediation, that an analysis specifying detailed causal relations can enhance our ability to make more nuanced explanations. In this sense, physical disability and education are not seen as control variables, but as meaningful factors in the process of aging, declining physical health, and depression.

Confounds, therefore, should be (and frequently are) thought of as variables that interact within a system of relationships that can help explain how certain dynamics unfold, as opposed to nuisance variables that need to be removed from models. For instance, Duckworth, Quinn, Lynam, Loeber, and Stouthamer-Loeber (2011) hypothesized that individual differences in test motivation have caused substantial overestimation of the relations among intelligence and life outcomes, such as education, employment, and criminal convictions. Rather than simply lumping test motivation into a model with other predictors, Duckworth et al. (2011) integrated it into a comprehensive structural equation model detailing their hypothesis of how this “confound” influences intelligence and life outcomes. By creating a latent construct of non-intellective traits that was, in part, measured by test motivation, they were able to correct for
its influence on life outcomes, while, at the same time, introducing a mechanism by which test motivation influences life outcomes. This method of interpreting confounds clearly adds more meaning to the interpretation of the phenomenon of interest, as well as the confound variable itself.

By considering and modeling the role of “confounds” in a system of variables, we get a strong understanding of how these variables relate to one another. However, this knowledge of association does not guarantee that these manifest variables accurately measure the constructs we intend to model. Investigators will often aimlessly include demographic variables that are typically used as confounds (e.g., age, sex, etc.) without any thought to how accurately these variables represent the variance they want to control for. The poor representation or conceptualization of what underlying construct we are ultimately trying to control for when incorporating confounds is the second problem related to use of confound variables.

The need to have a clear conceptualization of the variance for which statistical control is sought is integral to the use of proxy variables and how well they represent this underlying construct (however, this issue related to proxy variables is not limited to confounds but is relevant in any circumstance in which one variable is being used to represent another). The reason proxy variables play such a large part in this problem is that investigators often do not have direct access to the variables for which they want to control. In these cases, proxy variables are used to represent the intended variables based on their assumed correlation, and it is this assumed correlation that hypothetically allows the proxy variable to remove the same variance as the intended variable. For instance, many studies will use education as a control variable when income is not available because education and income are highly correlated.
However, correlations and established validity coefficients do not necessarily establish sound mapping between the indicator and the construct from which measurement is targeted (Sechrest, 2005). Education and income may indeed be correlated, but education may not capture the variance in income that is associated with design variables in a model.

When estimating or controlling for the effects of confounds in a research model, we must ensure that these measured variables accurately represent the underlying construct for which control is intended. Two issues, in particular, warrant determination when we evaluate the quality of this proxy relationship. The first is a detailed conceptualization of how well the observed variable represents the latent construct of primary interest. Many researchers, especially in the social sciences, who use income, for instance, as a control variable are not interested in the number of dollars a person earns per se; rather, they are interested in an underlying construct, such as resource availability or familial provisioning, that is relevant to the theoretical model and object of the research.

The second issue is a specific instance of the first issue and relates to how accurately the operationalized variable represents the intricacies of the construct it is intended to measure. Income may be tested as a predictor of quality of life; however, the effectiveness of the income measure depends on how closely its operationalization embodies the construct predicted to affect quality of life. If resource availability is a primary influence on quality of life, net income may capture the variance in this outcome better than gross income. Two people with the same gross income, for instance, may have vastly different resources available if one allocates a large percentage of his or her gross to family health insurance, while a single worker is able to retain a larger portion of his or her base salary.
The importance of adequately identifying and operationalizing a proxy variable is clearly an issue that spans many scientific disciplines. Demonstrated below are examples of how a single variable, income, and then education, can represent drastically different constructs across health, criminology, and education.

**Income in Health Research**

Wood, Boyce, Moore, and Brown (2012) evaluated competing theories used to explain the influence of income on mental health. Despite the focus on actual income figures, the authors were really interested in two mechanistic processes that underlie income and may directly influence mental illness. One such process that income can represent is access to resources and the ability to obtain medical treatment. The second dynamic is social status. Occupying low social status, according to evolutionary theory, can cause individuals to experience profound anxiety and physiological distress, such as reduced serotonin and dopamine that can alter brain and behavioral functioning.

Data from the British Household Panel Survey were analyzed, and a relationship between income and psychological distress was detected ($b = -.08$ to $-.13$). This relationship deteriorated ($b = -.04$ to $-.06$) with the inclusion of income rank ($b = -.29$ to $-.40$) to the statistical model, leading the authors to infer that, rather than simple access to medical services, social standing among community members can explain mental health. In this scenario, proxy variables need to measure social rank when investigators predict mental health. As the results indicated, income rank represents this construct better than absolute income.

Another example of linking proxy variables with latent constructs is the investigation by Cohen, Agree, Ahmed, and Naumova (2011) on how interactions with children can promote...
rates of pneumonia and influenza (P&I) infection in the elderly. Because increased interaction with children increases the exposure to P&I, Cohen et al. (2011) hypothesized that grandparents who are the primary caregivers of children will be more likely to contract such infections than grandparents who are not primary caregivers. Since grandparents who are primary caregivers are more often of lower income than grandparents who are not primary caregivers, income is hypothesized to predict rate of P&I in the elderly. Based on this logic, income can be interpreted as a proxy for a sociodemographic pattern in which low income grandparents, through primary caretaking activities, are in continued contact with virus-shedding children.

County level P&I claims were obtained from the Centers for Medicare and Medicaid Services, and county level median household income was obtained from the US Census Bureau. Income was negatively related to the proportion of primary caregiving grandparents ($r = -.52, p<.0001$). Seven Poisson regressions, varying in the predictor variables modeled, yielded a significant relationship between both low income and grandparental caregiving and P&I infection rates. Across the seven regressions, relative risks ranged from 1.31 to 1.68 for low income and 1.13 to 1.23 for grandparental caregiving. Cohen et al. (2011) offered a number of factors represented by income that are hypothesized to cause P&I infection rates in the elderly. High interpersonal density resulting from small living quarters can increase physical contact and, therefore, virus spreading, among household members. Furthermore, a likelihood of living in public housing or in houses with deteriorating ventilation is increased in low income families. Therefore, in addition to the always plausible lack of access to care that low income can represent, low income can characterize a state in which grandparents live in constant and close
contact with children and inhabit local environments that make viruses easier to spread. The hypothesized significant impact that this underlying construct has on P&I rates in the elderly underscores the importance of precisely modeling this construct with valid proxy variables.

**Income in Criminology Research**

Income can also represent a variety of distinct constructs in criminology and sociology research. Galloway and Skardhamar (2010), for instance, investigated the association between SES and crime. The hypothesized mechanism underlying this relationship is the increased resources and opportunities available to high SES individuals that allow them to accomplish their goals. In addition to materialistic resources, high SES can also relate to advantages and support bestowed by parents. Without access to material and familial support, low SES individuals may view criminal activities as an alternative strategy for attaining solvency and other financial goals and necessities. Furthermore, low SES can affect interpersonal interactions, in which financial and employment stress breeds a contentious home environment. Children experiencing such dysfunctional parental relationships may be more likely to engage in risk taking and aggressive behaviors (Galloway & Skardhamar, 2010).

Galloway and Skardhamar (2010) obtained individual-level data from various Norwegian registers containing income, crime, and education records. The authors used income to represent SES. Results showed that, although there was a strong association between income and crime, the relation between income and crime diminished with the inclusion of parental education in the model (hazard ratios of the income variable moved closer to one in models that included parental education). Galloway and Skardhamar (2010) interpreted this finding as evidence for the importance of educational resources and value, beyond income, in
determining criminal activity of children. This perspective is in accord with the hypothesis that crime is a mechanism to meet financial demands when other goal-required resources, such as education, are not available. Income itself is not seen as directly impacting crime; rather, a lack of resources required to smoothly traverse an employment landscape, along with a fractured familial support network, will promote an acceptance of illegal methods for generating income. Hence, education may more effectively capture this construct compared to income and will, therefore, be a superior proxy with which to model incidences of crime. By specifying this underlying construct accurately, investigators can better estimate the effect of access to resources, via parental education, on crime.

**Income in Education Research**

Income can also represent constructs related to academic performance. Beyond academic effort like study time, social and other external factors such as community, family, and friends, can influence academic success. Mullis, Rathge, and Mullis (2003), for example, noted that educated or wealthy parents can teach their children study strategies, provide materialistic resources, such as technology, and be active in the educational experience to help these students succeed.

Mullis et al. (2003) obtained student and school-level data from the National Educational Longitudinal Study to test a comprehensive model specifying social capital and resource capital to predict academic performance (grades). This source contained data on, among other demographic variables, parental income and education and the degree of resources available that make the home conducive to academic success (e.g., newspapers and work areas). These three variables were modeled as indicators of resource capital and are
interpreted as proxy causes of academic success via parents’ provisioning of academic support.

Results indicated that resource capital accounted for a large portion of the variance (23%) in academic performance, with income obtaining the largest beta loading of .73. The beta loadings for parental education (.68) and home resources (.67) were slightly lower.

By modeling multiple constructs (resource and social capital) in a structural equation model, with each construct having multiple indicators, Mullis et al. (2003) maximized the potency of proxy variables in modeling the underlying construct of interest. Income, by itself, did not contain the information necessary to explain the variance in academic success sufficiently.

**Education as a Proxy Variable**

Like income, education is also commonly found as a proxy for underlying constructs. For instance, Zhang et al. (2009) documented relation among SES and healthcare seeking behaviors in Chinese females with sexually transmitted infections. The effect of SES on health seeking behaviors was hypothesized to manifest from low SES females being less likely to seek services for preventative measures and not having access to legitimate healthcare. High SES females, with greater access to formal healthcare, would be more likely to seek services.

Zhang et al. (2009) obtained data from the Chinese Health and Family Life Survey, which comprised numerous health and demographic variables. Education, measured as levels of school grades completed, was used as an indicator of SES. Participants with high and low education levels had the lowest rates of healthcare seeking behaviors, whereas those with high school degrees had the highest rates of treatment sought. With “primary education or less” being the reference group, high school level of education had an odds ratio of 2.27 for
predicting health seeking behaviors. This odds ratio for high school was greater than that obtained for middle school (1.81) and college or above (1.27).

Zhang et al. (2009) detailed a number of underlying constructs represented by education that can explain this nonlinear finding. Level of education can represent, beyond SES, the knowledge a patient has of sexually transmitted infections and what course of action is required for treatment. Such knowledge can lead patients to seek treatment in more formal locations, such as hospitals, compared to private, unlicensed providers. Another influence on health seeking behavior is that highly educated women may feel a stigma of having contracted a sexually transmitted infection and may, therefore, eschew treatment. These two factors may also interact, producing a population of high school educated women with enough treatment seeking knowledge to obtain effective care, yet who do not feel a stigma that would cause them to avoid treatment.

Généreux, Auger, Goneau, and Daniel's (2008) investigation of the relationship between negative birth outcomes, such as preterm birth and low birth weight, and living in proximity to highways provides another example of education being used to represent a larger underlying construct. The underlying causal premise of this investigation was that living close to highways increases exposure to air pollution, which in turn, will cause adverse birth outcomes. Low SES mothers are more likely to both live close to highways and to experience negative health problems. Therefore, to get a clean estimate of the effect of living in proximity to highways on birth problems, Généreux et al. (2008) wanted to control for SES. However, SES was not directly measured, and maternal education was used as a proxy for this confound.
Généreux et al. (2008) obtained birth outcomes records from the Quebec birth registry, and the postal code of the mother was used to determine distance from highways. Results showed that distance from highways and adverse birth outcomes were negatively related. Odds ratios comparing measured education levels with the reference group, some university, in predicting adverse birth outcomes ranged from 1.23 to 1.45. However, and unexpectedly, of those in proximity to highways, high SES mothers were more likely to have negative birth outcomes than low SES mothers.

Généreux et al. (2008) offered two explanations for this unexpected and counterintuitive result, both of which detail valid underlying constructs that SES, and, therefore, education, can represent. First, in Canada, strong social service programs providing quality healthcare to low income individuals can help offset some of the negative effects of air pollution. Second, low SES mothers are more often associated with health risk behaviors and activities, such as smoking, and may experience negative birth outcomes independently of how close they live to highways. These other health risks may conceal a linear relationship between highway proximity and birth problems.

The examples outlined above demonstrate how a single variable can represent vastly different constructs when we predict outcomes. It is, therefore, essential for investigators to have a clear understanding of the constructs they wish to model and ensure that proxy variables adequately represent the latent phenomenon that is hypothesized to cause the outcome. Misrepresentation of underlying constructs by proxy variables can lead to inaccurate parameter estimates.
The following examples are of the second issue regarding the quality of proxy variables: accurate operationalization of measured variables in representing theoretically important latent constructs.

**When Income Measures Income**

As outlined above, Galloway and Skardhamar (2010) estimated the relationship between parental income and child criminal offense. The income used in statistical models was measured as yearly total earnings from employment and social security. Income from social services, such as welfare, was not included because it was unavailable. Measured income was, therefore, defined as earned income. This distinction, or at least the disclosure of the specific type of income used, is essential for evaluating which construct income is really measuring. Can welfare income be expected to play a part in predicting criminal offenses, and does the exclusive use of earned income reduce the amount of variability that can potentially be explained in crime rates? To answer this question, at least conceptually, we must contemplate what it is about income that relates it to crime and ensure that our measure of income adequately represents this latent factor.

In addition to their substantive research question, Galloway and Skardhamar (2010) also evaluated the effectiveness of different income measures in representing the income construct. The average annual parental income from the time a focal child was born through his 10th year was calculated to create a measure of long-term income. Many studies do not have access to such rich longitudinal data and must rely on single measures of income. To estimate the predicted increase in measurement error from using short-term income, Galloway and Skardhamar (2010) reran their analysis using the parental income measured during the child’s
While there were differences in point estimates between the two models estimating the relation between measures of income and crime, similar patterns between these two variables to the long-term income measure did emerge. This finding suggests that these two income measures may be capturing the same underlying construct that predicts criminal activity.

In addition to the methodological issue of income operationalization, Galloway and Skardhamar (2010) also noted that long-term measurements can capture an altogether different construct: a persistence in income class. This persistence can represent stability in income to support a family or remain in chronic poverty. Two parents with similar incomes in their child’s 10th year can have vastly different incomes in the years preceding. An educated and financially well compensated parent who lost his or her job in year nine will likely, based on the hypotheses, have different perspectives on crime than a parent who lived in poverty throughout the duration of the study. In this case, operationalizing income as long term captures a construct that has, based on the outcome, predictive value that differs from short term income.

In another example of variants in operationalization, McMillan, Henry, and Crosby (1995) discussed that SES is frequently measured or perceived as either unidimensional or multidimensional constructs. Different multidimensional measures will often vary in the degree that certain constructs or variables are emphasized. For instance, McMillan et al. (1995) noted how some multidimensional SES measures focus more on education than income, while other instruments reverse this priority. There will also be instances when a unidimensional measure is more effective or relevant than a multidimensional measure. When selecting a particular
instrument, we must have a thorough understanding of which predictive construct is best represented by the measure.

A final illustration of the variation in income operationalization is Ram's (2010) evaluation of the positive association between social capital and happiness. The weak relationship between these two variables obtained in previous studies prompted Ram (2010) to focus on the impact of income, which is typically understood to influence happiness, in possibly obscuring this relationship. Because life satisfaction data were grouped by country, GDP per capita, obtained from World Bank, was an appropriate proxy to represent income. Although a strong theoretical basis exists for modeling the effect of income on happiness, there was little discussion of how income actually affects happiness. Knowing this mechanism is essential for evaluating the suitability of using GDP as a proxy for income. Although GDP was found to be significantly related to happiness in various models (t values ranged from 5.89 to 7.89), the contribution of income varied dramatically when countries were grouped into high and low income categories (t values range from -.06 to 2.41). Ram (2010) suggested that familial and community structure in countries with a particular income status can provide a latent basis for observed levels of happiness. This type of interpretation is useful in identifying constructs that proxy variables can represent and determining if the measured variable is operationalized in accord with the desired mechanism of influence.

Effects of Imperfect Proxy Variables

When predictors are estimated without rigorous consideration of how and why they interact with other variables in a model, poor conceptualization and operationalization of the observed-latent trait relationship can occur. These conditions compromising the
representativeness of measured proxy variables are likely to cause inaccurate parameter
estimates because these imperfect proxies do not account for an adequate amount of the
construct’s influence on other select variables in the network.

What, then, is the effect of using a non-representative proxy variable? The study
detailed above by Wood et al. (2012) demonstrates the dramatic effect that an imperfect proxy
can have on parameter estimates. The authors hypothesized that low social status can cause
physiological changes in the brain, which may lead to psychological distress. Using income to
represent social status, they found that, across numerous models, significant beta values for
income ranged from .08 to .13. However, income may not have been ideal for capturing the
interpersonal competitive spirit that can cause changes in the brain during the struggle for
status. When income rank was included in the model, the beta for income deteriorated to half
its original value (ranging from .04 to .06), while income rank yielded a beta (ranging from -.29
to -.40) up to almost four times that of the original income estimate. Income rank may,
therefore, be effectively tapping into a competitive earning construct that causes physiological
changes in the brain. These estimates illustrate the discrepancy in the two income variables’
ability to measure, under the proposed hypothesis, causes of psychological distress.

In addition to the need for having a clear understanding of the construct that proxy
variables are intended to represent, we must also be aware of the strength of the relationship
between these two variables. Verropoulou and Tsimbos (2007) used home ownership as a
proxy variable for wealth when evaluating the effect of demographic and health variables on
depression. There was, however, no discussion or estimate of the strength of the relationship
between home ownership and wealth. Although validity coefficients between the proxy
variable and the construct it is intended to represent are often not known, estimates of this relationship can help evaluate findings. Home ownership lost statistical significance when additional demographic and health variables were included in the model. One can only wonder if wealth is not related to depression, or if home ownership is not a good proxy for wealth, and the essential variance of this construct was not modeled.

To address the common misuse of confounds discussed above, I will present and evaluate two methods for imbuing confound variables with meaning and interpretability. Such methods will show how approaching research with a richer understanding of the intent of the investigation and conceptualization of the variables involved can greatly improve both quantitative and substantive results. The findings of this demonstration can be used to base recommendations for the importance and need for a greater understanding of the process and content of quantitative research. The two methods to be evaluated are detailing the relationship that confounds play in a network of variables and specifying the representativeness of proxy variables.

The first method calls for the identification and conceptualization of confounds, not as nuisance variables, but as valid explanatory variables in a network of relationships. Understanding the part that a confound plays in a network of variables lets us more effectively model the scientific problem. In an attempt to demonstrate this role, I specified and estimated the effect of specific confound designations on three models that vary, both conceptually and mathematically, in how confounds are interpreted. This illustration sought to achieve three aims:
1. Detail how different model specifications relate to different conceptual meanings

2. Estimate how model parameter estimates change with different confound specifications

3. Present, mathematically, via path decompositions, the correction required to remove the effects of confounds on the relationship between the primary predictor and outcome ($r_c$)

The second method is one that calls for the understanding and accurate conceptualization of proxy variables. Investigators must have a clear understanding of what construct is being modeled with proxy variables and how closely these manifest variables represent this underlying construct. In an attempt to evaluate the effect of misspecifying the measured variable-latent trait relationship, I conducted a sensitivity analysis on models varying in the representativeness of measured variables. While this issue related to proxy variables is not limited to confound variables, it is demonstrated in models in which the control variable is a proxy for an intended construct.
2. METHOD

Data

I used archival data from Wave 1 of the Survey of Health, Ageing and Retirement in Europe (SHARE) for this demonstration. The SHARE project houses an extensive database of health, social, and SES measures of older Europeans and Israelis; thus, providing adequate information with which to estimate detailed structural models comprising multiple levels. The first wave of SHARE data was collected in 2004 and contains 28,517 records. I chose to limit the analysis to one country (Spain) to reduce the variability that can result from country-specific heterogeneity (Buber & Engelhardt, 2011). I selected Spain because it had a substantial sample size and contained correlations with the design variables (physical inactivity and depression) and a number of health and demographic variables. The final sample size was 2,396.

Measures

Many of the published SHARE studies involve influences on depression in aging adults, and these studies almost invariably control for standard confound variables such as age and health status. Having these studies to reference different ways of incorporating confounds in research made focusing on depression appropriate for the current demonstration. I drew upon this research for model development and selected the effect of physical inactivity on depression as the main substantive hypothesis to be tested. I have added a number of other confounds, such as age, health, and wealth, that have been used in depression research with SHARE data (Buber & Engelhardt, 2011; Ladin, 2008; Verropoulou & Tsimbos, 2007). The variables used in this present study were: assessment of current health on a five point Likert scale; number of chronic conditions that the participant was told by a physician that they had;
the number of signs and symptoms they report being bothered by; the number of drugs taken; presence or absence of any long-term illness; presence or absence of hopes for the future; presence or absence of feelings that the participant would rather be dead over the past month; amount of vigorous and moderate physical activity; number of daily living activities that the participant has difficulty performing; and ownership of property. See Appendices B through F for a complete list of questions that comprised these measures. The Euro-D and physical activity variables were obtained from the SHARE dataset of generated variables. This dataset contains variables created from multiple questions from the original SHARE survey data. The questions used to create these two generated variables used in this study are listed in the Appendices E and F, respectively.

When evaluating the effects of imperfect proxy variables, I created three constructs of primary interest that proxy variables were intended to represent. These constructs were calculated based on the standardized measured variables described above. These variables were standardized to have a mean of zero and standard deviation of one. I used the STANDARD procedure in SAS 9.3 to generate these standardized variables. I constructed the physical health construct using the number of chronic health conditions, number of health-related symptoms participant is bothered by, number of drugs taken, absence of long-term illness, and rating of current health. I created the mental health construct from the “hopes for the future” and the “feelings that the participant would rather be dead over the past month” measures. Finally, I used the number of Activities of Daily Living (ADL) and the number of Instrumental Activities of Daily Living (IADL) that the participant had difficulty performing to construct the physical and mental impairment construct. While the ADL measure is a survey of limitations that result
strictly from physical disability, the IADL measure contains a number of limitations that result from cognitive impairment, such as using a map and making telephone calls. Table 1 contains the alphas of these constructs.

Depression was measured by the EURO-D. This measure comprises a list of 12 symptoms of depression for which participants report a presence or absence, and scores range from 0 to 12, with higher scores indicating more severe depression (See Appendix E for symptoms used to create this measure). This measure was developed to compare depression across European countries, and its strong psychometric properties have been evaluated and published by (Prince et al., 1999).

Analysis

Structural Equation Models (SEM) were used to specify and estimate detailed networks of relationships among design variables and measured and latent confounds. SEM is preferred for this project because it permits the modeling of multiple independent and dependent variables constituting a complex and comprehensive network of relationships. Traditional use of removing confounds with univariate methods is not conducive to modeling multiple directions of variable influence simultaneously. This limitation is substantial when we evaluate confounds because there is no known way to simultaneously estimate the direct effect between the confound and predictor and the confound and the outcome with univariate methods.

SEM is also capable of parsing out common factor and specific variance, a feature that is essential to maintaining conceptual meaning to all of the variables in the system. Path decompositions from the SEM were determined to obtain the proper correction to remove the
effects of the confounding variable on the primary predictor and outcome. I used the CALIS procedure to execute this analysis with SAS 9.3. All estimates were made using Maximum Likelihood (ML), and because the intention of the demonstration is more methodological than substantive, standardized path coefficients are reported. Since many of the path models were designed to specify different conceptual interpretations of variable relationships, I adopted causal language to reflect these aims. Resultant statements are used to assess how the model variables’ role and meaning change with certain directional pathways and theoretical construct representation. The SHARE data used in the analysis is observational, and no control has been implemented to establish causal inference. Recognizing this limitation, I, in no way, intended to make causal inferences on the substantive results obtained by the SEMs.

The first step in the analysis was to estimate a Design Model, which represented the primary substantive research question of the effect of physical inactivity on depression (Figure 1). This model is meant to be an example of a typical comparison that investigators would test when conducting research on depression. This model contained no controls and is used as a reference point against which other models can be compared.

Confound Specification Analysis

Because there is an infinite number of ways that confounds can be specified in a model, I limited the analysis to models that represent three basic and central issues of confounded relationships among variables. The first two models are based on a simple three variable system with a primary predictor, primary outcome, and potential confound. Models representing this issue are spurious correlation and mediation. While mediating variables are not defined as a true confounds, they are often modeled as confounds in univariate analyses,
and their effects on both independent and dependent variables are relevant to this demonstration. The third model was intended to represent multiple confounds that have multiple effects in a system of relationships.

**Proxy Misspecification Analysis**

When estimating the effect of imperfect proxy variables, I used three different confound variables that differed in how well they represented the construct of interest. I focused on age as the main control variable. This variable is germane to demonstrate proxy variables’ ability to represent the construct of interest because the actual number of years one has been alive is usually not pertinent to research; rather, it is typically the health and activity patterns that are associated with age that produce the effects that researchers want to adjust for when they use age as a confound. Based on this reasoning, to accurately estimate the direct effect of physical inactivity on depression, I incorporated physical health, mental health, and physical and mental impairment as the effects that need to be removed and the constructs that are intended to be represented by age. Once parameter estimates were obtained with age as a confound, I re-ran the models twice more, each using a different confound with a different correlation to the constructs of interest. Wealth, measured by home ownership (following Verropoulou & Tsimbos, 2007) was included to evaluate the use of a poor proxy variable, and assessment of general health was included to evaluate the use of a good proxy variable. Age had a moderate correlation to the constructs of interest and allowed me to estimate the use of a moderately good proxy variable. Table 2 shows the correlations of proxy variables with constructs of interest.
Sensitivity analysis comprised the comparison of parameter estimates of models varying in the strength of the relationship between a measured proxy variable and the construct it was intended to represent. The parameter estimates of these proxy models were also compared with the parameter estimates from the model specified with the actual constructs of interest. The model containing the constructs is the Reference Model and is intended to represent the correctly specified model “in the eyes of God.” I left a nonsignificant path coefficient in this Reference Model to allow for comparisons with additional models with proxy variables. Although all beta weights for the design variables and the constructs were obtained through SEM, path coefficients for the constructs and indicators were generated with GLM. This two-step estimation process was conducted to stabilize the variance when I estimated the relationship between design variables and the constructs. Reference Models are included that contain the variable that is used as a proxy for the intended constructs. These Proxy Reference Models are included to estimate the direct effect of the particular proxy variable with the health and impairment constructs, and the unique direct effect of the proxy variable on the design variables. Standard SEM fit indices are compared for the Proxy Reference Model and the Proxy Model to demonstrate the effect of proxy misspecification. For the sensitivity analysis, I compared the standardized parameter estimates of the Design Model, the Reference Model, and those with the three different proxy variables. Comparison to the Design Model illustrates differences in parameter estimates as a result of controlling for confounds, and comparison to the Reference Model shows the error in estimation of path coefficients caused by a misspecification of the proxy variable.
3. RESULTS AND DISCUSSION

Because the intention of this paper is to evaluate different conceptual designations of covariates in a model and to identify and estimate the effects of misspecifying measured-latent trait relationships, and because the results are methodological rather than substantive, the presentation of results is most meaningful if it includes interpretation of how parameter estimates change with different conceptual specifications. Therefore, the results and discussion sections are combined.

The primary and substantive theoretical research question is the direct effect of physical inactivity on depression (Figure 1). Subsequent models incorporating confounds are compared against this model to detail how parameter estimates change with different methods of conceptualizing confounds. There was an observed significant effect of physical inactivity on depression, \( \beta = .21, t = 10.39, R^2 = .04. \)

**Method 1: Confound Specification**

**Spurious Correlation**

Figure 2 shows the standardized parameter estimates for the model testing the direct effect of physical inactivity on depression while controlling for physical disability, measured by limitations in Activities of Daily Living (ADL). The variables in this model were chosen because the relationship between physical inactivity and depression has been examined by many researchers using SHARE data, and in many of these studies, ADL has been classified as a confound. This former point is based on the hypothesis that exercise and general physical exertion have inhibiting effects on depression. The physical inactivity variable, however, does not tell us why a participant is physically active. One of the conditions we would like to control
for when estimating the effect of physical inactivity and depression is whether or not someone is physically unable to exercise or perform daily activities because of physical limitations.

Furthermore, physical limitations and the fact of being physically limited might well be one cause of depression. By this reasoning, physical disability, measured by ADL, would act as a true confound in that it could simultaneously cause physical inactivity and depression. By controlling for the variance in physical disability, I intended to get an estimate of the effect of physical activity on depression that is not related to physical disability.

In this spurious correlation model, the path coefficient of .05, with a $t$ value of 2.40, between physical inactivity and depression reveals that this relationship is spurious. This beta value is more than four times smaller than the beta weight obtained when no confounds were used (see Design Model, Figure 1). The Design Model in Figure 1, by not estimating a relevant variable, overestimated the relationship between physical inactivity and depression by .16 standardized units, compared to the spurious correlation model.

The variance explained in depression also increased dramatically from .04 in the Design Model to .25 in this spurious correlation model. Furthermore, .12 of the variance in physical inactivity was explained by the direct effect of physical disability, an estimation not attainable when, in a univariate GLM, physical inactivity can only be specified as either an independent or dependent variable. Comparing these parameter estimates to a univariate GLM in which physical inactivity and physical disability are predictors of depression, the SEM parameter estimates do not differ. In this sense, SEM does not provide a mathematical advantage or even a difference over the univariate GLM. The SEM, however, does provide an important conceptual advantage over GLM in model specification. In SEM, we could clearly model a spurious
correlation, in which one of the predictor models was specified as endogenous. The GLM does not allow us to discriminate spurious correlation and multiple predictors when independent variables are entered simultaneously.

For the spurious correlation model, the correction, obtained via path decomposition, for estimating the unconfounded effect between physical inactivity and depression, is \( r_{\text{cphydep}} = r_{\text{phydep} \mid \text{adl}} - P_{\text{adl,phy}} P_{\text{adl,dep}} \). By detailing the specific way that a confound interacts with other variables, we are able to make confounds more interpretable and identify the origin of the observed relationship between design variables. Lynam, Moffitt, and Stouthamer-Loeber (1993) provide an excellent example of meaningfully integrating confounds into a structural model. Evaluating the effects of IQ on delinquency, they modeled the influence of social class and test effort as confounds. Numerous hypotheses were specified regarding how these confounding variables interact with other variables in the model. Impulsivity and achievement were added as mediating variables when evaluating the effect of social class and test effort on delinquency. In addition to keeping the effects of these confounds constant, the authors explained how these confounds interact within a phenomenon of interest.

**Mediation**

Figure 3 specifies physical disability, measured by ADL, as a partially mediating variable in the relationship between physical inactivity and depression. In this scenario, physical disability has a direct effect on physical inactivity and depression, with physical inactivity having a direct effect only on depression. All parameter estimates are identical to the previous spurious correlation model. This is to be expected because the correlations between the two variables that have reversed causal direction from the previous model have not changed. This
finding does demonstrate that, despite conceptual differences, the spurious and mediation models are mathematically indistinguishable. Therefore, with only three variables in the equation, and not altering the variable pairs on which path coefficients are estimated, one cannot support one model over the other. Not being able to translate mathematical results for conceptual discrimination is a weakness, especially because the conceptual inferences supported by these two models differ substantially. Although it is possible that physical inactivity can cause physical disability, the reverse seems much more likely given contemporaneous data (the effect of physical inactivity on physical disability would be more plausible if we had longitudinal data showing that physical disability increased with high levels of prolonged physical inactivity). Because we cannot rely on mathematical results to discriminate the most likely causal specifications, an investigator must use both logic and a strong theoretical foundation to reach well-grounded conclusions. This type of critical thinking is necessary when mathematical results are equal but one model is logically sound and the other is neither reasonable nor plausible. For the present model, the correction, for estimating the unconfounded effect between physical inactivity and depression, is
\[
\rho_{\text{phydep}} = \rho_{\text{phydep}} - P_{\text{phy,adl}} P_{\text{adl,dep}}. 
\]

**Multiple Confounds**

This third model introduces an additional confound that is used to explain physical disability, measured by ADL. This confound is based on a hypothesis that being bothered by the symptoms associated with various health conditions causes physical disability. Being bothered by health-related symptoms also allows us to deconstruct physical disability variance that is not due to a current health condition. This variance partition is relevant to the current model that
specifies direct effects of disability on depression. If a participant has been disabled from a childhood illness or accident, he or she might not be as depressed years later as a result of this disability as a participant who is disabled as result of a host of contemporaneous health problems. Specifying this multiple confounds model, I added a path from the variable bothered by symptoms to physical disability. This model yielded identical beta weights compared to the spurious and mediation models, with an additional path estimate of .66 going from bothered by symptoms to physical disability. After adjusting this model, however, I located several additional pathways that, when modified, led to a model that was more efficient and yielded superior explanatory power compared to the univariate GLM approach.

For comparison, I ran a typical univariate GLM to evaluate the effect of physical inactivity on depression, while controlling for physical disability and bothered by health symptoms. Such a common model specification of design and confound variables yielded an $R^2 = .34$, and a beta weight of .22 for depression on physical disability, .07 for depression on physical inactivity, and .40 for depression on bothered by health symptoms. From these results, we can see that there is a direct relationship between physical inactivity and depression, but that bothered by health symptoms and physical disability have larger effects. The $R^2$ in the adjusted SEM is the same as the $R^2$ obtained in the GLM. If we think about what possible effects these confounds could have in the relationship of physical issues and depression, however, we get a much richer understanding of system dynamics using the SEM. Figure 4 displays the parameter estimates of a new model detailing the type of relationships that physical health problems can have on depression.
The paths between physical inactivity and depression and bothered by health symptoms and physical inactivity have been removed, due to the small beta values obtained. Although these paths were both statistically significant (likely caused by the large sample size), the beta values were small (.05 and -.08, respectively), and the Adjusted Goodness of Fit Index (AGFI) was slightly higher for this restricted model (.9739) than the full model where these paths were included (.9701).

In terms of the research question, there were a few interesting findings regarding the variables affecting depression. Although physical inactivity is related to physical disability, the variance of simply being physically inactive (i.e., not being constrained by physical limitations) is not strongly related to depression. Rather, being bothered by health related symptoms and physical disability are influences on depression. Furthermore, physical disability is strongly related to being bothered by health related symptoms. This model specification supports the hypothesis that physical disability is a sufficient cause of depression and provides, when the variance of physical inactivity is parsed out, a more comprehensive story, compared to the univariate GLM approach, of why someone would be depressed under these circumstances. We can, therefore, infer that being bothered by health related symptoms contributes to physical disability but not necessarily the physical inactivity that is unrelated to disability. By not explicitly defining the interactions among variables in a network (e.g., blindly adding variables in a GLM), we will likely not deconstruct variance estimates in a precise and accurate way. These results illustrate the exact purpose of meaningfully integrating “confounds” into a model. In this example, we can see how removing a specific type of variance can have a large effect on the type of explanations we can make. For this model, the correction for estimating the
unconfounded effect between physical inactivity and depression is \( r_{\text{phydep}} = r_{\text{phydep}} - \\
\text{P}_{\text{adl,phy}}\text{P}_{\text{adl,dep}}\text{P}_{\text{bb,adl}}\text{P}_{\text{bb,dep}}. \\
\)

While the above demonstration illustrated that properly defined SEMs can enrich our understanding, it is essential to remember that we are simply testing causal theories and that strong research design must be established when we test the validity of the obtained results and subsequent conclusions. The point of this method is that we need to think beyond just simply “controlling” for variables whose influence we want to remove and to introduce confounds into substantive interpretations. While there is a predominant focus on removing confounds, investigators must also be aware of the consequences of removing too much variance. This latter issue can occur when we are not clear on what type of variance we want to adjust for and can lead to inaccurate parameter estimates among primary design variables (Miller & Chapman, 2001).

Despite the advantages of enhancing the role of confounds in SEMs, adding more detail or causal pathways to a model forces us to choose among a greater number of potential explanations, thereby increasing the possibility that we select an incorrect model. With more complex models, we need to take additional precaution when concluding that a variable, such as physical disability, which was specified to be both a predictor and an outcome, was accurately modeled with directional arrows. For example, it is plausible that physical inactivity, at least if it were chronic before measurement, could have caused physical disability (as opposed to the reverse as specified in Figure 2). Similar errors, however, can also occur in univariate GLMs when variables are thrown into a model and conclusions are based solely on statistically significant associations with an outcome variable. Model fit statistics and traditional
parameter estimates generated by SEMs can help us select among well performing, competing models, and making risky predictions contributes to scientific advancement (Popper, 1962).

**Method 2: Proxy Misspecification**

A substantial number of SHARE researchers controlled for age when they investigated causes of depression (Brandt et al., 2012; Buber & Engelhardt, 2011; Ladin, 2008; Ploubidis & Grundy, 2009; Verropoulou & Tsimbos, 2007). Although many authors have acknowledged an observed relationship between age and depression, there has been no direct comment on what exactly age means as a confound in depression research. In this current demonstration, I focused on age as a confound, not because I am interested in the actual number of years that a participant has been alive, but because age is associated with many conditions that are likely to affect physical inactivity and depression. Authors using SHARE data to study depression have used current health, number of chronic diseases, physical disability, income, etc. as confounds because those variables have been shown to predict depression (Brandt et al., 2012; Deindl, 2013; Lindwall, Larsman, & Hagger, 2011). It is, therefore, reasonable to consider age as an effective confound that should remove variance from these health and SES sources because of its strong associations with these variables.

To begin evaluating the effect of age as a proxy variable, I first constructed a Reference Model that is intended to represent the correctly specified model “in the eyes of God” of the effect of physical inactivity on depression. This model controls for the effects of physical health, mental health, and physical and mental impairment and was used to compare against other models incorporating proxies for these three constructs (Figure 5). Because these constructs
were predicted to have a strong causal influence on both physical inactivity and depression, they were defined as true confounds.

The inclusion of physical health, mental health, and physical and mental impairment in the Reference Model renders the direct effect of physical inactivity on depression almost completely spurious, $\beta = -.01, t = .72$ (Figure 5). This substantially lower path coefficient between physical inactivity and depression compared to the one obtained in the Design Model (Figure 1) indicates that the Design Model does not correctly model the data. It is, therefore, essential to use confounds when we estimate this relationship between these two variables. In addition, the $R^2$ in the Reference Model surged to .55 for depression (compared to .04 in the Design Model) and emerged at .16 for physical inactivity. This finding is not surprising as the information in the three health constructs, obtained from a total of nine variables, used to explain the variance in depression (and physical inactivity), is greater than the information bestowed by the single variable, physical inactivity, in the Design Model. This Design Model, by not estimating a relevant variable, overestimated the relationship between physical inactivity and depression by .20 standardized units. Not modeling relevant confounds in this model will lead to a type I error, since the beta value for this relationship in the Design Model of .21 was statistically significant, $t = 10.39$.

**Moderate Representation of Proxy Variable: Age**

Although the Reference Model does a good job explaining the variance in depression, many investigators will lack comprehensive measures to thoroughly model the network of relevant variables. Researches in these circumstances must then rely on proxy variables to estimate and remove the influence of such confounding variables. Age was the first variable I
used in the sensitivity analysis of imperfect proxy variables. The ultimate purpose of age as a proxy variable, whether identified by an investigator or not, is that it should adequately represent the three health and impairment constructs used in the Reference Model. It is reasonable to consider that as one ages, negative physical health indicators such as chronic conditions and medications, negative mental health indicators, such as hopes for the future, and physical and mental health impairment indicators, such as inability to go grocery shopping and pay bills, should increase with age. Table 2 shows age as moderately correlated with physical health and physical and mental impairment and slightly correlated with mental health.

Figure 6 illustrates the Reference Model with age added to estimate the direct effect of age on the health and impairment constructs and the unique direct effect of age on the design variables. Age does not add any improvement in the model beyond the constructs in the Reference Model, and the path coefficients from the constructs to the design variables do not change much with the addition of age. Figure 7 contains the Age Proxy Model in which age is specified as the sole predictor of the design variables. Table 3 illustrates the drastic change in model fit between the Age Reference Model and the Age Proxy Model. The path coefficient of .17, with a \( t \) value of 8.45, reveals that, unlike the findings of the Reference Model, the relationship between physical inactivity and depression is not entirely spurious; however, compared to the Design Model, this beta value, controlling for age, has dropped almost 20%. Lowering the beta value is seen as an improvement over the Design Model since the Age Reference Model produced a small beta value (.02) between physical inactivity and depression.

The variance explained in depression increased slightly from .04 in the Design Model to .07 in this Age Proxy Model. Furthermore, .05 of the variance in physical inactivity was
explained by the direct effect of age, an estimate not attainable when physical inactivity was an exogenous variable in the Design Model. Although age is moderately correlated with the health and impairment constructs, the small change between the Design Model and the age proxy model makes it clear that age is not strongly correlated to physical inactivity and depression. This suggests that age does not capture the essential variance the three constructs explained in depression.

Although including age produced a slight improvement over the Design Model, there were still significant inaccuracies compared to the Reference Model. Using age as a proxy produced a model that overestimated the effect of physical inactivity on depression by .16 standardized units. Furthermore, compared to the Age Proxy Model, the $R^2$ for depression is almost eight times higher in the Reference Model and the $R^2$ for physical inactivity is over three times higher in the Reference Model. The beta value for the effect of age on depression ($\beta = .16, t = 8.09$) is lower than the beta for age on physical inactivity ($\beta = .22, t = 10.87$).

**Poor Representation of Proxy Variable: Wealth**

Wealth is another variable that has often been considered a confound in depression studies with SHARE data. Although wealth is frequently found to relate to health outcomes because of its correlation with access to care, etc., Verropoulou and Tsimbos (2007) found that this variable, measured by home ownership, quickly lost statistical significance in predicting depression when physical and mental health variables were added to their models. In fact, wealth was poorly related to the health and impairment constructs of the present study (Table 2). The use of wealth, therefore, demonstrates the effect of using a proxy that poorly represents the construct of interest.
The Wealth Reference Model (Figure 8) confirms the poor relationship between wealth and the health and disability constructs. The Wealth Proxy Model demonstrates the path coefficients when wealth is used as a proxy variable (Figure 9) and the lack of significant change in the direct effects of the health and impairment constructs on the design variables when wealth is added to the model. The large difference in fit indices between these two models show further demonstrate the ineffectiveness of wealth to represent the health and impairment constructs (Table 4).

When a poor proxy for health and impairment constructs were used, the relationship between physical inactivity and depression became similar to that observed for the Design Model, $\beta = .18, t = 7.5$, a drop in beta by only 3 standardized units. Furthermore, the $R^2$ for depression in this wealth proxy model (.04) is the same as that obtained in the Design Model.

Compared to the Reference Model, the obtained path coefficient for the effect of physical inactivity and depression was overestimated by .17 standardized units. Because wealth and physical inactivity do not correlate, wealth cannot, by definition, be a confound and must, therefore, only be viewed as just another predictor of depression. Including this poor proxy is mathematically akin to not using a control variable at all. This loss of precision is caused by the less potent effect that wealth has on physical inactivity and depression. The $R^2$ for depression in this model is also almost 14 times lower than the $R^2$ in the Reference Model, and the $R^2$ for physical inactivity in the wealth proxy model (.0003) is nowhere near this $R^2$ in the Reference Model (.16).

Although there are large differences in the beta values and $R^2$ for the wealth and age proxy variables on physical inactivity and depression, there is very little difference in the
parameter estimates for the primary research question (the effect of physical inactivity on depression) between these two proxy models, $\beta = .18$ (wealth) vs. .17 (age). This finding indicates that both variables do not represent the essential variance in health and disability constructs required for adequate statistical control. In other words, confound variables need to adjust for relevant variance captured by the constructs of primary interest, rather than just being correlated with the predictor and outcome variables in the model. This finding further demonstrates the importance of accurately specifying the essence of what control variables are intended to represent. Investigators must have a thorough understanding of the variance they need to control for, beyond simply knowing what variables are likely to be related to predictors and outcomes. As shown, age is related to both predictor and outcome, as well as the constructs of interest captured in the Reference Model; however, the specific variance that was needed to be removed to accurately estimate the relationship between physical inactivity and depression was not captured.

**Good Representation of a Proxy Variable: General Health**

I discussed above that age was moderately correlated with the health and impairment constructs that removed the specific variance necessary to accurately estimate the effect of physical inactivity on depression. Despite this moderate correlation, age did not capture this specific variance and, as a result, did not remove enough variance to reveal a spurious relationship between the two design variables. This specific variance appears to relate to general health. General health was measured on a five point Likert scale from poor to excellent with the question, “Would you say your health is...” The direct effects between general health and the design variables obtained in the General Health Reference Model show that there was
very little variance explained by general health in addition to the health and impairment
constructs (Figure 10). While general health is correlated with the three constructs, the mental
health and impairment constructs yield small path coefficients when the physical health
construct is estimated. The General Health Proxy Model is presented in Figure 11. The fit
indices for this model are much better than those obtained for the other proxy models (Table
5), showing that general health is a much better proxy variable than age and wealth. Unlike
these other proxy variables, general health partially removes spurious variance between
estimates of the effect of physical inactivity on depression (Figure 11). There is still a significant
association between these two design variables ($\beta = .11, t = 5.89$), but it is much smaller than
the parameter estimates in the Design Model ($\beta = .21, t = 10.39$). Despite this improvement, the
use of general health as a proxy variable overestimated the relationship between physical
inactivity and depression, compared to the Reference Model, by .10 standardized units. In
addition, the $R^2$ for depression is .22, substantially higher than the .04 obtained in the Design
Model. This $R^2$ for depression in the general health proxy model, however, is 2.5 times smaller
than the $R^2$ for depression in the Reference Model. The $R^2$ for physical inactivity with general
health modeled as a covariate is .05.

General health does a much better job in representing the desired variance of health
and impairment than age and wealth. The beta value for physical inactivity and depression
when I used general health as a proxy is about halfway between the high end obtained for the
wealth proxy model (.18) and the low value obtained for the Reference Model (.01).
Furthermore, the $R^2$ for depression is about three times higher when general health is used as a
proxy than when age or wealth is used as a proxy. The $R^2$ for physical inactivity is similar between the models using general health and age as a proxy.

Overall, general health is a good proxy for the health and impairment constructs modeled in the Reference Model. There is still a lot of variance that is not captured by general health, evidenced by the differences in beta and $R^2$ estimates; however, this proxy variable offers substantial improvements over the Design Model and the other models using less representative proxies.

Table 6 comprises a summary of beta weights of the pathways estimated from models with different proxy variables. It is clear that the better a proxy variable represents the health and impairment constructs required for adequate statistical control, the more accurate the beta estimates become. From this table, and Table 2 detailing the correlations of each proxy to the constructs they are intended to represent, we are able to quantify the effect of using imperfect proxy variables.

So what does it mean to use a good proxy variable? Many investigators will select a proxy variable because they do not have access to their desired variable, and their proxy is hypothetically correlated with the desired variable. Problems, however, can result when the investigators are not clear on what it is about the desired variable that they want to control for. The current demonstration illustrates that a good proxy represents specific confounding variance that needs to be removed from the estimated association between two design variables. The model using age as a proxy variable illustrated this finding clearly: Age was related to the general health constructs that needed to be modeled as confounds; however, age, did not capture the variance in these constructs that needed to be removed from physical
inactivity and depression. Investigators may be quick to add age as a confound because, intuitively, age is related to the physical and mental health and physical and mental impairment constructs. In this study, however, age does not appear to be related to the general health variance in these constructs and, therefore, did not remove this variance from the relationship between physical inactivity and depression. The inadequacy of age capturing specific general health variance may result from the large variability in individuals who age “successfully.” Brandt et al. (2012) identified a number of childhood SES and income factors relating to successful aging in older Europeans. This research supports the current findings that leading functional lives in late adulthood can be determined by variables independent of actual number of years one has been alive. It is, therefore, essential to identify what elements of age are associated with certain outcomes. What made general health a much better proxy variable than age is that it accounted for and removed specific health variance from the measured association between the two design variables. Ladin (2008) made a similar note in discussing that, although education and income are frequently correlated, each variable has unique components allowing them to link to and predict various outcomes. These finding speak to the importance of understanding exactly what elements need to be controlled in a model and then determining what variables, and to what degree these variables, capture this variance.
4. General Discussion

The first method presented in this paper demonstrated that confounds can have meaningful relationships within a network of design variables. The models are intended to illustrate some of the ways that confounds can shoulder a conceptual role in situations in which this role is not commonly recognized. The differences in parameter estimates between select models show how incorrectly specified confounds can have pronounced mathematical consequences. Even in the cases where the estimates either did not change or changed only slightly, we see that there are conflicts in conceptual interpretations that need to be resolved. In the simplest example, this resolution came in the form of reversing an arrow from a spurious relationship to a mediating one. However, in more complex models, we need to think about what concepts we need to hold constant.

The method of specifying a comprehensive network of relationships allows the investigator to gain considerable insight into how the system of variables behaves. An example of meaningful use of confounds can be seen with Ladin’s (2008) analysis of SHARE data to investigate the effects of education and additional SES variables on depression in aging Europeans. In addition to these design variables, demographic and health variables were included as control variables. These controls were modeled alongside the design variables in a path model to specify a network of influences on depression. Specifically, age, gender, and chronic diseases were defined to have a direct effect on education, income, and a few additional health-related variables. These endogenous variables were then hypothesized to either directly or indirectly cause depression. Although these relationships were not analyzed simultaneously in an SEM, the example is still relevant in terms of how confounds were given
substantive meaning and interpreted as an integral component of the explanatory model. Such clarity resulting from adequate conceptualization of confound variables can help the investigator generate auxiliary hypotheses, as the origin of unanticipated results can be easily identified. Finally, by granting confounds meaningful status in structural models, investigators can avoid making inaccurate scientific conclusions that may arise when essential model components are dismissed. Some of the problems that can lead to inaccurate conclusions are:

1. Misspecification of causality (endogenous variables specified as exogenous and vice versa, exogenous variables influencing each other)

2. Controlling for confounds that cannot be conceptually removed from the dynamic of interest (i.e., the confound is an inherent feature of the predictor).

Dennis et al. (2009) noted that IQ is often controlled statistically or experimentally to get a cleaner estimate of the relationship between cognition and neurodevelopmental disorders. Only by understanding the reticulate nature of IQ and such neurological disorders will an investigator know that separating such effects is scientifically invalid.

3. A variable may be (correctly) identified as a confound, but its actual influence in yielding spurious correlations among design variables may be overstated.

The second component of the paper identified the need to understand the conceptual and operational meaning of proxy variables. When a proxy variable is used, we must determine how this variable parallels the effects of the latent construct. When a latent construct has a strong influence on other measured variables, we must take extra precautions to ensure that variables representing this construct share its potency. As demonstrated above, the effects,
both mathematical and conceptual, that an imperfect proxy can have are substantial, especially when the imperfect proxy is claimed or intended to represent the primary confound of interest perfectly. Additionally, the question must be asked if researchers are even aware that they are using an imperfect proxy. Many investigators including confounds in a model make no mention of the existence of underlying constructs that are really causing the spurious correlations.

Silvia (2008), for instance, found that openness to experience partially accounted for the effects of intelligence on creativity. Although lack of research uniting personality and intelligence was cited, this author did not try to explain what underlying factors inherent in openness to experience permit it to partially confound the relationship between intelligence and creativity. Furthermore, investigators who throw standard control variables such as race and SES into a model should consider, beyond just their correlations with outcomes, how these variables interact with other design variables and whether or not these variables are being used to represent a larger construct (e.g., access to resources, social inequality). If a scientific problem calls for an estimate of an underlying construct, and this construct is not adequately represented by a proxy variable, then investigators must concede that their parameter estimates may be biased by virtue of an imperfect proxy.

Golden et al. (1982), for example, estimating the relationship between intelligence of brain damaged patients and degree of brain damage, determined that measures of pre-damage intelligence were required for adequate statistical control. In the absence of any possible measure of pre-damage intelligence, the authors used pre-damage education as a proxy. Their response to using this measure is appealing, in that they not only recognized education as an imperfect proxy, but they assessed how it would likely overestimate the relationship between
measured intelligence in brain damaged patients and degree of affliction. This type of responsible qualification is necessary when we work with limited data.

Realizing the underlying influences that interact in a network of variables is essential for accurately choosing relevant proxy variables and specifying models. This understanding will allow us to quantify the correct necessary to estimate and adjust for the effect of an imperfect proxy variable. Furthermore, we can better examine if proxy variables are pulling in additional, unwanted variance if we are clear on how the proxy variable relates to the construct of interest. Detailing the relationships between latent constructs and measured variables should help our analytic methods become concordant with our conceptual hypotheses. We are, therefore, getting more accurate estimates by precisely modeling relationships that are scientifically credible and hypothesized to exist. Ultimately, this adds up to getting more information out of our data.

Limitations

Because many of the interpretations in this work relied on comparisons to a “Reference Model,” I proceeded under the working assumption that this Reference Model was correctly and fully specified. This, of course, was not the case, as I was limited by the data available in the SHARE dataset. The specific degree to which imperfect proxy variables over- or under-estimated path coefficients, although accurate in a relative sense, were made relative to a model that was not complete. This inability to obtain a true Reference Model speaks to a central point in this work, namely, that validity coefficients of true underlying constructs of interest are rarely known. Researchers will always use age, income, education, etc. as proxy confounds because some purer variance of interest is not available in existing data sets or is too
difficult to measure. These very reasons that constructs of interest are not available are, however, the reasons we can only estimate, using theory and previous research, the true validity coefficient. As we have a better conceptualization of the intended constructs we wish to model, we can prospectively plan to measure such variance to reduce their reliance on potentially imperfect proxy variables. Even in circumstances where measurements will never be available (i.e., prospective planning will not be of benefit), published research can provide an excellent resource from which to examine the specific variance that needs to be modeled. This understanding can help us evaluate the effectiveness of available proxies and guide our formulation of the necessary statistical correction.

Another limitation was the small range of correlations between the latent constructs and the proxy variables. This restricted range led, in part, to path coefficients that did not always differ dramatically when poor and moderate proxies were used (as discussed above, a large reason for this lack of difference is because the desired variance was simply not removed by the proxy variable). It would have been ideal, at least for demonstration purposes, to have proxy variables comprising correlations with latent traits that ranged from about .10 to .85. Similarly, many of the variables included in the models did not have a substantial correlation with the primary outcome, depression. This finding may result from the fact that there are varied influences on depression in the elderly (see citations of SHARE research above) and selecting a handful of physical and mental health measures may not be sufficient to explain the observed variance in depression. While the SHARE dataset is relatively large, the lack of a method to prospectively measure influences on depression likely contributed to incomplete data. Furthermore, when we evaluate patterns from small path coefficients, changes in
parameter estimates may seem more dramatic when they drop to near zero than when they drop to a range that is likely to be statistically significant. The limitations of this dataset should not detract from the main points of this dissertation; however, they are worth considering when estimating the effects that confound misspecification can have in other investigations.
APPENDIX A: TABLES AND GRAPHS

Table 1. Alphas of items for constructs.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Health</td>
<td>.86</td>
</tr>
<tr>
<td>Mental Health</td>
<td>.48</td>
</tr>
<tr>
<td>Physical and Mental Impairment</td>
<td>.76</td>
</tr>
</tbody>
</table>
Table 2. Correlations of proxy variables with constructs they are intended to represent.

<table>
<thead>
<tr>
<th></th>
<th>Physical Health</th>
<th>Mental Health</th>
<th>Physical/ Mental Impairment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.34</td>
<td>.22</td>
<td>.44</td>
</tr>
<tr>
<td>Wealth</td>
<td>&lt;.01</td>
<td>-.06</td>
<td>-.04</td>
</tr>
<tr>
<td>General Health</td>
<td>.75</td>
<td>.31</td>
<td>.51</td>
</tr>
</tbody>
</table>
Table 3. Fit indices for the Age Reference Model and the Age Proxy Model.

<table>
<thead>
<tr>
<th>Fit Index</th>
<th>Age Reference Model</th>
<th>Age Proxy Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$ (df, p)</td>
<td>1190.43 (3, &lt;.001)</td>
<td>3222.30 (9, &lt;.001)</td>
</tr>
<tr>
<td>GFI</td>
<td>.86</td>
<td>.24</td>
</tr>
<tr>
<td>RMSEA (90% CI)</td>
<td>.41 (.39, .43)</td>
<td>.39 (.38, .40)</td>
</tr>
<tr>
<td>CFI</td>
<td>.72</td>
<td>.25</td>
</tr>
<tr>
<td>NFI</td>
<td>.72</td>
<td>.25</td>
</tr>
<tr>
<td>NNFI</td>
<td>-.39</td>
<td>-.25</td>
</tr>
</tbody>
</table>
Table 4. Fit Indices for the Wealth Reference Model and the Wealth Proxy Model.

<table>
<thead>
<tr>
<th>Fit Index</th>
<th>Wealth Reference Model</th>
<th>Wealth Proxy Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^2$ (df, $p$)</td>
<td>1058.83 (3, &lt;.001)</td>
<td>2624.69 (9, &lt;.001)</td>
</tr>
<tr>
<td>GFI</td>
<td>.83</td>
<td>.64</td>
</tr>
<tr>
<td>RMSEA (90% CI)</td>
<td>.46 (.44, .48)</td>
<td>.42 (.41, .43)</td>
</tr>
<tr>
<td>CFI</td>
<td>.61</td>
<td>.02</td>
</tr>
<tr>
<td>NFI</td>
<td>.61</td>
<td>.03</td>
</tr>
<tr>
<td>NNFI</td>
<td>-.97</td>
<td>-.63</td>
</tr>
</tbody>
</table>
Table 5. Fit Indices for the General Health Reference Model and the General Health Proxy Model.

<table>
<thead>
<tr>
<th>Fit Index</th>
<th>General Health Reference Model</th>
<th>General Health Proxy Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^2$ (df, $p$)</td>
<td>N/A*</td>
<td>1603.98 (&lt;.001)</td>
</tr>
<tr>
<td>GFI</td>
<td>1.0*</td>
<td>.85</td>
</tr>
<tr>
<td>RMSEA (90% CI)</td>
<td>N/A*</td>
<td>.34 (.33, .35)</td>
</tr>
<tr>
<td>CFI</td>
<td>1.0*</td>
<td>.72</td>
</tr>
<tr>
<td>NFI</td>
<td>1.0*</td>
<td>.72</td>
</tr>
<tr>
<td>NNFI</td>
<td>N/A*</td>
<td>.30</td>
</tr>
</tbody>
</table>

* Estimated from a just-identified model.
Table 6. Summary of path coefficients ($\beta$) across the four different models.

<table>
<thead>
<tr>
<th>Reference Model</th>
<th>Age (Design Proxy)</th>
<th>Wealth (Poor Proxy)</th>
<th>General Health (Strong Proxy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Inactivity to Depression</td>
<td>.01</td>
<td>.17</td>
<td>.18</td>
</tr>
<tr>
<td>Confound to Physical Inactivity</td>
<td>N/A</td>
<td>.22</td>
<td>-.02</td>
</tr>
<tr>
<td>Confound to Depression</td>
<td>N/A</td>
<td>.16</td>
<td>-.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.43</td>
</tr>
</tbody>
</table>
Figure 1. Design Model estimating the effect of physical inactivity on depression.
Figure 2. Model specifying spurious relationship between physical inactivity and depression.
Figure 3. Model specifying physical disability as a mediator of physical inactivity and depression.
Figure 4. Model specifying multiple confounds.

- Physical Inactivity
- Depression
- Physical Disability
- Bothered by Symptoms

Arrows and coefficients:
- Physical Inactivity to Physical Disability: .39
- Physical Inactivity to Depression: .24
- Physical Disability to Bothered by Symptoms: .66
- Depression to Bothered by Symptoms: .39
Figure 5. Reference Model.

- Physical Inactivity
- Mental Health
- Depression

Connections:
- Physical Health (0.86)
- Physical/ Mental Impairment (0.31)
- Hopes for Future (0.81)
- Rather be Dead (0.81)
- Chronic Health Conditions (0.82)
- Bothered by Symptoms (0.75)
- General Health (0.73)
- Number Medications (0.85)
- Long Term Illness (0.89)
- ADL (0.90)
- IADL (0.16)

Path Coefficients:
- 0.42
- 0.51
- 0.01
Figure 6. Age Reference Model.
Figure 7. Age Proxy Model.
Figure 8. Wealth Reference Model.
Figure 9. Wealth Proxy Model.

- Physical Inactivity
- Chronic Health Conditions
- Rather be Dead
- Hopes for Future
- Bothered by Symptoms
- General Health
- Number Medications
- Long Term Illness
- ADL
- IADL
- Mental Health
- Depression
- Physical Health
- Physical/Mental Impairment

Correlation coefficients:
- Physical Inactivity to Wealth: 0.18
- Physical Health to Depression: 0.07
- Physical Health to ADL: 0.73
- Physical Health to IADL: 0.90
- Mental Health to Hopes for Future: 0.81
- Mental Health to Rather be Dead: 0.81
- Mental Health to Physical Health: 0.86
- Physical Health to Chronic Health Conditions: 0.82
- Physical Health to Bothered by Symptoms: 0.75
- Physical Health to General Health: 0.85
- Physical Health to Number Medications: 0.89
- Physical Health to Long Term Illness: 0.90
- Depression to Physical Health: 0.04
- Depression to Mental Health: 0.06
- Depression to Physical Inactivity: -0.02
- Depression to Chronic Health Conditions: -0.07
- Depression to Rather be Dead: -0.06
- Depression to Hopes for Future: -0.01
Figure 10. General Health Reference Model.
Figure 11. General Health Proxy Model.
APPENDIX B: PHYSICAL HEALTH QUESTIONS FROM SHARE SURVEY USED IN STRUCTURAL MODELS

Labels for each question correspond to the SHARE Questionnaire Codebook.

**PH006_ DOCTOR TOLD YOU HAD CONDITIONS**

Please look at card 6. Has a doctor ever told you that you had any of the conditions on this card? Please tell me the number or numbers of the conditions.

1. A heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure
2. High blood pressure or hypertension
3. High blood cholesterol
4. A stroke or cerebral vascular disease
5. Diabetes or high blood sugar
6. Chronic lung disease such as chronic bronchitis or emphysema
7. Asthma
8. Arthritis, including osteoarthritis, or rheumatism
9. Osteoporosis
10. Cancer or malignant tumour, including leukaemia or lymphoma, but excluding minor skin cancers
11. Stomach or duodenal ulcer, peptic ulcer
12. Parkinson disease
13. Cataracts
14. Hip fracture or femoral fracture
15. None
16. Other conditions, not yet mentioned

**PH010_ BOTHERED BY SYMPTOMS**

Please look at card 7. For the past six months at least, have you been bothered by any of the health conditions on this card? Please tell me the number or numbers.

1. Pain in your back, knees, hips or any other joint
2. Heart trouble or angina, chest pain during exercise
3. Breathlessness, difficulty breathing
4. Persistent cough
5. Swollen legs
6. Sleeping problems
7. Falling down
8. Fear of falling down
9. Dizziness, faints or blackouts
10. Stomach or intestine problems, including constipation, air, diarrhoea
11. Incontinence or involuntary loss of urine
12. None
13. Other symptoms, not yet mentioned

**PH002_ HEALTH IN GENERAL QUESTION V 2**

Would you say your health is ....

1. Excellent
2. Very good
3. Good
4. Fair
5. Poor

**PH011_ CURRENT DRUGS AT LEAST ONCE A WEEK**

Our next question is about the medication you may be taking. Please look at card 8. Do you currently take drugs at least once a week for problems mentioned on this card?

1. Drugs for high blood cholesterol
2. Drugs for high blood pressure
3. Drugs for coronary or cerebrovascular diseases
4. Drugs for other heart diseases
5. Drugs for asthma
6. Drugs for diabetes
7. Drugs for joint pain or for joint inflammation
8. Drugs for other pain (e.g. headache, backpain, etc.)
9. Drugs for sleep problems
10. Drugs for anxiety or depression
11. Drugs for osteoporosis, hormonal
12. Drugs for osteoporosis, other than hormonal
13. Drugs for stomach burns
14. Drugs for chronic bronchitis
15. None
16. Other drugs, not yet mentioned

**PH004_ LONG-TERM ILLNESS**

Some people suffer from chronic or long-term health problems. By long-term we mean it has troubled you over a period of time or is likely to affect you over a period of time. Do you have any long-term health problems, illness, disability or infirmity?

1. Yes
2. No
APPENDIX C: MENTAL HEALTH QUESTIONS FROM SHARE SURVEY USED IN STRUCTURAL MODELS

Labels for each question correspond to the SHARE Questionnaire Codebook.

MH003_ HOPES FOR THE FUTURE

What are your hopes for the future?

1. Any hopes mentioned
2. No hopes mentioned

MH004_ FELT WOULD RATHER BE DEAD

In the last month, have you felt that you would rather be dead?

1. Any mention of suicidal feelings or wishing to be dead
2. No such feelings
APPENDIX D: ACTIVITIES OF DAILY LIVING QUESTIONS (ADL) AND INSTRUMENTAL ACTIVITIES OF DAILY LIVING (IADL) QUESTIONS FROM SHARE SURVEY USED IN STRUCTURAL MODELS

Labels for each question correspond to the SHARE Questionnaire Codebook.

**PH048_ HEALTH AND ACTIVITIES (ADL)**

Please look at card 9. We need to understand difficulties people may have with various activities because of a health or physical problem. Please tell me whether you have any difficulty doing each of the everyday activities on card 9. Exclude any difficulties that you expect to last less than three months. (Because of a health problem, do you have difficulty doing any of the activities on this card?)

1. Walking 100 metres
2. Sitting for about two hours
3. Getting up from a chair after sitting for long periods
4. Climbing several flights of stairs without resting
5. Climbing one flight of stairs without resting
6. Stooping, kneeling, or crouching
7. Reaching or extending your arms above shoulder level
8. Pulling or pushing large objects like a living room chair
9. Lifting or carrying weights over 10 pounds/5 kilos, like a heavy bag of groceries
10. Picking up a small coin from a table
11. None of these

**PH049_ MORE HEALTH AND ACTIVITIES (IADL)**

Please look at card 10. Here are a few more everyday activities. Please tell me if you have any difficulty with these because of a physical, mental, emotional or memory problem. Again exclude any difficulties you expect to last less than three months. (Because of a health or memory problem, do you have difficulty doing any of the activities on card 10?)

1. Dressing, including putting on shoes and socks
2. Walking across a room
3. Bathing or showering
4. Eating, such as cutting up your food
5. Getting in or out of bed
6. Using the toilet, including getting up or down
7. Using a map to figure out how to get around in a strange place
8. Preparing a hot meal
9. Shopping for groceries
10. Making telephone calls
11. Taking medications
12. Doing work around the house or garden
13. Managing money, such as paying bills and keeping track of expenses
14. None of these
APPENDIX E: QUESTIONS USED TO CONSTRUCT THE EURO-D SCALE

Labels for each question correspond to the SHARE Questionnaire Codebook.

**MH002_ SAD OR DEPRESSED LAST MONTH**
In the last month, have you been sad or depressed? IWER: IF PARTICIPANT ASKS FOR CLARIFICATION, SAY 'BY SAD OR DEPRESSED, WE MEAN MISERABLE, IN LOW SPIRITS, OR BLUE'

1. Yes
2. No

**MH003_ HOPES FOR THE FUTURE**
What are your hopes for the future? IWER: NOTE ONLY WHETHER HOPES ARE MENTIONED OR NOT

1. Any hopes mentioned
2. No hopes mentioned

**MH004_ FELT WOULD RATHER BE DEAD**
In the last month, have you felt that you would rather be dead?

1. Any mention of suicidal feelings or wishing to be dead
2. No such feelings

**MH005_ FEELS GUILTY**
Do you tend to blame yourself or feel guilty about anything?

1. Obvious excessive guilt or self-blame
2. No such feelings
3. Mentions guilt or self-blame, but it is unclear if these constitute obvious or excessive guilt or self-blame
**MH006_ BLAME FOR WHAT**

So, for what do you blame yourself? IWER: NOTE - ONLY CODE 1 FOR AN EXAGGERATED FEELING OF GUILT, WHICH IS CLEARLY OUT OF PROPORTION TO THE CIRCUMSTANCES. THE FAULT WILL OFTEN HAVE BEEN VERY MINOR, IF THERE WAS ONE AT ALL. JUSTIFIABLE OR APPROPRIATE GUILT SHOULD BE CODED 2.

1. Example(s) given constitute obvious excessive guilt or self-blame
2. Example(s) do not constitute obvious excessive guilt or self-blame, or it remains unclear if these constitute obvious or excessive guilt or self-blame

**MH007_ TROUBLE SLEEPING**

Have you had trouble sleeping recently?

1. Trouble with sleep or recent change in pattern
2. No trouble sleeping

**MH008_ LESS OR SAME INTEREST IN THINGS**

In the last month, what is your interest in things?

1. Less interest than usual mentioned
2. No mention of loss of interest
3. Non-specific or uncodeable response

**MH009_ KEEPS UP INTEREST**

So, do you keep up your interests?

1. Yes
2. No

**MH010_ IRRITABILITY**

Have you been irritable recently?

1. Yes
2. No
**MH011_ APPETITE**

What has your appetite been like?

1. Diminution in desire for food
2. No diminution in desire for food
3. Non-specific or uncodeable response

**MH012_ EATING MORE OR LESS**

So, have you been eating more or less than usual?

1. Less
2. More
3. Neither more nor less

**MH013_ FATIGUE**

In the last month, have you had too little energy to do the things you wanted to do?

1. Yes
2. No

**MH014_ CONCENTRATION ON ENTERTAINMENT**

How is your concentration? For example, can you concentrate on a television programme, film or radio programme?

1. Difficulty in concentrating on entertainment
2. No such difficulty mentioned

**MH015_ CONCENTRATION ON READING**

Can you concentrate on something you read?

1. Difficulty in concentrating on reading
2. No such difficulty mentioned
**MH016_ ENJOYMENT**

What have you enjoyed doing recently?

1. Fails to mention any enjoyable activity
2. Mentions ANY enjoyment from activity

**MH017_ TEARFULNESS**

In the last month, have you cried at all?

1. Yes
2. No
APPENDIX F: PHYSICAL ACTIVITY QUESTIONS FROM SHARE SURVEY USED IN STRUCTURAL MODELS

Labels for each question correspond to the SHARE Questionnaire Codebook.

**BR015_ SPORTS OR ACTIVITIES THAT ARE VIGOROUS**

We would like to know about the type and amount of physical activity you do in your daily life. How often do you engage in vigorous physical activity, such as sports, heavy housework, or a job that involves physical labour?

1. More than once a week
2. Once a week
3. One to three times a month
4. Hardly ever, or never

**BR016_ ACTIVITIES REQUIRING A MODERATE LEVEL OF ENERGY**

How often do you engage in activities that require a low or moderate level of energy such as gardening, cleaning the car, or doing a walk?

1. More than once a week
2. Once a week
3. One to three times a month
4. Hardly ever, or never
APPENDIX G: GLOSSARY OF TERMS

**Confound**: A variable that has a simultaneous direct effect on an independent variable and the dependent variable that the independent variable is modeled to predict.

**Misspecification**: The use of a proxy variable that does not capture the same variance as the construct it is intended represent. This proxy variable, in other words, is not similar to the intended construct and should not be used to represent this construct.

**Proxy Variable**: A measured variable that is used to represent an unmeasured variable or construct. This proxy variable is often used in a statistical model when the unmeasured variable is not available.

**Spurious Correlation**: An observed direct effect of an independent variable on a dependent variable where the independent variable does not in fact cause the dependent variable. Rather, this observed effect is a result from another variable having a simultaneous direct effect on both the independent variable and dependent variable, and when this third variable is modeled, the observed direct effect between the independent variable and the dependent variable becomes statistically non significant.
REFERENCES


doi:10.1073/pnas.1018601108


doi:http://dx.doi.org.ezproxy2.library.arizona.edu/10.1177/1477370810376569


doi:10.1136/jech.2007.066167


doi:10.1177/0898264308321002


doi:10.1089/jwh.2009.1394