A Mega-Analysis of Personality Predictions: Robustness and Boundary Conditions

Emorie D. Beck

Washington University in St. Louis

Short Abstract

Personality predicts important life outcomes. However, the presence, direction, and magnitude of personality-outcome associations vary cross-study. I mega-analytically examine their robustness using 10 longitudinal studies, 14 personality predictors, 14 outcomes, and seven moderators using propensity score matching and specification curve analysis to test boundary conditions and reduce selection bias.

**A Mega-Analysis of Personality Predictions: Robustness and Boundary Conditions**

Personality traits are relatively stable, dispositional patterns that differentiate people from one another (Roberts, Wood, & Caspi, 2008). Moreover, personality traits predict many important life outcomes, such as marriage (Kelly & Conley, 1987; Malouff, Thorsteinsson, Schutte, Bhullar, & Rooke, 2010; Specht, Egloff, & Schmukle, 2011), life expectancy (Jackson, Connolly, Garrison, Leveille, & Connolly, 2015; Jokela et al., 2013; Martin, Friedman, & Schwartz, 2007; Turiano, Chapman, Gruenewald, & Mroczek, 2015), and health (Hampson, 2012; Weston, Hill, & Jackson, 2015). Much research has examined what personality predicts (see Ozer & Benet Martinez, 2006, and Soto, 2019, for reviews), including what the strongest predictors of different outcomes are (see Roberts, Kuncel, Shiner, Caspi, & Goldberg., 2007).

Despite numerous studies and meta-analyses that demonstrate personality prospectively predicts consequential life outcomes, most studies fail to capture true selection effects for two main reasons: reverse causality as well as a combination of researcher degrees of freedom, publication bias, and a lack of understanding of boundary conditions. Reverse causality could occur because many of outcomes that personality prospectively predicts also prospectively predict personality. For example, Conscientiousness predicts both work success (Judge, Higgins, Thoresen, & Barrick, 1999; Rothmann & Coetzer, 2003) and health (e.g. Weston, Hill, & Jackson, 2015). However, socioeconomic status also predicts both work success (e.g. Ensminger, Fothergill, Bornstein, & Bradley, 2003) and health (e.g. Adler, Boyce, Chesney, Folkman, & Syme, 1993), and work success and health also predict each other (e.g. Conger & Donnellan, 2007). Thus, the direction of the relationships among personality, consequential outcomes, and potential moderators make it difficult to tease apart their relationships. Thus, the existent work examining selection effects of personality on life events cannot disentangle selection effects of personality from selection bias of other background characteristics, like health, work factors, and socioeconomic status, which may explain why evidence of personality predicting life events has been mixed. In other words, people who experience the life events that personality predicts might differ from those who do not in important ways that also influence personality.

In addition, researcher degrees of freedom occur for a number of reasons. One important indicator of researcher degrees of freedom is publication bias, which favors results that are surprising or find evidence for links between predictors and outcomes (e.g. Franco, Malhotra, & Simonovits, 2014). In other words, it is possible that the specifications of published personality-outcome models favor significant results at the expense of replicable, reliable, and generalizable results. Although recent work indicates that past work linking personality and outcomes cross-sectionally are highly reproducible (see Soto, 2019), this merely means that similar results can be found given similar procedural and analytic choices. However, reproducibility means the results are reliable, not valid, which is different than either saying that these results are robust to other procedural or analytic choices or from saying that there is a true causal relationship between personality and outcomes.

Moreover, the published personality prediction literature has largely tackled basic prediction questions, rather than testing the boundaries of such prediction. Indeed, arbitrarily chosen covariates coupled with publication bias may account for the observation that across different studies, both what personality predicts and the strength with which personality predicts outcomes is not always consistent. As a result, despite seemingly reliable patterns of prediction, we have almost no evidence concerning the boundaries of and moderators of personality prediction, which is a critical concern in understanding both what personality predicts and why it does so.

In the present study, using 10 longitudinal panel studies, I will examine whether 14 measures of personality at a baseline assessment predict the future experience of 14 life events and outcomes that occur throughout the lifespan independently of selection bias of other background characteristics and of procedural and analytic procedures. Previous work has examined selection effects of personality on life events using these samples, but the present study (1) tests whether selection effects persist after accounting for background characteristics via propensity score matching and (2) specification curve techniques to examine the different boundary conditions where personality-outcome associations exist, each (3) using a mega-analytic procedure to test the robustness of effects.

**Selection Effects**

Personality is reliable predictor of a number of outcomes, including major life events, such as divorce (e.g. Solomon & Jackson, 2014; Specht et al., 2011), career success (e.g. Judge, Higgins, Thoresen, & Barrick, 1999; Rothmann & Coetzer, 2003), and major health events (e.g. Weston, Hill, & Jackson, 2015). More Extraverted people are more likely to move in with a partner (Specht et al., 2011), to have children (van Scheppingen et al., 2016), and to enter into romantic relationships (Wagner, Becker, Lüdtke, & Trautwein, 2015), while more Agreeable people are more likely to become unemployed and less likely to separate from a partner (Specht et al., 2011) or to enter into military service (Jackson, Thoemmes, Jonkmann, Lüdtke, & Trautwein, 2012). Conscientiousness and Neuroticism, in particular, have been linked to health, both cross-sectionally (e.g. Hampson, 2012) and longitudinally (e.g. Weston et al., 2015). In one study, for example, Conscientiousness predicted the onset of high blood pressure, diabetes, stroke, and arthritis, with Conscientiousness serving as a protective factor against each of these health conditions (Weston et al., 2015).

But there are a number of reasons to interpret selection effects of personality and life events with caution. First, most studies examine cross-sectional group differences among those who have or have not experienced events, making it impossible to tease apart whether personality predicts life events or life events predict personality (i.e. reverse causality). For example, people who are more Extraverted are more likely to have social and enterprising occupational interests (Barrick, Mount, & Judge, 2001; Larson, Rottinghaus, & Borgen, 2002), have fewer cardiovascular problems (Miller, Smith, Turner, Guijarro, & Hallet, 1996), and to start romantic relationships (Wagner et al., 2015). But there is longitudinal evidence that work (e.g. Lüdtke, Roberts, Trautwein, & Nagy, 2011), chronic illness (e.g. Mueller, Wagner, & Gerstorf, 2017) and romantic relationships (e.g. Mund & Neyer, 2014; Neyer, Mund, Zimmermann, & Wrzus, 2014) may influence personality, which calls into question whether personality influences who experiences life events or vice versa. Without understanding the direction of the relationship, reverse causality remains a possible explanation of observed personality-outcome associations – that is, they are not selection effects because personality predicts experiences and is not simply associated with them.

Second, and perhaps most critically, almost no studies account for baseline factors that may influence both personality and the likelihood of experiencing an event. Personality has been linked to a number of demographic and background factors, including socioeconomic status (Roberts et al., 2007), cognitive ability (e.g. Moutafi, Furnham, & Paltiel, 2005), age (e.g. Donnellan & Lucas, 2008; Soto, John, Gosling, & Potter, 2011), parental education (e.g. Sutin, Luchetti, Stephan, Robins, & Terracciano, 2017), marital satisfaction (e.g. Kelly & Conley, 1987; Malouff et al., 2010), health (e.g. Hampson, 2012; Roberts et al., 2007), and geographic region (e.g. Rentfrow, Jokela, & Lamb, 2015), among others. Although most studies have controlled for a small number of these background characteristics, particularly age and gender (e.g. Specht et al., 2011), this is problematic. The small sets of background characteristics do not account for selection bias based on other characteristics that personality has been linked to. Indeed, in the small number of studies that have accounted for broader ranges of background characteristics (Jackson et al., 2012; Nieß & Zacher, 2015; van Scheppingen et al., 2016; Wagner et al., 2015), personality-related selection effects have been much more limited. In one study, for example, single young adults were less extraverted and had lower self-esteem than young adults who entered into romantic relationships (Wagner et al., 2015). After accounting for selection bias through matching, single young adults had higher mean self-esteem than those in relationships. In other words, matching individuals on background characteristics can greatly impact personality selection effects on life events – in *direction* as well as magnitude. Without matching, selection effects that are seemingly driven by personality may be driven by the direct or indirect influence of other factors.

Third, which background characteristics are controlled for are often somewhat arbitrary and may introduce many researcher degrees of freedom when not consistent across studies. In other words, one study may control for age and gender and find no effect of personality predicting an outcome, while another study using the same or a different sample may find different results when also controlling for SES. Such inconsistency in which covariates are included can produce inconsistent results and make true hypothesis tests difficult. Recently, Simonsohn and colleagues (2015) proposed specification curves as a possible solution to inconsistencies in model specification (Simonsohn, Simmons, & Nelson, 2015). Importantly, specification curves make it possible not just to test how different sets of covariates influence results but also how other specifications, including how indicators are operationalized and even the form of a model, influence results. By testing all possible combinations of reasonable and theoretically valid tests, researchers can test the robustness of an effect by assuming that robust associations should persist across specifications. Thus, in the present study, I can test how the selection of covariates influences the relationship between personality and outcomes across all studies as well as within specific studies.

Fourth, the arbitrariness with which covariates are selected has resulted in unclear boundary conditions of when and for whom personality predicts important outcomes. Indeed, a survey of the literature on essentially any personality-outcome association will reveal that most to all will include one to five covariates. Some, like age and gender, tend to be included in most, but others, like socioeconomic status, health status (in many distinct forms), marriage status, and many others tend to be represented inconsistently and much less often. The result is that it personality appears to predict outcomes in some, but not all investigations, but there is no understanding of why this is the case and arbitrary covariate inclusion cannot be eliminated as a cause. Thus, there is no understanding of the boundary conditions of when, how, and for whom personality predicts outcomes, which is imperative for understanding why this occurs and using it potentially change negative or create positive outcomes.

In the present study, I address concerns about reverse causality, researcher degrees of freedom, and the boundary conditions of personality-outcome associations by examining selection effects of 14 personality characteristics (1) in 14 broad outcomes and life events (2) longitudinally (3) while accounting for more than 50 background characteristics in 10 large longitudinal panel studies. In addition, because of the large sample size, both within studies as well as across all of them, I will be able to test important moderators of selection effects. Doing so will allow me to more closely approximate personality’s unique role in predicting relatively common life experiences and events as well as its consistency across samples, which I argue will push research toward better understanding the mechanisms through which personality influences the life course.

Given that life events cannot be experimentally manipulated and subjective evaluations of life events are often unavailable, research examining life events and personality prediction must use alternative methods to improve upon previous work. Given that personality both predicts who experiences specific events (e.g. Neyer & Lehnart, 2007; Specht et al., 2011) as well as behavioral responses to them (Turiano, Hill, Roberts, Spiro III, & Mroczek, 2012; Weston & Jackson, 2015, 2016), teasing apart selection bias from selection effects using propensity score matching may partially approximate teasing apart how personality uniquely predicts individuals’ experiences and outcomes. Moreover, by using a specification curve technique, I will be able to observe which indicators may most strongly account for selection bias and test whether observed personality-outcome associations persist reliably above chance levels regardless of covariate inclusion. Thus, together, both propensity score matching and specification curve analysis will allow me address both the overall robustness of personality-outcome associations, as well as to identify characteristics that may threaten the inferences made when testing such associations.

**Processes of Selection Effects**

Traits are often thought to influence important life outcomes, like health and relationship satisfaction, through accumulation – that is, traits influence outcomes indirectly through the behaviors that traits also predict (e.g. Hampson, Edmonds, Goldberg, Dubanoski, & Hillier, 2015) . But to understand how personality might influence the events that people experience through behavior requires a true understanding of what the relationship between personality traits and behaviors is. Traits, as they are traditionally considered in trait-outcome associations, are relatively stable patterns of thoughts, beliefs, and behavior that differentiate people from one another (a nomothetic approach). This alone is descriptive, describing how people differ from one another on a number of core dimensions rather than *why* they differ, which means that traits are not good causal candidates for *why* people engage in certain behaviors. Instead, understanding why people differ from one another – that is, why core dimensions of personality can be reliably recovered – requires first understanding a person (an idiographic approach; Beck & Jackson, 2019ab; Borsboom, 2003; Molenaar, 2004). In part, this is due to the observation that the structure of personality within a person, measured as states or behaviors repeatedly measured over short intervals, rarely adheres to the between person structure (Beck & Jackson, 2019ac; Borkenau & Ostendorpf, 1998; Molenaar, 2004). Instead, within a person, these states and behaviors form unique combinations and patterns. In other words, individuals appear to exhibit unique traits, much as the “father” of personality, Gordon W. Allport, suggested nearly a century ago (Allport, 1937).

Thus, because nomothetic personality traits do not appear to exhibit “top down” influence on behaviors, another explanation is needed to understand how traits are linked to behavior. But there have been few attempts to examine a more “bottom up” approach, in which traits are the outcome of patterns of behavior rather than the other way around, despite several calls for such approaches (Baumert et al., 2017; Beck & Jackson, 2019ab; Borsboom et al., 2003; Cramer et al., 2012; Mischel & Shoda, 1995). Recently, I expanded on previous proposals that emergence underscores the link between within- and between-person personality structure (Baumert et al., 2017; Cramer et al., 2012) by differentiating idiographic, within-person, and between-person approaches to personality as well as the links between them (Beck, 2019). Moreover, I argued that doing so has important implications for description, prediction, and explanation, with idiographic, within-, and between-person approaches having different utility and applicability at different levels. A summary of these levels is shown in Table 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 1  *Summary of Idiographic, Within-Person, and Nomothetic Levels* | | | | | |
| **Level** | **Utility** | **Bandwidth** | **Units** | **Examples** | **Methods** |
| Idiographic | Explanation | Micro | Unique socioemotional cognitive *processes* | Affect  Attention | Dynamical Systems |
| Within-Person | Description,  Prediction | Narrow | Unique Socioemotional behavioral *states* | Emotions  Behaviors  Beliefs | P-technique; Vector Autoregressive Models |
| Between-Person | Description,  Prediction | Broad | Shared aggregated socioemotional behavioral *tendencies* | The Big 5  HEXACO | R-technique  Network Psychometrics |

Specifically, I argue that starting with the idiographic level, each of these levels sequentially produces the next. In other words, the idiographic level underlies the within-person level, which in turn underlies the nomothetic, between-person level. Importantly, just as Allport (1937) argued, only the idiographic level has causal properties. The within- and between-person levels themselves have no causal properties and are better described as different levels of aggregation, as described by Cattell’s (1946) data box, that are incredibly useful for description and prediction. Instead,nomothetic personality traits, which rely on reflective models, capture patterns of correlated behaviors and are not good candidate explanatory mechanisms because they suggest that there should be correspondence, or *common causes*, underlying those traits if they are to explain behavior – that is, the structure should reflect the causal processes that underlie it. In other words, if Agreeableness is to be a candidate cause of politeness, then the cause of politeness should be the same across people. However, this rarely, if ever, holds. Evidence suggests, for example, that some people are motivated to be polite because of a desire for power, some because of a favorable view of others, and some out of social convention (Baumert et al., 2017). In each case, this is a different individual-level causal theory that results in the same structure when measured and observed nomothetically.

That unique idiographic processes produce unique within-person states that themselves underscore a higher-order common nomothetic traits suggests why each level has different utility for description, prediction, and explanation. In prediction, we are typically most interested in building the most simple and parsimonious model. Idiographic processes, which are micro phenomena, are too fine grained and too often joined by relationships that are too complex to be parsimoniously modeled and understood, making them non-optimal for both predicting and describing broader phenomena. Idiographic processes are, however, excellent candidate explanations of within-person states and, through such within-person states, nomothetic traits.

Within-person states, in turn, are much more parsimonious than idiographic processes, making these narrower individual-level dimensions useful for both the prediction and description. Description at this level, like at the idiographic level, is unique to the individual, meaning it is not always possible to make references to other individuals (i.e. no rank ordering on shared dimensions). Moreover, the breadth of the prediction is smaller, however, because the breadth and timing of the phenomena is smaller (e.g. my heartrate right now may predict whether I am exercising but may not predict how much I have exercised this year).

Finally, common nomothetic traits are useful for both broad description and prediction. Descriptively, broad traits can parsimoniously describe how individuals differ from one another on shared dimensions, which is useful when comparing individuals. Predictively, broad traits are strong predictors of broad outcomes as they represent similar levels of aggregation because both represent multiply determined phenomena that have evolved over time. Importantly, implicit in this is that the broad, nomothetic traits, themselves are products of the dynamics of idiographic processes that influence observable within-person states and behaviors whose patterns underlie these broad, nomothetic traits. Thus, the causal true causal link between nomothetic traits and outcomes goes back to the lowest, idiographic level. Despite this, however, the power of nomothetic traits lies in their ability to parsimoniously describe large populations of people and to predict important broad outcomes, like health, longevity, and academic and career success.

**From persons to persons in situations**. Thus far, I have argued that unique idiographic processes give rise to within-person patterns of behaviors and experiences that underlie and explain nomothetic traits. In other words, the focus has been on how person-level properties are related to these outcomes to the neglect of considering the importance of situations and context. However, according to the correspondence principle of personality development, the correspondence between persons and the situations they encounter (or select into) are the foundation of the stability of personality – people select into situations that align with person-level characteristics, and those situations, in turn, reaffirm (or solidify) those characteristics (Roberts et al., 2008). This applies both to smaller, more everyday life events, like choosing a lunch venue, as well as major life events, like having a child. In other words, the process through which people experience life events is predictable. People are neither passive victims of their environment nor fully active controllers of it. Instead, individuals and their environment mutually influence one another through person-environment transactions (Roberts et al., 2008). Person-environment transactions capture the dynamic interplay through which persons and their environment jointly and continuously influence one another. Of course, persons and situations are not always on equal footing – depending on the person and the context, one may have greater influence.

Past research indicates that there are at least three kinds of transactions (Roberts et al., 2008). First, active person-environment transactions occur when individuals seek out contexts that align with their personalities. For example, people higher in Extraversion tend to seek out social stimulation, like attending parties. Second, reactive person-environment transactions occur when individuals in the same situation respond differently. Essentially, when the context is the same but the response is different, the context has drawn out a response from each individual that is aligned with their personalities. For example, while at a party, someone higher in Extraversion is likely to talk and spend time with large groups, while someone lower in Extraversion will likely converse with a smaller group or single person. Finally, evocative person-environment transactions highlight how individuals can draw out or influence environmental contexts in ways that may influence their behavior through reciprocal processes. For example, the Extraverted individual at a party might draw more sociable behaviors out of others, encouraging them to act still more Extraverted.

Thus, persons and environments are constantly influencing one another through combinations of active, reactive and evocative person-environment transactions, at least in the short-term. But such person-environment changes can also influence major life events. For example, through active person-environment transactions, a more Extraverted person may have more opportunities to meet potential romantic partners, which might make them more likely to marry earlier. Similarly, through reactive person-environment transactions, someone lower in Neuroticism may act more calmly and answer questions more effectively in an interview than someone higher in Neuroticism, making it more likely they will start a job. Again, the explanation for these patterns is not the nomothetic personality trait itself but regularities in within-person patterns reliably produces nomothetic traits that differentiate people in ways that are associated with these outcomes. Thus, to the extent that background characteristics, like SES and age, are unrelated or inconsistently related to idiographic processes, there is reason to believe that personality should predict the life events that people experience even when considering other background characteristics. The characteristic signature of a person that is “left” after controlling for those variables should still influence situation selection.

**Moderators of Personality-Outcome Associations**

But understanding the processes through which nomothetic personality may be connected to outcomes does not necessarily elucidate the boundary conditions of when and for whom it does. Rather, these processes elucidate why personality prediction is plausible – or how the processes and patterns could plausibly combine in ways that could influence events. Fully understanding the boundary conditions of personality-outcome associations requires understanding how other demographic, social, economic, and health variables, among others, are related to both personality and outcomes. Moreover, despite evidence that personality-outcome associations are reproducible when following the same procedures (Soto, 2019), there is evidence of cross-sample heterogeneity in personality-outcome associations in large scale panel studies (see Jokela et al., 2013).

There are a number of pathways through which many factors beyond personality may influence personality-outcome associations. First, as noted previously, such measures may be confounded with personality itself through reverse causality. Early childhood experiences, for example, may impact the pathways that individuals follow in their lives, which, in turn, influences personality development. For example, early childhood interventions, like the Perry Preschool Project, have been shown to predict early boosts in intelligence (IQ) and behavior. However, most intelligence differences between those in the program and controls level out by about age 11, while participants show lower rates of externalizing behaviors and criminal activity (Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Schweinhart, Montie, Xiang, Barnett, Belfield, & Nores, 2005). Such results suggest that while early childhood interventions may not have lasting effects on intelligence, they do impact other aspects of personality in ways that can have strong consequences for other important outcomes.

Second, personality may have a differential relationship with and impact on both outcomes and covariates throughout the lifespan. For example, there is well-documented evidence that personality differentially predicts health behaviors throughout the lifespan, with Conscientiousness predicting different health behaviors to different degrees for individuals above and below age 30 (Bogg & Roberts, 2004). Moreover, at different ages, certain health behaviors may be differentially important for health, with physical activity playing a large, cumulative effect across the whole lifespan, nutrition playing its largest role in early childhood, and medication adherence being an important health predictor primarily in older adulthood. Such age differential patterns of personality-outcome associations underlie the importance of investigating moderators of such associations (Hill, Edmonds, & Jackson, 2019). In the case of medication adherence, for example, a full lifespan sample may show little association between personality and health indicators influenced by medication adherence because emerging (18-25) and young adults (26-40) show little association between personality and such outcomes, while middle (40-60) and older adults (60+) show stronger associations. Without explicitly accounting for different personality-outcome associations for different ages, these may appear null. In contrast, a sample with only emerging and young adults would show no association while a middle and older adult population may show strong effects. Critically, such heterogeneity in the effect may lead some to believe it is neither robust nor reliable, when in fact it is, just as a function of age. Thus, understanding moderators of personality is imperative to understanding personality-outcome associations.

Third, concerns about replication within psychology highlight the importance of understanding how cross-study variance can influence the direction and magnitude of personality-outcome associations. As a recent example, Watts, Duncan, and Quan (2018) replicated the results of Shoda, Mischel, and Peake’s (1990) test of prospective delay of gratification-behavior / achievement outcome associations. In their sample, they found that zero-order correlations were approximately half the magnitude of the original sample and those effect sizes were reduced by a factor of approximately two-thirds when controlling for background characteristics. They reasoned that one possible reason for such attenuated results could be task differences because the sample used in Shoda and colleague’s (1990) initial study, which included children recruited from the Stanford University community, who likely had higher parental income and education than many children in the United States and elsewhere. The higher overall SES raised the possibility that SES, not delay of gratification, could drive the link between delay and behavioral and achievement outcomes given well-established links between SES and such outcomes. However, Watts and colleagues (2018) found that SES did not moderate the association between delay of gratification and composites of behavioral or achievement outcomes. In other words, it appears that delay of gratification and a number of background characteristics predict important achievement and behavioral outcomes, but that these effects appear statistically independent.

Even though SES does not appear to moderate prospective delay of gratification-outcome associations, however, Watts and colleagues’ (2018) study highlights the importance of both replication and examining variation across different samples, studies, and designs. As with Shoda and colleagues (1990) original sample, which may have had a restricted range of SES, many studies of personality-outcome associations, even those using large-scale panel studies, may have properties that do not allow them to (1) test potential within- and cross-sample moderators of personality-outcome associations and (2) generalize across different populations. Instead, testing such questions requires pooling or meta-analytic techniques to systematically test whether effects are similar across studies. However, because such studies are time and resource intensive, they remain relatively rare. But a recent pooled analysis of six longitudinal studies that investigated personality-mortality links suggests why such studies are important (Jokela et al., 2013). Across the studies, Agreeableness, for example, showed no meta-analytic effect overall. However, one study showed a positive Agreeableness-mortality association, one a negative association, and the others no association, highlighting very real heterogeneity – even in studies with large samples – as well as the importance of looking at cross-study variance in effects.

In sum, within- and cross-study moderators can greatly threaten the reproducibility of personality-outcome associations and also mask both the boundary conditions of such associations and the processes underlying them. But large, systematic investigations of personality-outcome associations using multiple samples offer insight into previously untested questions concerning moderators and boundary conditions of personality prediction.

**Researcher Degrees of Freedom**

Although a number of studies, reviews, and even replication studies suggest that personality-outcome associations are relatively robust, these reflect the published literature, which may overestimate the number of significant personality-outcome associations. Publication bias occurs because of an emphasis on positive, surprising, and novel results, sometimes at the expensive of more solid methods and measures.

Thus, results that are novel or positive make up the bulk of the published literature, which has led to an increase in researcher degrees of freedom, or the many choices that must be made in any study that may not be adequately documented. Although researcher degrees of freedom can occur at any point in a study, from designing questions to writing up the results, in the present paper, I will focus on two forms of researcher degrees of freedom: operationalizing variables (including data cleaning procedures) and choosing covariates.

Psychology, and the study of personality, in particular, have long invested great effort into psychometrics, or creating measures of psychological phenomena that have good measurement properties, such as internal consistency and test-retest consistency. One outcome of this emphasis is that there are many freely available measures of psychological constructs, especially individual difference measures, like the Big Five, self-esteem, depression, and other constructs. However, many studies of personality-outcome associations use scales that have not been validated, subsets of indicators within scales that do not have the same psychometric properties as the full scale, or other data cleaning or analytical choices, such as binarizing measures into “high-low” subsets (Flake & Fried, 2019). Importantly, however, different operationalizations of personality characteristics can have strong consequences for estimates of personality-outcome associations, either exaggerating or attenuating the estimates, and invalidating conclusions about the robustness of these associations.

Even if constructs are operationalized and data are cleaned similarly, researcher degrees of freedom can creep in through model building and the selection of covariates. Due to publication bias, researcher degrees of freedom creep in when researchers choose model specification without clear theoretical forethought or basis or based on significance, which may result in different studies using different combinations of covariates, some theoretically dubiously linked to either key predictors or outcomes. Importantly, coupled publication bias, this results in a bias toward significant results rather than robust, valid, and replicable results.

To demonstrate the great heterogeneity in covariate inclusion, consider the subset of papers reviewed by Roberts and colleagues (Roberts et al., 2007) on the relationship between personality characteristics (e.g. hostility, optimism, the Big Five) and mortality (see Table 2 in Roberts et al., 2007). The papers reviewed were heterogeneous in many ways, but one of the largest and most obvious is in which covariates (or controls) were included in each study, including age, biological sex, race, blood pressure, smoking, marital status, cohabitation status, physician ratings of health, depression, IQ, education, income, SES, and expected survival, to name just a few. In total, the reported studies included a total of more than 50 unique covariates, with few studies using the same covariates. Although most of these covariates have some plausible link to mortality, some inclusion standards, such as including varying combinations of maternal and paternal background characteristics, seem arbitrary, indicating likely researcher degrees of freedom. More troublesome, however, is that despite the published literature indicating some 50 covariates of interest, there has been no systematic study of to date how such covariates influence personality-outcome associations.

Recently, Simonsohn and colleagues proposed specification curve analysis as a solution to the effect of publication bias on estimating the robustness of statistical tests (Simonsohn et al., 2015). Specification curve analysis proceeds in three parts: identifying plausible specifications, testing all combinations of plausible specifications, and using an inferential permutation-based test to test whether the phenomena is robust under the null of no effect. Each of these stages will be discussed in more detail below. Specification curve analysis has been successfully used, for example, in testing the robustness of birth order effects (Rohrer, Egloff, & Schmukle, 2017) and how black-sounding names predict hiring decisions (Simonsohn et al., 2015). But despite its ability to test both the robustness of personality-outcome associations and to identify covariates that have the largest impact on such associations, to date, specification curve has been applied in almost no cases within personality psychology.

**The Present Study**

Together, the aforementioned evidence suggests that previous investigations of personality-outcome associations have not adequately accounted for reverse causality and publication bias, leaving the robustness and boundary conditions of personality-outcome associations unknown. Moreover, previous investigations have not systematically addressed which covariates are fundamentally influence personality-outcome associations, how study-level characteristics may influence them, or how researcher degrees of freedom influence them. The present study addresses selection bias and researcher degrees of freedom influence the boundaries of estimates of personality prediction using propensity score matching and specification curve analysis in 10 longitudinal studies using 14 personality characteristics and 14 outcomes.

**Method**

**Participants**

**Ad Health**. The National Study of Adolescent to Adult Health (Ad Health; Harris & Udry, 2018) is an ongoing longitudinal study of adolescents in the United States that began as a response to a federal mandate to better understand adolescent health. The data are available online at <https://www.icpsr.umich.edu/icpsrweb/DSDR/studies/21600>.

The initial sample of participants included approximately 20,000 students who completed at home administrations the study. Four waves of data collection (1994-1995, 1996, 2001-2002, and 2008) have been completed. The latest release contains data through 2008. Another wave of collection began in 2016 but has not yet been released. More documentation of the data are available at <https://www.icpsr.umich.edu/icpsrweb/content/DSDR/add-health-data-guide.html#intro>.

Sample sizes vary by year, from 14,738 (1996) to 20,745 (1994-1995). This provides 99% power to detect a correlation effect size of ~.03.

**BHPS**. The British Household Panel Study (BHPS; University of Essex, 2018) is a longitudinal study of households in the United Kingdom. These data are available online, through application, from https://www.iser.essex.ac.uk/bhps/about/latest-release-of-bhps-data.

Participants were recruited from more than 15,000 individuals from approximately 8,000 households in the United Kingdom. Data have been collected annually since 1991 from approximately 10,000 individuals (5,500 households) in Great Britain but expanded to include Scotland and Wales in 1999 and Northern Ireland in 2001. In 2010, the BHPS stopped data collection, but 6,700 of the current 8,000 participants were solicited to become part of the broader Understanding Society study (University of Essex, 2019). Participants can be matched across studies, so I will use additional data on the original BHPS participants from the Understanding Society study for additional waves of outcome data.

Sample sizes vary by year, ranging from 10,264 (1991) to 14419 (2008). This provides 99% power to detect a zero-order correlation effect size of ~.05, two-tailed at alpha .05.

**GSOEP**. The German Socioeconomic Panel Study (GSOEP; Socio-Economic Panel, 2017) is an ongoing longitudinal study of German collected by the German Institute of Economic Research (DIW Berlin). The data are freely available at https://www.diw.de/soep by application.

Data have been collected annually since 1984 (the latest data release includes data up to 2017). Participants have been recruited from more than 11,000 households, which are nationally representative of private German households. 20,000 individuals are sampled each year, on average. It is critical to note that the GSOEP samples households, not individuals, and the households consist of individuals living in both the “old” and “new” federal states (the former West and East Germany), foreigners, and recent immigrants to Germany.

Sample size varies by year, ranging from approximately 10,000 (1989) to 31,000 (2013). This provides 99% power to detect a zero-order correlation effect size of ~.06, two-tailed at alpha < .05.

**HILDA**. The Household Income and Labour Dynamics in Australia (HILDA; Wilkins, Laß, Butterworth, & Vera-Toscano, 2019) study is an ongoing longitudinal study of Australian households. These data are available through application from https://melbourneinstitute.unimelb.edu.au/hilda/for-data-users.

Participants were recruited from more than 17,000 individuals. Data have been collected annually since 2001. The latest data release includes 17 waves of data from 2001 to 2017. More documentation can be found in the HILDA data dictionary at https://www.online.fbe.unimelb.edu.au/HILDAodd/srchSubjectAreas.aspx.

Sample sizes vary by year, ranging from 12,408 (2004) to 17,693 (2016). This provides 99% power to detect a zero-order correlation effect size of ~.03, two tailed at alpha .05.

**HRS**. This study uses the Health and Retirement Study (HRS; Juster & Suzman, 1995) data. The HRS is an ongoing longitudinal study of households in the United States. These data are available at https://hrs.isr.umich.edu by creating a free account.

Participants were recruited from more than 35,000 individuals from the financial households of individuals born between 1931 and 1941 in the US. Data have been collected biannually since 1992. The latest data release includes data up to 2016. On average, 10,000 individuals are sampled each wave More information on the HRS can be found at https://hrs.isr.umich.edu/documentation/survey-design, but, in short, the HRS is a nationally representative sample of adults over 50 in the US. It is critical to note that the HRS samples households of the original cohort and follows individuals and their spouses or partners until their death.

Sample size varies by year, ranging from approximately 7,500 (2014) to 15,500 (1992). (https://hrs.isr.umich.edu/sites/default/files/biblio/ResponseRates\_2017.pdf). This provides 99% power to detect a zero-order correlation effect size of ~.04, two-tailed at alpha .05.

**LISS**. The Longitudinal Studies for the Social sciences (LISS; Scherpenzeel, Das, Ester, & Kaczmirek, 2010) is an ongoing longitudinal study of households in the Netherlands. These data are online, through application, from https://statements.centerdata.nl/liss-panel-data-statement.

Participants were approximately 8,000 Dutch-speaking individuals permanently residing in the Netherlands from 5,000 households. Data have been collected annually since 2007. The latest data release includes 11 waves of data from 2008 to 2018. More documentation are available at https://www.dataarchive.lissdata.nl/study\_units/view/1.

Sample sizes vary by year, ranging from 5,021 (2018) to 6808 (2008). This provides 99% power to detect a correlation effect size of ~.04, two-tailed at alpha .05.

**MIDUS**. The Midlife in the United States (MIDUS; Brim, Ryff, & Kessler, 2004; Ryff et al., 2012, 2016) study is an ongoing longitudinal study of adults in the United States. These data are available at http://www.icpsr.umich.edu by making a free account.

Participants included more than 10,000 individuals aged 25 or older from the United States. The present study uses data from MIDUS I, II, and III. MIDUS I was collected in 1995-1996. MIDUS II was the follow-up to MIDUS I and was collected from 2004-2006. MIDUS III was an additional follow-up conducted from 2013-2014. More information can be found at http://midus.wisc.edu/findings/Understanding\_Data\_Collection\_in\_MIDUS.pdf.

Sample size varies by wave, with 7,108 (MIDUS I), 4,963 (MIDUS II), 3,294 (MIDUS III). This provides 99% power to detect a zero-order correlation effect size of ~.06, two-tailed at alpha .05.

**NLSY**. The Children to Young Adults Study (CNLSY; Bureau of Labor Statistics, 2017) is an offshoot study of the National Longitudinal Study of Youth (NLSY79), which is an ongoing longitudinal, nationally representative study of more than 12,500 individuals in the United States that began in 1979. The CNLSY includes the biological children of the NLSY79 participants and began in 1986. Children (10 years and older) completed separate inventories from children (or “young adults”) aged 15 and above. Mothers of children 10 and below also completed surveys on the children prior to age 10. All participants were interviewed in addition to surveys.

Sample sizes vary by year, ranging from approximately 1,331 (1979) to 11,530 (2016). This provides 99% power to detect a zero-order correlation effect size of ~.05.

The present study included data from a subsample of 7,736 CNLSY participants who had at least one wave of measurement for *each* of the following: matching variables (see Supplementary Materials; partial data allowed), life event / outcome variables (e.g. marriage), and personality variables (e.g. the Big Five).

**SHP**. The Swiss Household Panel Study (SHP; Voorpostel et al., 2016) “Living in Switzerland” is an ongoing longitudinal study of households in Switzerland. These data are available online, through application from https://forsbase.unil.ch/project/study-public-overview/15632/0/.

Participants were recruited from more than 10,000 individuals from the households whose members represent the non-institutional resident population of Switzerland. Data have been collected annually since 1999. The latest data release includes data up to 2018. On average, about 5,000 individuals are sampled at each wave. More documentation can be found at LINK, but, in short, the SHP is a nationally representative sample of Swiss citizens.

Sample sizes vary by year, ranging from 5,220 (2003) to 13,295 (2013). This provides 99% power to detect a zero-order correlation effect size of ~.06, two tailed at alpha .05.

**WLS**. The Wisconsin Longitudinal Study (WLS) is an ongoing longitudinal study of individuals who graduated from Wisconsin high schools in 1957 and were born between 1937 and 1940 as well as their siblings.

Graduates were randomly recruited from Wisconsin high schools in 1957 and born between 1937 and 1940. In 1977, at least one sibling of the original graduates from 2,100 families were also invited to participate in the study. As such, the study is representative of older, white Americans who have at least a high school education. Graduate data have been collected in in 1957, 1964, 1975, 1992, 2004, and 2011, and sibling data have been collected in 1977, 1994, 2005, and 2011. Personality data were initially collected in 1992 for graduates and 1994 for siblings. More documentation can be found at https://www.ssc.wisc.edu/wlsresearch/.

Sample sizes vary by wave, from 9,681 (2011) to 10,317 (1957). This provides 99% power to detect zero-order correlation effect sizes of ~.06, two-tailed at alpha .05.

**Measures**

In the proposed study, I will test how 14 personality characteristics are associated with 14 life events and outcomes, while controlling for more than 50 background (matching) characteristics (propensity score matching) or covariates and operationalizations (specification curve analysis), testing seven potential moderators of personality predicting such outcomes. For a full overview of which personality, outcome, and moderator measures are available across data sets, see Table 2.

**Personality**. I will examine 14 personality characteristics: the Big Five (Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience), self-esteem, optimism / pessimism, subjective well-being, locus of control, social support, life satisfaction, negative affect, positive affect, depression, and intelligence. Full information on the scales used for each of these measures for each study is presented Table 3. Because none these data sets have been accessed in the design of this study, descriptive statistics for each scale are not available at this stage. However, as is clear in Table 3, many of the measures were on different scales, so all personality indicators will be operationalized as *P*ercentages *O*f the *M*aximum *P*ossible score (POMP) in the mega-analytic procedure (Cohen, Cohen, Aiken, & West, 1999). Unlike standardization procedures, that have a mean of zero and unit variance and can be misleading when data are skewed, POMP does not rescale sample variance based on the observed data, which overly relies on deviations from the mean. Instead, POMP relies on the ratio between the difference between a score and the minimum and the maximum and minimum, or

.

**Life events and outcomes**. I will investigate whether personality predicts 14 life events and outcomes. A full list of outcomes can be seen in Table 2. The outcomes chosen are those frequently tested in other studies and most likely to be included in large panel studies. For each outcome, participants reported whether the outcome had occurred in the survey year or years prior or, in some cases, the year an outcome or event first occurred. Responses will be coded as "1" for that event if participants reported experiencing it anytime between the year after the utilized personality measure to the latest available wave and "0" otherwise. Participants who experienced events prior to the first personality measure will be excluded.

**Moderators**. In addition to matching for common demographic variables, like age and gender, I will additionally test whether these, as well as survey year, ethnicity, socioeconomic status (SES), personality measure reliability, and study moderate the relationship between personality and outcomes. Age and SES will be coded at the time of measured personality. Age, and SES will be test as continuous Level 1 moderators, while ethnicity and gender will be tested as binary Level 1 moderators. Reliability and survey year will be tested as continuous Level 2 moderators. The effect of study will be tested by examining the Level 2 Variance of personality-outcome associations. Significant variance will indicate that study moderates the effect.

**Matching variables and specification curve covariates**. Matching variables are those to be used in the propensity score analysis to match those who did or did not experience different life events or outcomes. Target variables can be roughly broken down into eight categories: demographics (e.g. sex), activities (e.g. volunteering), financial (e.g. gross wages), household (e.g. number of household members), health (e.g. BMI), psychological (e.g. life satisfaction), relationship (e.g. relationship with father), and social (e.g. visits to friends). In order to construct more reliable measures and not exclude participants who did not respond to surveys in the same year as the personality measures, matching variables were pooled across time, using all available data from the earliest wave of the study to the year of the utilized personality measure for each person. Cross-study covariates will be operationalized using POMP, with the exception of core variables with meaningful scales (e.g. age, gender, etc.). The full details of the construction of these variables will be available in supplemental files.

**Analytic Plan**

**Study 1: Propensity Score Matched mega-analysis of longitudinal studies.** Confirmatory analyses will be tested using a series of multilevel Bayesian logistic regression models implemented using the brms (Bürkner, 2017, 2018) package in R (R core team, 2018). I will use the default priors for all fixed effects, which are "uninformative" priors meant to regularize the models. However, given the sample sizes in the present model, the data are likely to overwhelm these priors. I will use half Cauchy priors for Level 2 variances and LKJ Cholesky covariance priors for Level 2 covariances.

The analysis phase will consist of three main parts, with interim steps to link these together: multiple imputation, propensity score matching, and tests of selection effects using multilevel Bayesian logistic regression models.

First, I will use multiple imputation to impute missing data for the matching variables separately for each study. Before doing so, I will first create composites of the matching variables prior to the year for which we will use personality data for each study. I elected to use composites rather than survey responses from the year prior to personality measures due to irregularities in survey construction and responses that would severely restrict the number of observations. To ensure transparency, I will conduct all analyses using the raw data imported directly from the data files obtained from data maintainers for each study, and all steps in creating the composites are documented in an extensive codebook containing the item lists, text, scales, and recoding of all variables for all studies. Moreover, all steps will be documented in files and code shared on the Open Science Framework and GitHub.

The composite matching variables will then be used in multiple imputation and propensity score matching, which requires completely non-missing data. Multiple imputation will be conducted using the amelia package in R (Honaker, King, & Blackwell, 2011). I will impute10 data sets.

Second, I will use the multiply imputed data to calculate propensity scores for each of the multiply imputed data sets for each outcome and study separately. In addition, I will conduct the procedure separately for each individual difference characteristic, such that each propensity score matched set will be matched on all matching variables and all personality variables except for the target variable that will be used to predict each outcome, making this an extremely conservative test of the relationship between personality and outcomes. Finally, separate propensity score matched sets will also be generated to test each of the Level 1 moderating questions (age, gender, SES, and ethnicity).

The propensity score matching procedure attempts to equate those who did or did not experience an outcome by assigning each person a risk score based on a number of background factors. Then each person who experienced the outcome is matched with someone else in the control group who had a similar "risk" of experiencing the outcome. Matching will be done using the matchit packages in R (Ho, Imai, King, & Stuart, 2011). Because the sample sizes of the groups of people who experience specific outcomes are much smaller than the individuals who did not experience them, we choose to use propensity score matching rather than propensity score weighting. I will begin by using "nearest neighbor" matching and a ratio of 2 to 1 and a caliper width of .25 (Guo & Fraser, 2015) and iteratively increase the ratio for outcomes that were not balanced using these criteria.

Third, I will test for selection effects using a series of multilevel Bayesian logistic regression models using the brms package in R (Bürkner, 2017, 2018). In all models, the "no outcome" group will be considered the reference group. Using unmatched data sets and matched data sets that did not account for personality or moderators at the wave of personality assessment, I will predict outcomes from baseline personality. With the exception of moderator analyses, I will not include additional covariates (e.g. age, gender) in these models because these should be effectively controlled for in the propensity score matching procedure. The basic form of the model is as follows:

Level 1:

Level 2:

,

where is the average log odds of experiencing the outcome across all studies and multiple of log odds change associated with a one-unit change in the percentage of the maximum of the possible (POMP) personality score. All results will be presented both as log odds and as odds ratios (OR) with 89% uncertainty intervals (UI). indicates the difference between the average estimate of log odds of experiencing an outcome and the estimate for each study (i.e. the study-specific estimate of the log odds of each outcome), and indicates the difference between the average multiple of log odds associated with a one unit change in POMP personality score and the estimate for each study (i.e. the study-specific estimate of the personality-outcome relationship). Each of these will be presented as forest plots showing both study-specific and average effects.

Moderator analyses will extend the form of the core analyses by additional terms at Level 1 and ; age, ethnicity, gender) or Level 2 and ; reliability, survey year). For Level 1 moderators, I will include random effects that will capture the study-specific effects for each moderator term.

**Study 2: Specification curve analysis of longitudinal studies**. As a second test of the robustness of prospective personality-outcome associations, I will conduct a specification curve analysis for each personality-outcome combination (Simonsohn et al., 2015).

Specification curve analysis is carried out in three main steps. First, the researcher defines the set of reasonable model and variable specifications. Second, the researcher estimates all of these reasonable specifications and represents them using a specification curve. Finally, the researcher constructs an inferential specification curve using join statistical analyses (Simonsohn et al., 2015). Constructing the specification curve in the second step serves both to show the full range of how specifications influence the statistical results as well as which specifications are most consequential for the results, while using the specification curve inferentially in the third step allows the researcher to make a statistical inference about whether the curve is inconsistent with a null hypothesis of no effect of personality on outcomes. This is most simply done with a permutation test in which the consequential variable (in the context of this paper, whether an outcome occurred for a participant) is shuffled. Shuffling the consequential outcome and re-estimating the full set of possible specifications, except now when there is no reason to expect a relationship between predictors or covariates and outcomes, produces a distribution of specification curves under the null of no personality-outcome relationship.

***Defining specifications***. The first step in specification curve analysis involves defining the set of valid specifications. Because I am testing the relationship between different personality characteristics and outcomes, the set of valid specifications will differ across outcomes. Thus, a unique set of specifications was established for each outcome separately based on (1) covariates used in previous studies to predict those outcomes, (2) other covariates that I identified as plausibly related to each outcome based on theory, (3) using the other personality characteristics that are not the focal part of the target personality-outcome association, and (4) different operationalizations of covariates used in previous studies. Different operationalizations include whether variables are treated as continuous or binarized or trichotomized, as was sometimes done in previous studies. In addition, the chosen specifications include whether the covariates are cross-sectional (i.e. measured in the same year as personality) or composites of multiple waves of measures of the covariate (i.e. measured prior to personality). Because the focus is on covariates that may predict the outcome, the same set of covariates will be used to test all personality-outcome associations for each outcome. Table 4 presents the full set of covariates and operationalizations that will be used for each outcome, as well as the total number of specifications that result from the set of specifications.

***Defining the specification curve***. The next step in specification curve analysis is to run the model, specifying all combinations of the specifications from the previous step. Once this is done, the target personality-outcome association is extracted from each model and ordered from strongest negative to strongest positive to define the specification curve. As can be seen in Figure 1 in the sample specification curve from Rohrer et al. (2017), a visual representation of which specifications were used in each model is included below the curve. Thus, I will define a total of 196 specification curves (14 personality characteristics x 14 outcomes). Although I could compute separate models for each study, the number of specifications would be so large as to be intractable, so I will restrict the model to multilevel logistic regression. Because conducting the models in a Bayesian framework would require resources likely well beyond the scope of this project, I will fit in a frequentist multilevel model framework using lme4.

A close up of text on a white background

Description automatically generated

*Figure 1*. Sample specification curve of birth order predicting positive reciprocity from Rohrer et al. (2017). The top panel includes the specification curve with points ordered by effect size, while the bottom panel demonstrates which covariates and operationalization were included in each specification. Longer, red lines indicate significant birth order-positive reciprocity effects, while shorter, black lines indicate non-significant effects.

The basic form of the model is the same as for the propensity score matched mega-analysis, but with additional covariates to ):

Level 1:

Level 2:

…

,

where to represent Level 1 main effects with no random slope (i.e. study level slope).

**Permutation-Based Inferential Test.** The final step in specification curve analysis involves conducting a permutation-based test to determine whether the observed specification curve differs from the null of no relationship between any predictors and the outcome. Thus, the outcome variable is shuffled and the specification curve procedure from the second step is repeated a large number of times. The observed specification curve can then be plotted against the median permuted curves and the 95% interval of the permuted curves to demonstrate how the observed curve differs from the null.

Because none of the specifications are independent because all use some overlapping variables, traditional tests that assume independence are not appropriate. Instead, I will base decisions on whether curves differ from the null on three tests: (1) the median overall point estimate within each specification curve, (2) the percentage of specifications that are of the dominant sign, and (3) the percentage of specifications with the dominant sign that are also significant. Each of these results in *p* value constructed from the number of permutations that meet each criterion divided by the number of permutations.

References

Adler, N. E., Boyce, W. T., Chesney, M. A., Folkman, S., & Syme, S. L. (1993). Socioeconomic inequalities in health: no easy solution. *JAMA*, *269*(24), 3140-3145.

Allport, G. W. (1937). Personality: A psychological interpretation.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9* (1-2), 9–30.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.[doi:10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., ... & Jayawickreme, E. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality*, *31*(5), 503-528.

Beck, E. D. (2019) Why do states make traits? A three-level model of personality Unpublished manuscript.

Beck, E. D. and Jackson, J. J. (2019a). Idiographic personality: A return to Allportian approaches to personality. *Current Directions in Psychological Science*.

Beck, E. D. and Jackson, J. J. (2019b). Network approaches to representing and understand- ing psychological dynamics. In Wood, D., Read, S., Harms, P., and Slaughter, A., editors, *Measuring and* *Modeling the Person and Situation*. Elsevier, 1st edition.

Beck, E. D. and Jackson, J. J. (2019c). Consistency and change in idiographic personality net- works: A longitudinal ESM study*. Journal of Personality and Social Psychology*.

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: a meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin*, *130*(6), 887.

Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time?. *Journal of Research in Personality*, *32*(2), 202-221.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203.

Brim, O. G., Ryff, C. D., & Kessler, R. C. (2004). The MIDUS national survey: An overview. *How healthy are we*, 1-36.

Bureau of Labor Statistics, U.S. Department of Labor, and National Institute for Child Health and Human Development. Children of the NLSY79, 1979-2014. Produced and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH: 2017.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1-28.

Bürkner, P. C. (2018). *Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal, 10 (1), 395–411. doi: 10.32614*. RJ-2018-017.

Cattell, R. B. (1946). Personality structure and measurement. I. The operational determination of trait unities. *British Journal of Psychology*, *36*, 88–102.

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*(3), 315-346.

Conger, R. D., & Donnellan, M. B. (2007). An interactionist perspective on the socioeconomic context of human development. *Annu. Rev. Psychol.*, *58*, 175-199.

Cramer, A. O., Van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., ... & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*(4), 414-431.

Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the big five across the life span: Evidence from two national samples. *Psychology and Aging*, *23* (3), 558.

Ensminger, M. E., Fothergill, K. E., Bornstein, M. H., & Bradley, R. H. (2003). A decade of measuring SES: What it tells us and where to go from here. *Socioeconomic status, parenting, and child development*, *13*, 27.

Flake, J. K., & Fried, E. I. (2019, January 17). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. https://doi.org/10.31234/osf.io/hs7wm..

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502-1505.

Guo, S., & Fraser, M. W. (2015). *Propensity score analysis*. Sage.

Hampson, S. E. (2012). Personality processes: Mechanisms by which personality traits ?get outside the skin? *Annual Review of Psychology*, *63*, 315–339.

Hampson, S. E., Edmonds, G. W., Goldberg, L. R., Dubanoski, J. P., & Hillier, T. A. (2015). A life-span behavioral mechanism relating childhood conscientiousness to adult clinical health. *Health Psychology*, *34*(9), 887.

Harris, K. M., & Udry, J. R. (2018). National longitudinal study of adolescent to adult health (add health), 1994-2008 [Public Use]. *Ann Arbor, MI: Carolina Population Center, University of North Carolina-Chapel Hill [distributor], Inter-university Consortium for Political and Social Research [distributor]*, 08-06.

Hauser, R. M., & Sewell, W. H. (2012). Wisconsin Longitudinal Study (WLS)[graduates, siblings, and spouses]: 1957–2019, Version 13.07.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of public Economics*, *94*(1-2), 114-128.

Hill, P. L., Edmonds, G. W., & Jackson, J. J. (2019). Pathways Linking Childhood Personality to Later Life Outcomes. *Child Development Perspectives*, *13*(2), 116-120.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, http://gking. harvard. edu/matchit*.

Honaker J, King G, Blackwell M (2011). “Amelia II: A Program for Missing Data.” Journal of Statistical Software, **45**(7), 1–47.<http://www.jstatsoft.org/v45/i07/>.

Jackson, J. J., Connolly, J. J., Garrison, S. M., Leveille, M. M., & Connolly, S. L. (2015). Your friends know how long you will live: A 75-year study of peer-rated personality traits. *Psychological Science*, *26*(3), 335–340.

Jackson, J. J., Thoemmes, F., Jonkmann, K., Lüdtke, O., & Trautwein, U. (2012). Military training and personality trait development: Does the military make the man, or does the man make the military? *Psychological Science*, *23* (3), 270–277.

Jokela, M., Batty, G. D., Nyberg, S. T., Virtanen, M., Nabi, H., Singh-Manoux, A., & Kivimäki, M. (2013). Personality and all-cause mortality: Individual-participant meta-analysis of 3,947 deaths in 76,150 adults. *American Journal of Epidemiology*, *178*(5), 667–675.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, *52*(3), 621-652.

Juster, F. T., & Suzman, R. (1995). An overview of the Health and Retirement Study. Journal of Human Resources, 30, S7–S56.

Kelly, E. L., & Conley, J. J. (1987). Personality and compatibility: A prospective analysis of marital stability and marital satisfaction. *Journal of Personality and Social Psychology*, *52*(1), 27.

Larson, L. M., Rottinghaus, P. J., & Borgen, F. H. (2002). Meta-analyses of big six interests and big five personality factors. *Journal of Vocational Behavior*, *61*(2), 217–239.

Lüdtke, O., Roberts, B. W., Trautwein, U., & Nagy, G. (2011). A random walk down university avenue: Life paths, life events, and personality trait change at the transition to university life. *Journal of Personality and Social Psychology*, *101* (3), 620.

Malouff, J. M., Thorsteinsson, E. B., Schutte, N. S., Bhullar, N., & Rooke, S. E. (2010). The five-factor model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality*, *44* (1), 124–127.

Martin, L. R., Friedman, H. S., & Schwartz, J. E. (2007). Personality and mortality risk across the life span: The importance of conscientiousness as a biopsychosocial attribute. *Health Psychology*, *26*(4), 428.

Miller, T. Q., Smith, T. W., Turner, C. W., Guijarro, M. L., & Hallet, A. J. (1996). Meta-analytic review of research on hostility and physical health. *Psychological Bulletin*, *119*(2), 322.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*(2), 246.

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201-218.

Moutafi, J., Furnham, A., & Paltiel, L. (2005). Can personality factors predict intelligence? *Personality and Individual Differences*, *38*(5), 1021–1033.

Mueller, S., Wagner, J., & Gerstorf, D. (2017). On the role of personality in late life. In *Personality development across the lifespan* (pp. 69–84). Elsevier.

Mund, M., & Neyer, F. J. (2014). Treating personality-relationship transactions with respect: Narrow facets, advanced models, and extended time frames. *Journal of Personality and Social Psychology*, *107*(2), 352.

Neyer, F. J., & Lehnart, J. (2007). Relationships matter in personality development: Evidence from an 8-year longitudinal study across young adulthood. *Journal of Personality*, *75*(3), 535–568.

Neyer, F. J., Mund, M., Zimmermann, J., & Wrzus, C. (2014). Personality-relationship transactions revisited. *Journal of Personality*, *82* (6), 539–550.

Nieß, C., & Zacher, H. (2015). Openness to experience as a predictor and outcome of upward job changes into managerial and professional positions. *PloS One*, *10*(6), e0131115.

Nores, M., Belfield, C. R., Barnett, W. S., & Schweinhart, L. (2005). Updating the economic impacts of the High/Scope Perry Preschool program. *Educational Evaluation and Policy Analysis*, *27*(3), 245-261.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.*, *57*, 401-421.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rentfrow, P. J., Jokela, M., & Lamb, M. E. (2015). Regional personality differences in great britain. *PLoS One*, *10*(3), e0122245.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345.

Roberts, B. W., Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. *Handbook of Personality: Theory and Research*, *3*, 375–398.

Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2017). Probing birth-order effects on narrow traits using specification-curve analysis. *Psychological Science*, *28*(12), 1821-1832.

Rothmann, S., & Coetzer, E. P. (2003). The big five personality dimensions and job performance. *SA Journal of Industrial Psychology*, *29*(1), 68-74.

Ryff, C., Almeida, D. M., Ayanian, J., Carr, D. S., Cleary, P. D., Coe, C., & Williams, D. (2012). Midlife in the United States (MIDUS 2), 2004-2006. *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]*, 11-20.

Ryff, C., Almeida, D., Ayanian, J., Binkley, N., Carr, D. S., Coe, C., & Williams, D. (2016). National survey of midlife development in the United States (MIDUS 3), 2013–2014. ICPSR36346-v4. *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi. org/10.3886/ICPS R*, *36346*, v4.

Scherpenzeel, A., Das, J. W. M., Ester, P., & Kaczmirek, L. (2010). 'True' longitudinal and probability-based internet panels: Evidence from the Netherlands. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet: Advances in applied methods and research strategies*. New York: Routledge. 77-103.

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental psychology*, *26*(6), 978.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification curve: Descriptive and inferential statistics on all reasonable specifications. *Available at SSRN 2694998*.

Socio-Economic Panel (SOEP), data for years 1984-2017, version 34, SOEP, 2019, doi: 10.5684/soep.v34.

Solomon, B. C., & Jackson, J. J. (2014). Why do personality traits predict divorce? Multiple pathways through satisfaction. *Journal of Personality and Social Psychology*, *106*(6), 978.

Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science*, *30*(5), 711-727.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, *100* (2), 330.

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the big five. *Journal of Personality and Social Psychology*, *101* (4), 862.

Sutin, A. R., Luchetti, M., Stephan, Y., Robins, R. W., & Terracciano, A. (2017). Parental educational attainment and adult offspring personality: An intergenerational life span approach to the origin of adult personality traits. *Journal of Personality and Social Psychology*, *113*(1), 144.

Turiano, N. A., Chapman, B. P., Gruenewald, T. L., & Mroczek, D. K. (2015). Personality and the leading behavioral contributors of mortality. *Health Psychology*, *34*(1), 51.

Turiano, N. A., Hill, P. L., Roberts, B. W., Spiro III, A., & Mroczek, D. K. (2012). Smoking mediates the effect of conscientiousness on mortality: The veterans affairs normative aging study. *Journal of Research in Personality*, *46* (6), 719–724.

University of Essex. Institute for Social and Economic Research. (2018). *British Household Panel Survey:* Waves 1-18, 1991-2009. [data collection]. *8th Edition*. UK Data Service. SN:5151, <http://doi.org/10.5255/UKDA-SN-5151-2>

University of Essex, Institute for Social and Economic Research. (2019). Understanding Society: Innovation Panel, Waves 1-11, 2008-2018. [data collection]. 9th Edition. UK Data Service. SN: 6849, http://doi.org/10.5255/UKDA-SN-6849-12.

van Scheppingen, M. A., Jackson, J. J., Specht, J., Hutteman, R., Denissen, J. J., & Bleidorn, W. (2016). Personality trait development during the transition to parenthood: A test of social investment theory. *Social Psychological and Personality Science*, *7*(5), 452–462.

Voorpostel, M., Tillmann, R., Kuhn, U., Lipps, O., Ryser, V. A., & Schmid, F. (2016). *Swiss house-hold panel user guide (1999–2016)*. Lausanne.

Wagner, J., Becker, M., Lüdtke, O., & Trautwein, U. (2015). The first partnership experience and personality development: A propensity score matching study in young adulthood. *Social Psychological and Personality Science*, *6*(4), 455–463.

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, *29*(7), 1159-1177.

Weston, S. J., Hill, P. L., & Jackson, J. J. (2015). Personality traits predict the onset of disease. *Social Psychological and Personality Science*, *6* (3), 309–317.

Wilkins, R., Laß, I., Butterworth, P, & Vera-Toscano, E. (2019) *The Household, Income and Labour Dynamics in Australia Survey: Selected Findings from Waves 1 to 17*. Melbourne Institute: Applied Economic & Social Research, University of Melbourne.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2  *Personality and Outcome Measures Across Data Sets* | | | | | | | | | | | |
|  |  | **Ad Health** | **BHPS** | **GSOEP** | **HILDA** | **HRS** | **LISS** | **MIDUS** | **NLSY** | **SHP** | **WLS** |
| **Personality** | Extraversion | X | X | X | X | X | X | X | X | X | X |
|  | Agreeableness | X | X | X | X | X | X | X | X | X | X |
|  | Conscientiousness | X | X | X | X | X | X | X | X | X | X |
|  | Neuroticism | X | X | X | X | X | X | X | X | X | X |
|  | Openness to Experience |  | X | X | X | X | X | X | X | X | X |
|  | Self-Esteem | X |  | X |  |  | X | X | X | X | X |
|  | Optimism / Pessimism |  | X | X |  | X |  | X |  | X | X |
|  | Satisfaction with Life |  | X | X | X | X | X | X | X | X |  |
|  | Positive Affect | X |  | X | X | X | X | X |  | X | X |
|  | Negative Affect | X |  | X | X | X | X | X |  | X | X |
|  | Locus of Control | X | X | X | X | X |  | X | X | X | X |
|  | Social Support |  | X | X |  | X | X | X | X |  |  |
|  | IQ | X |  |  | X | X | X | X | X |  | X |
|  | Depression | X | X | X | X | X | X | X | X | X | X |
| **Outcomes** | Marriage | X | X | X | X | X | X | X | X | X | X |
|  | Move in with a partner | X | X | X | X |  | X | X | X | X |  |
|  | Divorce | X | X | X | X | X | X | X | X | X | X |
|  | Child Birth | X | X | X | X |  | X | X | X | X | X |
|  | Mortality | X | X | X | X | X |  | X | X | X | X |
|  | First Job | X | X | X | X |  | X | X | X | X |  |
|  | Retirement |  | X | X | X | X | X | X |  | X | X |
|  | Unemployment | X | X | X | X | X | X | X | X | X | X |
|  | Volunteering | X | X | X | X | X | X | X | X | X | X |
|  | Child Moves Out |  | X | X | X | X | X |  | X | X |  |
|  | Major Health Event | X | X | X | X | X | X | X | X | X | X |
|  | Mental Health Event | X | X | X | X | X | X | X | X | X | X |
|  | Criminality | X |  |  | X |  |  | X | X |  | X |
|  | Higher Ed | X | X | X | X | X | X | X | X | X | X |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 3  *List of Personality Measures, Scales, and Available Waves Across Studies* | | | | |
| Measure | Study | Source | Scale | Used (Available) |
| Big 5 | Ad Health | 6 (N), 3 (E), and 4 (C) item scales (Young & Beaujean, 2011) | \*\*1 "strongly agree" to 5 "strongly disagree" | 1995 (1995, 1996, 2001) |
| BHPS | 15 item BFI-S (Donnellan & Lucas, 2008; John, Naumann, & Soto, 2008) | 1 "does not apply at all" to 7 " applies perfectly" | 2005 (2004) |
| GSOEP | 15 item BFI-S (John, Naumann, & Soto, 2008, and Lang, Lüdtke, & Asendorpf, 2001) | 1 "does not apply at all" to 7 " applies perfectly" | 2005 (2005, 2009, 2013, 2017) |
| HILDA | 30 items from the TDA-40 (Saucier, 1994) | 1 "does not describe me at all" to 7 "describes me very well" | 2005 (2005, 2009, 2013, 2017) |
| HRS | 25 adjectives (Lachman & Weaver, 1997) | \*\*1 "a lot" to 4 "not at all" | 2006/8 (2006/8, 2010/12, 2014/16) |
| LISS | 50 items IPIP-50 (Goldberg, 1992) | 1 "very inaccurate" to 5 "very accurate" | 2008 (2008-2018) |
| MIDUS | 25 adjectives (Lachman & Weaver, 1997) | \*\*1 "a lot" to 4 "not at all" | 1994 (1994, 2004, 2013) |
| NLSY | 10 item TIPI (Gosling, Rentfrow, & Swann, 2003) | 1 "disagree strongly" to 7 "agree strongly" | 2006 (2006-2018) |
| SHP | 10 item TIPI (Gosling, Rentfrow, & Swann, 2003) | 0 "not at all" to 10 "completely" | 2009 (2009, 2010, 2011) |
| WLS | 34 items from the "Big Five" inventory (John, Donahue, & Kentle, 1991) | \*\*1 "agree strongly" to 6 "disagree strongly" | 1992/3 (1992/3, 2003/4) |
| Self-Esteem | Ad Health | 6 items (e.g. "You have a lot of good qualities") | \*\*1 "strongly agree" to 5 "strongly disagree" | 1995 (1995, 1996, 2001) |
| GSOEP | 1 Item (positive attitude toward myself) | 1 "does not apply" to 7 "applies fully" | 2010 (2010, 2015, 2016, 2017) |
| LISS | 10-item Rosenberg Self-Esteem Scale (Rosenberg, 1965) | 1 "totally disagree" to 7 "totally agree" | 2008 (2008-2018) |
| MIDUS | 7 items (e.g. "Feel no good at all at times") | \*\*1 "agree strongly" to 7 "disagree strongly" | 2004 (2004, 2013) |
| NLSY | 10 item Rosenberg Self-Esteem Scale (Rosenberg, 1965) | 1 "strongly disagree" to 4 "strongly agree" | 1994 (1994-2018) |
| SHP | 2 items (uselessness and satisfaction with self) | 0 "not at all" to 10 "completely" | 2009 (2009, 2012, 2015, 2018) |
| WLS | 9 items (e.g. "In general, I feel confident and positive about myself") | \*\*1 "agree strongly" to 6 "disagree strongly" | 1992/3 (1992/3, 2003/4) |
| Optimism / Pessimism | BHPS | 2 items ("life is full of opportunities" and "future looks good") | \*\*1 "often" to 4 "never" | 2001 (2001, 2006) |
| GSOEP | 1 item (attitude towards future) | \*\*1 "optimistic" to 4 "pessimistic" | 2005 (2005, 2009, 2013, 2017) |
| HRS | Life Orientation Test (Scheier, Carver, & Bridges, 1994) | 1 "strongly disagree" to 6 "strongly agree" | 2006/8 (2006/8, 2010/12, 2014/16) |
| MIDUS | 6 items (e.g. "optimistic about my future") | \*\*1 "agree a lot" to 5 "disagree a lot" | 2004 (2004, 2013) |
| SHP | 1 item "Are you often plenty of strength, energy and optimism?" | 0 "never" to 10 "always" | 1999 (1999-2018) |
| WLS | 5 items (e.g. "In uncertain times, I usually expect the best," Scheier, Carver, & Bridges, 1994) | \*\*1 "strongly agree" to 4 "strongly disagree" | 2003/4 (2003/4) |
| Satisfaction with Life | BHPS | 1 item ("Satisfaction with life overall") | 1 "not satisfied at all" to 7 "completely satisfied" | 1996 (1996-2008) |
| GSOEP | SWLS (Diener, Emmons, Larsen, & Griffin, 1985) | 0 "low" to 10 "high" | 2005 (1984-2017) |
| HILDA | 1 item ("How satisfied are you with your life?") | 0 "totally dissatisfied" to 10 "totally satisfied" | 2005 (2001 - 2017) |
| HRS | SWLS (Diener, Emmons, Larsen, & Griffin, 1985) | 1 "strongly disagree" to 7 "strongly agree" | 2006/8 (2006/8, 2010/12, 2014/16) |
| LISS | SWLS (Diener, Emmons, Larsen, & Griffin, 1985) | 1 "strongly disagree" to 7 "strongly agree" | 2008 (2008-2018) |
| MIDUS | 2 items ("How satisfied with life now" and "Satisfied with self") |  | 1994 (1994, 2004, 2013) |
| SHP | 1 item ("Satisfaction with life in general") | 0 "not at all satisfied" to 10 "completely satisfied" | 2001 (2001-2018) |
| Positive Affect | Ad Health | 2 items ("happy", "hopeful") | \*\*1 "strongly agree" to 5 "strongly disagree" | 1995 (1995, 1996, 2001) |
| GSOEP | 1 item ("Frequency of being happy in the last 4 weeks") | 1 "very seldom" to 5 "very often" | 2007 (2007-2017) |
| HILDA | 2 items (e.g. "Been a happy person") | \*\*1 "all of the time" to 6 "none of the time" | 2001 (2001-2018) |
| HRS | PANAS-X (Watson & Clark, 1994) | \*1 "very much" to 5 "not at all" | 2006 (2006-2016) |
| LISS | 10 items (e.g. "interested") | 1 "not at all" to 7 "extremely" | 2008 (2008-2018) |
| MIDUS | 6 items (e.g. "cheerful"; Mroczek & Kolarz, 1998) | \*\*1 "all of the time" to 5 "none of the time" | 1994 (1994, 2004, 2013) |
| SHP | 1 item ("joy") | 0 "never" to 10 "always" | 2006 (2006-2018) |
| WLS | 4 items from the Spielberger anger and anxiety indices (Siblings only; Spielberger, 1980; Spielberger et al., 1970) | 0 days to 7 days | 1993 (1993, 2003/4) |
| Negative Affect | Ad Health | 2 items ("fearful", "sad") | \*\*1 "strongly agree" to 5 "strongly disagree" | 1995 (1995, 1996, 2001) |
| GSOEP | 3 items (angry, sad, worried) | 1 "very seldom" to 5 "very often" | 2007 (2007-2017) |
| HILDA | 3 items (e.g. "Felt down") | \*\*1 "all of the time" to 6 "none of the time" | 2001 (2001-2018) |
| HRS | PANAS-X (Watson & Clark, 1994) | \*1 "very much" to 5 "not at all" | 2006 (2006-2016) |
| LISS | 10 items (e.g. "distressed") | 1 "not at all" to 7 "extremely" | 2008 (2008-2018) |
| MIDUS | 6 items (e.g. "nervous"; Mroczek & Kolarz, 1998) | \*\*1 "all of the time" to 5 "none of the time" | 1994 (1994, 2004, 2013) |
| SHP | 3 items ("angry", "sad", "worried") | 0 "never" to 10 "always" | 2006 (2006-2018) |
| WLS | 16 items from the Spielberger anger and anxiety indices (Siblings only; Spielberger, 1980; Spielberger et al., 1970) | 0 days to 7 days | 1993 (1993, 2003/4) |
| Locus of Control | Ad Health | 1 item "When you get what you want, it's usually because you worked hard for it." | \*\*1 "strongly agree" to 5 "strongly disagree" | 1995/6 (1995/6) |
| BHPS | 7 items (e.g. "Life is full of opportunities") | \*\*1 "often" to 4 "never" | 2001 (2001, 2006) |
| GSOEP | 8 items (e.g. "No one can escape their destiny" | \*\*1 "agree completely" to 4 "does not apply" | 1994 (1994, 1995, 2013) |
| HILDA | Pearlin Mastery Scale (Pearlin & Schooler, 1978) | 1 "strongly disagree" to 7 "strongly agree" | 2003 (2003, 2004, 2007, 2011, 2015) |
| HRS | 10 items (e.g. "What happens in my life is often beyond my control." Lachman & Weaver, 1998) | 1 "strongly disagree" to 6 "strongly agree" | 2006 (2006-2016) |
| MIDUS | 12 item Sense of Control Scale (Lachman & Weaver, 1998) | \*\*1 "agree strongly" to 7 "disagree strongly" | 1994 (1994, 2004, 2013) |
| NLSY | Pearlin Mastery Scale (Pearlin & Schooler, 1978) | 1 "strongly disagree" to 4 "strongly agree" | 1994 (1994-2018) |
| SHP | Perceived Constraints subscale from the MIDI Sense of Control Scale (Lachman & Weaver, 1998) | 0 "completely disagree" to 10 "agree" | 2009 (2009, 2012, 2015, 2018) |
| WLS | 2 items (e.g. "I have difficulty arranging my life in a way that is satisfying to me") | \*\*1 "agree strongly" to 6 "disagree strongly" | 1992/3 (1992/3, 2003/4) |
| Social Support | BHPS | 10 items (e.g. "importance of friends") | \*\*1 "daily" to 6 "never" | 2001 (2001, 2002, 2006) |
| GSOEP | 2 items (sum score of how many people support career and are confidants) | variety of codes indicating relationship of person | 2006 (2006, 2011, 2016) |
| HRS | 7 items (e.g. "How much can you rely on them if you have a serious problem") | \*1 "A lot" to 4 "Not at all" | 2006 (2006-2016) |
| LISS | 12 items (e.g. "Did you receive any counsel or advice from your father, over the past 3 months?") | 1 "no", 2 "once", 3 "several times" | 2008 (2008-2017) |
| MIDUS | 24 items (e.g. "Family members/friends/partner care(s) about you") | \*\*1 "a lot" to 4 "not at all" | 1994 (1994, 2004, 2013) |
| NLSY | 5 items (e.g. "How much can you open up to relatives if you need to talk about worries") | 1 "not at all" to 5 "a great deal" | 2008 (2008-2018) |
| SHP | 10 items (practical and emotional support from colleagues, neighbors, friends, close relatives, and partners) | 0 "not at all" to 10 "a great deal" | 1999 (1999-2010, 2013, 2016) |
| IQ | Ad Health | 1 Subtest of the WAIS (Picture Vocabulary) | Raw Score | 1995 (1995) |
| HILDA | 3 Subtests of the WAIS (Digits Backward, Symbol Digits, and Word Pronunciation) | Sum Score | 2012 (2012, 2017) |
| HRS | Vocab, Recall, and Mental Status | Sum Score (range 0 to 45) | 1996/8 (1996/8-2010/12) |
| LISS | BSI problems in cogntive functions | T-Score | 2008 (2008-2011) |
| MIDUS | 6 Subtests of the WAIS (Word Recall (Immediate and Delayed), Digits Backward, Category Fluency, Number Series, Backward Counting) | Sum Score (range 0 to 185) | 2004 (2004, 2013) |
| NLSY | Digit Span, Piat Reading and Math, Word Recall | Raw Score |  |
| Depression | Ad Health | 18 items from the CESD (Radloff, 1977) | 0 "never or rarely" to 3 "most of the time or all of the time" | 1995 (1995, 1996, 2001, 2008) |
|  | BHPS | Mental Health subscale from the SF-36 Scale (4 items; e.g. feeling run-down, melancholy) | \*\*1 "better than usual" to 4 "much less than usual" | 1991 (1991-2008) |
|  | GSOEP | Mental Health subscale from the SF-12 Scale (4 items; e.g. feeling run-down, melancholy, Ware, Kosinski, & Keller, 1995) | \*\*1 “Always” to 5 “Never” | 2002 (2002-2016) |
|  | HILDA | Mental Health subscale from the SF-36 Scale (4 items; e.g. feeling run-down, melancholy) | 1 "yes", 2 "no | 2001 (2001-2017) |
|  | HRS | 8 items (e.g. "Lose appetite when sad", Kessler, Mickelson, & Williams, 1999) | \*\*1 "All or almost all of the time" to 4 "none or almost none of the time" | 1992/4 (1992/4-2012/4) |
|  | LISS | 5 items (e.g. "I felt very anxious") | 1 "never" to 6 "continuously" | 2008 (2008-2018) |
|  | MIDUS | 8 items (e.g. "Lose appetite when sad", Kessler, Mickelson, & Williams, 1999) | 1 "yes", 2 "no | 1994 (1994, 2004, 2013) |
|  | NLSY | CESD (Radloff, 1977) | 0 "rarely or none of the time/1 day" to 3 "most or all of the time/5-7 days" | 1994 (1994-2018) |
|  | SHP | 1 item "Do you often have negative feelings such as having the blues, being desperate, suffering from anxiety or depression" | 0 "never" to 10 "always" | 1999 (1999-2018) |
|  | WLS | Mental Health subscale from the SF-12 Scale (4 items; e.g. feeling run-down, melancholy, Ware, Kosinski, & Keller, 1995) | 0 days to 7 days | 1992/3 (1992/3, 2003/4) |
| Note: \*\* indicates values are all reverse coded such that higher values indicate more likely, frequent, or characteristic on a measure. | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4  *Specification Curve Analysis Specifications for Each Outcome* | | | | | | | | | | | | | | | | | | | |
| **Measure** | **S** | **T** | **Marr-iage** | **Move in with Partner** | **Div-**  **orce** | **Child Birth** | **Mortality** | **First Job** | **Retire** | **Unem-ployed** | **Volun-teer** | **Child Moves Out** | **Health Event** | **Mental Health Event** | | **Criminal Behavior** | | | **Higher Ed** |
| **Personality** | | | | | | | | | | | | | | | | | | | |
| Extraversion | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Agreeableness | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Conscientiousness | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Neuroticism | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Openness to Experience | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Self-Esteem | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Optimism / Pessimism | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Subjective Well-Being | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Locus of Control | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Social Support | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| IQ | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| **Moderators** | | | | | | | | | | | | | | | | | | | |
| Age | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Age Moderator | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Education | N | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** |  | |
| Education Moderator | N | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** |  | |
| Gender | B | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Gender Moderator | B | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| SES | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
|  | C | P | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| SES Moderator | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
|  | C | P | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| **Other Covariates** | | | | | | | | | | | | | | | | | | | |
| Disease | B | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  | **X** | | **X** | | **X** | **X** | |
|  | B | P | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  | **X** | | **X** | | **X** | **X** | |
| Self-Rated Health | C | C | **X** | **X** | **X** | **X** | **X** |  | **X** |  | **X** |  | **X** | | **X** | |  |  | |
|  | C | P | **X** | **X** | **X** | **X** | **X** |  | **X** |  | **X** |  | **X** | | **X** | |  |  | |
| Smoking | B | C | **X** | **X** | **X** | **X** | **X** |  |  |  |  |  | **X** | | **X** | | **X** |  | |
|  | B | P | **X** | **X** | **X** | **X** | **X** |  |  |  |  |  | **X** | | **X** | | **X** |  | |
| Alcohol Consumption | C | P | **X** | **X** | **X** |  | **X** |  |  | **X** |  |  | **X** | | **X** | |  |  | |
|  | C | C | **X** | **X** | **X** |  | **X** |  |  | **X** |  |  | **X** | | **X** | |  |  | |
| Eating Habits | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  | **X** | | **X** | | **X** |  | |
|  | C | P | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  | **X** | | **X** | | **X** |  | |
| Activity Levels | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  |  | **X** | | **X** | | **X** | **X** | |
|  | C | P | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  |  | **X** | | **X** | | **X** | **X** | |
| BMI | C | C |  |  |  | **X** | **X** |  | **X** |  |  |  | **X** | | **X** | | **X** |  | |
|  | C | P |  |  |  | **X** | **X** |  | **X** |  |  |  | **X** | | **X** | | **X** |  | |
| Marital Status | N | C |  |  |  | **X** | **X** |  | **X** | **X** |  | **X** | **X** | | **X** | | **X** | **X** | |
|  | N | P |  |  |  | **X** | **X** |  | **X** | **X** |  | **X** | **X** | | **X** | | **X** | **X** | |
| Age at Marriage | C | C |  |  | **X** | **X** | **X** |  |  |  | **X** | **X** | **X** | | **X** | |  | **X** | |
| Years Married | C | C |  |  | **X** | **X** | **X** |  |  |  |  | **X** | **X** | | **X** | |  | **X** | |
| Number of Children | N | C |  |  | **X** |  |  |  | **X** |  |  | **X** |  | |  | | **X** | **X** | |
| Parental Divorce | B | C | **X** | **X** | **X** | **X** | **X** | **X** |  |  |  |  |  | |  | | **X** | **X** | |
| Father's Education | N | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Mother's Education | N | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Functional Limitations | B | C |  |  | **X** | **X** | **X** |  | **X** | **X** |  |  | **X** | | **X** | |  |  | |
|  | B | P |  |  | **X** | **X** | **X** |  | **X** | **X** |  |  | **X** | | **X** | |  |  | |
| Income | C | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
|  | C | P | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | | **X** | | **X** | **X** | |
| Religion | B | C | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  | **X** | **X** | **X** | | **X** | |  | **X** | |
|  | B | P | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  | **X** | **X** | **X** | | **X** | |  | **X** | |
| *Note*: Column S indicates whether variables are continuous (C), binary (B), or nominal (N), while column T indicates concurrent (C) or prior measurement to personality. | | | | | | | | | | | | | | | | | | | |