Personality Measurement and Assessment in Large Panel Surveys

Brent Roberts

Joshua J. Jackson

Angela Lee Duckworth

University of Pennsylvania, duckwort@psych.upenn.edu

Katherine Von Culin

University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/psychology_papers

Part of the Demography, Population, and Ecology Commons, Personality and Social Contexts Commons, and the Quantitative Psychology Commons

Recommended Citation


This paper is posted at ScholarlyCommons. https://repository.upenn.edu/psychology_papers/2
For more information, please contact repository@pobox.upenn.edu.
Personality Measurement and Assessment in Large Panel Surveys

Abstract
Personality tests are being added to large panel studies with increasing regularity, such as the Health and Retirement Study (HRS). To facilitate the inclusion and interpretation of these tests, we provide some general background on personality psychology, personality assessment, and the validity of personality tests. In this review, we provide background on definitions of personality, the strengths and weaknesses of the self-report approaches to personality testing typically used in large panel studies, and the validity of personality tests for three outcomes: genetics, income, and health. We conclude with recommendations on how to improve personality assessment in future panel studies.

Keywords
health and retirement study

Disciplines
Demography, Population, and Ecology | Personality and Social Contexts | Psychology | Quantitative Psychology
Personality Measurement and Assessment in Large Panel Surveys*

Brent Roberts, Joshua J. Jackson, Angela L. Duckworth, and Katherine Von Culin

Abstract

Personality tests are being added to large panel studies with increasing regularity, such as the Health and Retirement Study (HRS). To facilitate the inclusion and interpretation of these tests, we provide some general background on personality psychology, personality assessment, and the validity of personality tests. In this review, we provide background on definitions of personality, the strengths and weaknesses of the self-report approaches to personality testing typically used in large panel studies, and the validity of personality tests for three outcomes: genetics, income, and health. We conclude with recommendations on how to improve personality assessment in future panel studies.

KEYWORDS: Health and Retirement Study

*Please send correspondence to Brent W. Roberts, Department of Psychology, University of Illinois, Urbana-Champaign, 603 East Daniel Street, Champaign, IL, 61820; Email: bwrobrts@illinois.edu.
Personality psychology concerns itself with variation across individuals and how the individual differences shape people’s lives and society’s structures. Personality is clearly a multifaceted system and can be conceptualized at many different levels of analysis (John, Robins, & Pervin, 2008). We prefer to simplify personality by dividing it into four correlated, but conceptually distinct categories. Specifically, most individual differences can be thought of as falling into the domains of Abilities, Traits, Motives, and Narratives (Roberts & Wood, 2006). In a nutshell, abilities concern what people are capable of, traits reflect what people typically think, feel, and do, motives subsume what people want or desire, and narratives reflect the particular stories of peoples’ lives. For the remainder of this paper, we will concentrate on personality traits. This is not to say the other domains of personality are irrelevant. Rather, it is tacit acknowledgement that personality traits are the domain most often discussed when considering the inclusion of personality variables into any study, such as the HRS. It is also an acknowledgment that with the exception of cognitive ability, we know more about personality traits than any of the domains of personality described above.

Personality traits are defined as the relatively enduring patterns of thoughts, feelings, and behaviors that differentiate individuals from one another and are elicited in trait affording situations (Roberts, 2009). Currently, most personality psychologists accept the Big Five (Extraversion, Agreeableness, Conscientiousness, Emotional Stability/Neuroticism, and Openness to Experience) as an adequate working taxonomy of personality traits. According to three recent reviews, personality traits matter for many important life outcomes, but in particular for outcomes related to health, work, and relationships (e.g., Caspi, Roberts, & Shiner, 2005; Ozer & Benet-Martinez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007).

In terms of health and longevity, personality traits predict objective health outcomes, such as life expectancy, and the range of mechanisms thought to affect health and life expectancy (Adler & Matthews, 1994). Traits such as conscientiousness are critical for the health-related behaviors that are known to undermine or promote health, such as tobacco, alcohol, and drug consumption, risky sexual activities, risky driving, healthy eating and activity level (Bogg & Roberts, 2004). Numerous studies of hostility have shown that it not only predicts problematic health behaviors, but also physiological factors such as cardiovascular reactivity, which plays a significant role in heart disease (Smith, 2006). And, in a testament to the fact that personality traits predict multiple health factors, most personality traits are associated with mortality at levels similar to or higher than socioeconomic status (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007).

Achievement outcomes were long thought to be the exclusive purview of cognitive ability (Heckman, Stixrud, & Urzua, 2006). Numerous studies have
shown that personality traits, such as conscientiousness, predict grades and overall GPA above and beyond cognitive ability in elementary school (Abe, 2005), secondary school (Duckworth & Seligman, 2005), and college (Chamorro-Premuzic & Furnham, 2003; Conard, 2005; Noftle & Robins, 2008). Similarly, personality traits predict a whole suite of work-related outcomes. Personality traits predict how individuals go about the job search process (Wanberg, Glomb, Song, & Sorenson, 2005), and which types of jobs people are interested in, therefore helping to shape the career pathways people choose (Mount, Barrick, Scullen, & Rounds, 2005). Moreover, personality traits predict work behaviors such as absenteeism (Ones, Viswesvaran, & Schmidt, 2003), job satisfaction (Judge, Heller, & Mount, 2002), leadership (Judge, Bono, Ilies, & Gerhardt, 2002), and counterproductive work behaviors (Roberts, Harms, Caspi, & Moffitt, 2008). Finally, personality traits predict immediate outcomes, such as job performance (Hogan & Holland, 2003), as well as long-term occupational attainment and income above and beyond cognitive ability (Heckman et al., 2006; Judge, Higgins, Thoresen, & Barrick, 1999).

Personality traits also play an important role in shaping relationships and marriages. People who are more neurotic are less securely attached (Noftle & Shaver, 2006), and tend to be less satisfied with their partners—regardless of who their partners are (Robins, Caspi, & Moffitt, 2000). In contrast, people who are more agreeable and open to experience have more satisfied partners (Donnellan, Conger, & Bryant, 2004). Given these findings, it is not too surprising to note that people who are more agreeable, open, and less neurotic are perceived as more desirable mates (Botwin, Buss, & Shackelford, 1997). It turns out these preferences are adaptive. The personality traits of agreeableness and emotional stability, as well as conscientiousness predict a significantly lower probability of experiencing divorce (Roberts et al., 2007).

Given these findings alone, we believe that HRS and other panel studies should invest more in assessing personality with a particular focus on personality traits. Nonetheless, some of the evidence cited above is not directly relevant to particular panel studies and may be construed as less than definitive for methodological reasons (e.g., it is predominantly cross-sectional). Being sensitive to the particular set of issues that confront research on populations, such as retirees in the HRS, we have compiled a more focused review of the literature on three topics: The genetics of personality traits, personality traits and wealth, and personality traits and health.

We have refined our review with the following ideas in mind. First, there is often an implicit if not explicit circularity in personality research (if not psychological research in general). For example, researchers often ask people whether they work hard in general and then use answers to these questions to predict whether they work hard in a particular setting (e.g., their job). Economists
describe this as the endogeneity problem. To this end, we have tried to review research that reduces the endogeneity problem, typically through employing more rigorous methodological designs. For example, we have emphasized prospective longitudinal studies, and where possible, those studies that bridge key transitions, such as using personality assessed before people enter the labor market to predict labor market outcomes. We also focus on studies that employ multiple methods, such as observer reports of personality, which significantly diminish the circularity problem. We emphasize research that uses objective criteria that are not gathered using self-report techniques. Finally, we focus on research that is most relevant to the age period of the HRS and other ongoing panel studies.

**Personality and Molecular Genetics**

From our perspective, personality traits provide a clear phenotypic conduit through which genetic variation will affect important life outcomes, such as health and wealth (Roberts & Jackson, 2008). Behavior genetics studies have shown repeatedly that personality traits are genetically influenced, with estimates of heritability ranging from 40 to 60% (Krueger & Johnson, 2008). Of course, heritability estimates are fraught with indeterminacy, so the argument that personality traits are ideal phenotypic conduits of genetic factors would be bolstered by research showing that genetic polymorphisms reliably predict individual differences in personality traits.

Personality traits were first associated with specific genetic polymorphisms almost fifteen years ago (Benjamin et al., 1996; Ebstein et al., 1996). Since then, a large number of genetic polymorphisms have been associated with personality traits. In this section we review the accumulated evidence linking genetic polymorphisms to personality traits, and offer future directions in the search for the genetic architecture of personality. To guide our review we will focus on genetic polymorphisms associated with the Big Five—extraversion, agreeableness, conscientiousness, neuroticism, and openness (Goldberg, 1993). Most studies in the personality genetics literature rely either on single dimension scales (e.g., the sensation seeking scale; Zuckerman et al., 1978) or on two omnibus personality measures: The Temperament and Character Inventory (TCI; Cloninger et al., 1993) and the NEO (either the shorter FFI version or the longer NEO-PI-R; Costa & McCrae, 1992). The TCI does not neatly overlap with the more accepted Big Five model (Farmer & Goldberg, 2008) and thus interpreting the results of the TCI within the framework of the Big Five is difficult (for a review of the findings of TCI/TPQ, see Ebstein, 2006; Munafo et al., 2003; Noblett & Coccoro, 2005; Reif & Lesch, 2003).
Extraversion

The personality trait of extraversion refers to an energetic approach toward the social and material world and includes lower order traits such as sociability, activity, assertiveness, and positive emotionality (John & Srivastava, 1999). Theoretical models of the neurobiology of extraversion posit that the dopamine and opioid systems likely govern sensitivity to reward and positive emotions (Depue & Collins, 1999; Depue & Morrone-Strupinsky, 2005). Correspondingly, most studies focus on polymorphisms associated with dopamine transport and reuptake. For example, the exon III repeat polymorphism in the DRD4 gene is linked to both sensation seeking and extraversion more generally (Benjamin et al., 1996; Bookman, Taylor, Adams-Campbell, & Kittles, 2002; Eichhammer et al., 2005; Golimbet et al., 2007; Ozkaragoz & Noble, 2000; Tochigi et al., 2006). Other dopamine receptors (e.g. DRD2 & DRD3) have also been associated with extraversion (Ratsma et al., 2001).

Additionally, the Val/Val repeat polymorphism in the COMT gene—which catabolizes dopamine at a faster rate than the met/met counterpart and results in a significant reduction of synaptic dopamine—is associated with greater levels of extraversion (Golimbet et al., 2007; Reuter & Hennig, 2005; Wacker et al., 2010). Moreover, a recent study using a haplotype approach (combing many unlinked Single Nucleotide Polymorphisms or SNPs) examined 36 markers within the COMT gene and found multiple markers associated with extraversion (Stein et al., 2005). A recent application of the multiple marker approach associated polymorphisms in multiple dopamine genes (DDC, DAT1, DBH) with sensation seeking (Derringer et al., 2010). While less work has been done with opioid related genes, preliminary evidence suggests that the opioid receptor gene OPRM1 is associated with extraversion (Luo et al., 2008).

Agreeableness

The personality trait of agreeableness contrasts a prosocial orientation towards others with antagonism and includes lower order traits such as altruism, compliance, trust, and modesty (John & Srivastava, 1999). Most work with the personality trait of agreeableness focuses on antisocial or aggressive behavior on the negative end and trust on the positive. Less evidence exists regarding the biological substrate of agreeableness so no neurobiological systems have been studied extensively. That said, there is evidence that agreeableness is associated with multiple biological systems. Similar to extraversion, Val/Val repeat polymorphism in the COMT gene is associated with aggression and antisocial behavior (Volavka et al., 2004; Tiitinen et al., 1999). Additionally, the serotonin system is associated with agreeableness (Carver & Miller, 2006). Variation in the
promoter region of the serotonin transporter gene has also been associated with agreeableness (Canli & Lesch, 2007; Jang et al., 2001; Wand et al., 2002).

Given that agreeableness is related to substance abuse (Martin & Sher, 1994), it is not surprising that genes directly related to substance abuse are associated with agreeableness. Variation in the M2 cholinergic receptor (CHRM2), which is responsible for the Acetylcholine release, is associated agreeableness (Luo et al., 2007a). A polymorphism in the receptor region of the ADH4 gene, which influences the expression of the enzyme that metabolizes acetaldehyde, is also associated with agreeableness (Luo et al., 2007b). The Cannabinoid receptor 1 gene (CNR1) gene has been related to levels of agreeableness (Juhasz et al., 2009). Additionally, a recent genome wide association found the Clock gene was associated with agreeableness (Terracciano et al., 2010). In sum, there are numerous examples that agreeableness is associated with genetic polymorphisms. However, the specific neurobiological systems involved are not entirely clear. Future investigations of agreeableness would profit to focus on genes also implicated in substance abuse.

**Conscientiousness**

Conscientiousness is the tendency to be controlled, task- and goal-directed, norm following, responsible, planful, and organized (John & Srivastava, 1999). Similar to agreeableness, there is less theory about the underlying biological mechanisms involved in the personality trait of conscientiousness. However, also similar to agreeableness, conscientiousness is associated with substance abuse and other psychiatric disorders that are characterized by low levels of impulse control, such as substance abuse and ADHD (Nigg et al., 2003; Widiger, 2008). Thus, many of the associations with conscientiousness involve polymorphisms already associated with substance dependence or psychiatric disorders.

Consistent with this the DRD4, COMT, GABRA1, GABRA6, TPH1, CHRM2 and MAOA polymorphisms have all been associated with impulsivity and conscientiousness more broadly (Kreek et al., 2005; Dragan & Oniszczenko, 2007; Rosenberg et al., 2006; Tochigi et al., 2006; Luo et al., 2007). These are the same polymorphism linked to ADHD and other disorders marked by impulsivity (e.g., Dick et al., 2006; 2008; Swanson et al., 2000).

A recent investigation found that the Dopamine-β-hydroxylase (DBH) gene was associated with levels of conscientiousness (Hess et al., 2009). Links between conscientiousness and the serotonin system also exist (Carver & Miller, 2006). For example, variations in the serotonin receptor 2A gene (5-HT2A) and serotonin transporter gene (5-HTT) are related to conscientiousness (Heck et al., 2009; Tochigi et al., 2006). All in all, the genes related to conscientiousness appear to be genes that are also related to disorders that index a lack of control—
or conversely, that are marked by impulsivity. Thus, investigating the neurophysiology of impulsivity should also inform the neurophysiology of conscientiousness.

**Neuroticism**

The personality trait of neuroticism contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad, and tense (John & Srivastava, 1999). Many studies examine the neurobiology of neuroticism because of the overlap with anxiety and many psychiatric disorders. Perhaps the most commonly studied gene associated with neuroticism is a repeat in the promoter region of the serotonin transporter gene (5-HTTLPR) first reported by Lesch et al., (1996). A number of meta-analyses of the association between this gene and neuroticism exist, though the results are equivocal. One meta-analysis found no effect; another found an effect and a third found that the effect depended on the measure used (Schinka et al., 2003; Sen et al., 2004a; Munafo et al., 2003). Regardless of the specific meta-analysis, it appears that the higher bound effect of the serotonin transporter gene on neuroticism does not exceed 3% (Munafo et al., 2007).

A number of other genes have been associated with Neuroticism, many of which are involved with the serotonin system. For example, genes associated with the transport, reuptake, catabolization or reception of serotonin such as COMT, 5-HT1A, 5-HT2A, and TCAT have been associated with neuroticism (Stein et al., 2004; Strobel et al., 2003). Additionally, BDNF, DRD2, the tyrosine hydroxylase repeat polymorphism (TCAT), and GABA6 were all associated with neuroticism (Ebstein, 2006; Perrson et al., 2000; Sen et al., 2004b). In contrast to extraversion, which is associated primarily with the dopaminergic genes, neuroticism appears to be primarily associated with serotogernic genes.

**Openness**

Openness to Experience is defined as preferring depth, originality, and complexity in an individual’s mental and experiential life (John & Srivastava, 1999). Even though the heritability of openness is the same as the other Big Five traits, fewer studies have investigated the neurobiology of openness. However, given that openness is reliably associated with intelligence, the neurobiology of intelligence may offer insights on which neurobiological system is involved in the trait of openness (DeYoung et al., 2009). The COMT gene, which has previously been associated with memory, is associated with the personality trait of openness (Harris et al., 2005). Additional studies also find that the serotonin transporter
(Harro et al., 2009) and the corticotropin-releasing hormone receptors 2 (CRHR2) genes are associated with openness (Tochigi et al., 2006).

While the findings from research linking specific polymorphisms to personality traits are promising, a number of challenges exist in the field of personality genetics. First, many of these genes are associated with multiple personality traits as well as other psychological variables like IQ (e.g. COMT), implying that polymorphisms associated with psychological variables are pleiotropic. Second, the replication rate is poor. The small number of meta-analyses conducted to date find that polymorphisms explain very little variance in personality traits, typically just a few percent (e.g., Munafo et al., 2003). A number of alternative approaches to examining specific polymorphisms have been proposed, including genome-wide association studies, combining multiple methods of assessment to estimate phenotypes, and gene-by-environment interactions, which we review below.

**Genome wide association studies**

The studies reviewed above rely on a priori hypothesized associations with personality traits. In contrast, a genome wide association study (GWAS) takes a less hypothesis driven approach by analyzing numerous SNPs across the entire genome (usually around 500,000 markers). Large-scale GWAS studies are touted as optimal research strategies to uncover the genetic basis for complex phenotypes. Currently, there are a small number of genome wide association studies for personality traits (Terraciano et al., 2010; 2010; Shifman et al., 2008; Krueger et al., 2010; Gillespie et al., 2008). Unfortunately, the results of GWAS studies fail to replicate the same polymorphisms reviewed above and also mostly fail to replicate each other in a recent meta-analysis of GWAS studies (DeMoor et al., in press). Moreover, the SNPs that are significantly associated with personality traits in GWAS studies rarely account for more than 1% of the variance. If all of the significant SNPS are aggregated together the amount of explained variance is around 5%. However, the inability to identify replicable genetic polymorphisms that are strongly associated with personality traits is not unique to personality traits and extends to any polygenic trait. For example, the replication rate and the amount of variance explained are on par with GWASs of heart disease and even height, both of which are highly heritable (Krueger et al., 2010).

A future direction for GWAS data sets is to combine many genetic markers together at the same time in a theory driven approach. The relative number of SNPs assessed in GWAS studies allows one to aggregate polymorphisms thought to influence the same neurophysiological system instead of focusing on one polymorphism at a time (Plomin, Haworth, & Davis, 2009). If
the initial GWAS studies and meta-analyses are correct, then a single SNP will not be able to explain more than a one or two percent of the variance in any outcome. Combining multiple SNPs together into what is known as a SNP set allows the investigation of many numerous genes at the same time. Recently, this approach was adopted in an examination of dopamine system and its relation to sensation seeking. Roughly three hundred SNPS from twelve genes were identified as involved in the dopamine system and possible related to sensation seeking. Together the SNP set comprising 12 SNPS correlated .20 with sensation seeking (Derringer et al., 2010; McCrae et al., 2010).

Multiple methods

Another potential way to enhance the personality signal that is being detected by genetic approaches is to use additional methods other than self-reports. For example, to our knowledge no study of genetic polymorphisms has used observer reports of personality. Additionally, an approach that focuses on more basic biological markers that mediate the pathways from gene to behavior may find stronger links to genetic polymorphisms. This endophenotype approach has been successfully used in the substance abuse literature (Iacono et al., 2008) and may prove useful for personality traits. These endophenotypes are usually assessed by endocrinological or neurophysiological methods, such as EEG. Interestingly, a recent GWAS study found SNPs that explained almost 9% of the variance in resting EEG beta waves (Hodgskinson et al., 2010). Neuroimaging (fMRI) also offers a way to examine the mediating pathways of gene effects. For example, the serotonin transporter gene is associated with less grey matter in ACC and amygdala for short allele carriers compared to long allele carriers (Pezawas et al., 2005).

Environments

Thus far the majority of studies are predicated on the assumption of additive genetic effects, which assume a one-to-one association between genotype and phenotype. The additive model of genetic effects may lack efficacy or generalizability because increasing evidence points to the conditional nature of the genome (Robinson, Fernald & Clayton, 2008). That is, it may be more common that the genome interacts multiplicatively with environmental factors, in a process in which specific genes are turned on at various times in the life course, and possibly turned off thereafter (Rutter, Moffitt, & Caspi, 2006). Thus combining an informed assessment of relevant environmental experiences with genetic assessments is likely to enhance our understanding of how genes influence personality and consequentially important life outcomes.
Integrating environments with genetic testing may be accomplished many ways but the two most useful approaches are to investigate how genes lead a person to experience different environments (gene-environment correlations) and how people respond differently to an environment (gene-environment interaction). Many relevant environments and environmental experiences are already assessed in the HRS and other panel studies, such as poor health or low SES. These experiences could be utilized in gene-environment interaction research where some genetic diathesis interacts with environmental experiences to predict personality traits and other outcomes.

Moreover, the gene-environment interplay approach is consistent with the previous recommendations for how to successfully incorporate genetic information into ongoing longitudinal research. For example, environments have been successful incorporated into self-report association studies (Caspi et al., 2002); neuro-imaging (Canli, 2008) and even with GWAS studies. For example, a recent study using a GWAS found that the effect of each SNP depended on the environment one was raised in (Dick et al., 2010). If this is common—and evidence suggests that it is (e.g., Moffitt, Caspi & Rutter, 2005)—then incorporating environments in GWAS studies could greatly increase the overall variance explained in phenotypes of interest, such as personality.

Despite the promise of these alternative approaches, a number of difficulties remain for the field of personality genetics. The uniformly large estimates of the heritability of personality combined with the unimpressive results of candidate gene and GWAS approaches to detecting the underlying genes calls into question some of the initial assumptions behind broad-based genetic testing. For example, the first generation of research focusing on specific candidate genes and GWAS was based on the assumption that relatively common genetic variants will have strong and direct causal links to personality traits or other complex phenotypes. Clearly, this assumption has proven overly optimistic at best as was anticipated by geneticists a decade ago (Terwilleger, 2001). The most consistent finding across all efforts to detect and replicate specific genetic effects on personality (or other phenotypes) is that the effect of any given polymorphism, whether direct, or in interaction with an environment, is remarkably small. As such, alternative models of genetic effects must further be explored. Many rare variants of genes may be responsible for complex phenotypes, or other forms of genetic variation, such as copy number variants, may play a more important role than SNPs. Future efforts to link genetic variation to complex phenotypes, such as personality traits, would be wise to keep the remarkably small effect sizes of these initial efforts in mind. Detecting and replicating gene-personality association, or even gene-environment interplay will be hampered by low power, even in studies as large as the HRS.
Personality and Economic Outcomes

The notion that dispositions to act, think, and feel in certain ways likely influence earnings and wealth seems intuitive. One need only consider acquaintances who do better or worse, in economic terms, to generate hypotheses about the traits that might reliably aid or impair economic performance. Nevertheless, labor economists have traditionally treated human capital as isometric with cognitive ability and knowledge. That is, the idea that personality traits might also influence productivity and success in the labor market is relatively new to most economists (Borghans et al., 2008).

To date, interest has outpaced actual research studies on the role of personality traits in determining economic outcomes. For example, we were able to identify only one published article relating Big Five personality traits to lifetime saving and borrowing behavior. Nyhus and Webley (2001) used a survey of 734 Dutch households which included questions about household assets and debt as well as personality. More emotionally stable and introverted individuals saved more and borrowed less. More agreeable individuals, in contrast, saved less and borrowed more. Contrary to the authors’ hypotheses, conscientiousness was not related to either saving or borrowing. Notably, the personality traits of both household heads and partners explained variance in outcomes, though the effects were typically not symmetric.

More research has examined associations between personality traits and earnings. A meta-analysis by Ng et al. (2005) identified 7 studies reporting associations between Big Five traits and salary prior to 2003. We identified an additional 12 studies, representing a total sample of $k = 19$, $N = 22,652$. Table 1 summarizes correlations from this updated meta-analysis. Notably, all estimated effect sizes were relatively small in magnitude but, given the sample size, statistically significant.¹

As shown in Table 1, $Q$-statistics for heterogeneity in effect size were significant, and $I^2$ estimates for proportion of this variance that is systematic were high for all personality trait-salary associations. We therefore undertook two moderator analyses to explain the systematic variance in effect size estimates. First, given the wage gap separating women from men (Weinberg, 2007) and gender differences in personality (Costa, Terracciano, & McCrae, 2001), as well as the fact that women represented only 36% of our meta-analytic sample, we tested gender as a continuous moderator for each trait-earnings correlation. Second, given considerations of reverse causality and third-variable confounds in cross-sectional studies, we tested study design ($k = 11$ cross-sectional studies vs. $k$)

¹ A description of each of these studies and notes on our analytic approach are available on request.
= 8 longitudinal studies) as a categorical moderator for each trait-earnings correlation. Three out of ten moderator analyses reached statistical significance, and we discuss these findings in the context of the main analyses below.

As shown in Table 1, emotional stability was the strongest Big Five correlate of earnings, \( r = .13 \), corrected \( \rho = .14 \). This finding is consistent with Hogan and Holland (2003), who found adjustment, a measure of emotional stability, to be the most potent predictor of occupational performance. The positive association between emotional stability and earnings is also congruent with research on core self-evaluations, a trait shown theoretically and empirically to overlap with emotional stability (Judge, Erez, Bono, & Thoresen, 2002) and defined as one’s subjective assessment of one’s capabilities and control over the environment. Judge and Hurst (2007) have shown that core self-evaluations potentiate the effect of early life advantages on mid-life income. Specifically, youth with more positive core self-evaluations were more likely to capitalize on early advantages such as academic achievement and family socioeconomic status. Positive core self-evaluations also predict better job performance, higher job satisfaction, lower levels of stress and conflict, and better coping with setbacks (Judge, 2009). A disposition toward sadness, anxiety, hostility, and other negative emotions would be expected to undermine performance at work. Indeed, emotional stability reliably predicted job performance (corrected \( \rho = .13 \), 90% CI \([.01, .22]\)) in a quantitative summary of 5 independent meta-analyses examining associations between Big Five personality traits and job performance (Barrick et al., 2001). We speculate that, in addition, more emotionally stable individuals may earn higher salaries because they are more likely to pursue lucrative but stressful professions.

It is also possible, of course, that higher earnings contribute to higher levels of emotional stability. Indeed, in a longitudinal study of Baltimore residents, income measured in middle adulthood prospectively predicted increases in emotional stability a decade later (Sutin, Costa, Miech, & Eaton, 2009). Finally, the observed association between emotional stability and earnings could reflect unmeasured third variables associated with both variables. For example, certain vocational interests are related to particular Big Five personality traits (Larson, Rottinghaus, & Borgen, 2002) and certain vocations (e.g., investment banker) tend to have higher salaries than others (e.g., painter). However, emotionally stability is not robustly related to any of the six vocational interests in the RIASEC\(^2\) taxonomy (Larson et al., 2002).

---

The six domains of vocational interests in the RIASEC model are Realistic, Investigative, Artistic, Social, Enterprising, and Conventional
Table 1

Meta-analytic Associations between Big Five Personality Traits and Salary

<table>
<thead>
<tr>
<th>Big Five Factor</th>
<th>r</th>
<th>95% CI</th>
<th>ρ</th>
<th>95% CI</th>
<th>Q</th>
<th>I2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Stability</td>
<td>.13*</td>
<td>.10 - .16</td>
<td>68.17*</td>
<td>.14*</td>
<td>.11 - .17</td>
<td>73.60%</td>
<td>83.79***</td>
<td>77.56%</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.10*</td>
<td>.06 - .14</td>
<td>117.41*</td>
<td>.11*</td>
<td>.07 - .16</td>
<td>84.67%</td>
<td>152.37***</td>
<td>88.03%</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.06*</td>
<td>.03 - .09</td>
<td>52.43*</td>
<td>.07*</td>
<td>.04 - .10</td>
<td>65.67%</td>
<td>66.22***</td>
<td>72.60%</td>
</tr>
<tr>
<td>Openness</td>
<td>.06*</td>
<td>.02 - .10</td>
<td>129.91*</td>
<td>.06*</td>
<td>.01 - .12</td>
<td>86.14%</td>
<td>224.54***</td>
<td>91.96%</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>-.04*</td>
<td>-.07 - -.01</td>
<td>61.20*</td>
<td>-.04*</td>
<td>-.07 - -.01</td>
<td>70.59%</td>
<td>91.18***</td>
<td>79.26%</td>
</tr>
</tbody>
</table>

Note. *p < .05. There were 19 samples with an aggregate sample of N = 22,652. CI = confidence interval; r = correlation; ρ = correlation corrected for scale reliability; Q = Cochran’s measure of homogeneity; I2 = Higgins and Thompson’s (2002) measure of heterogeneity.
Table 1 shows that extraverts earned higher salaries, $r = .10$, corrected $\rho = 11$. Extraverts do not, however, reliably receive higher job proficiency ratings (corrected $\rho = .15, 90\%\ CI\ [-.03, .27]$) (Barrick et al., 2001). Moreover, moderator analyses indicated that the effect of extraversion is inversely proportional to the percentage of women in the sample, $r = -.19, p < .001$. Given that our meta-analytic sample was disproportionately male, it seems reasonable to assume that in a balanced sample of men and women, the association between extraversion and earnings would be substantially diminished. In our view, therefore, the evidence that more sociable, gregarious, assertive, and lively individuals generally perform better than their more introverted colleagues is equivocal. Consistent with the comparatively high Q-statistic for extraversion-earnings correlations, it is possible that dominance and sociability facets of extraversion are differentially related to both job performance and earnings, and we recommend future studies using facet-level personality measures test this hypothesis. Future studies might also test the possibility that extraverts select into more highly paid professions (e.g., sales; management positions) than do introverts, possibly because extraverts are more likely to hold enterprising vocational interests ($r = .41$; Larson et al., 2002).

Conscientiousness was positively related to salary ($r = .06$, corrected $\rho = .07$) but more modestly than expected in overall analyses. Moderator analyses indicated that only conscientiousness was a significantly better predictor of earnings in longitudinal studies ($r = .14$, corrected $\rho = .16$) than in cross-sectional studies ($r = .03$, corrected $\rho = .04$). The stronger estimates in longitudinal studies are consistent with the finding that conscientiousness predicts job performance (corrected $\rho = .27, 90\%\ CI\ =\ [.10, .35]$) better than any other Big Five trait (Barrick et al., 2001). We can only speculate as to why longitudinal findings differ from cross-sectional findings for conscientiousness but not for other Big Five traits. One possibility is that conscientious individuals tend to hold themselves to higher standards, and, thus, rate themselves lower than they should on self-report questionnaires. Such a bias should attenuate associations with any outcome, including earnings. Consistent with this speculation, individuals in East Asian cultures, including Korea, China, and Japan, rate themselves lower in conscientiousness than individuals in North America, South America, Europe, and Africa (Schmitt, Allik, McCrae, & Benet-Martínez, 2007).

What else might explain the relatively modest association between conscientiousness and earnings in cross-sectional studies? Conscientiousness is not robustly associated with any dimension of vocational interest in the RIASEC taxonomy; but it is nevertheless possible that more conscientious individuals select themselves into less lucrative professions. Another possibility is that conscientious individuals perform better on the job yet fail to realize income commensurate with their superior performance, at least earlier in their careers.
Finally, distinct facets of conscientiousness could have opposing effects on salary – for instance, a positive association between salary and industriousness might negate a negative association between salary and traditionalism? Published research cannot resolve these important questions, suggesting future studies are needed to unravel the potentially complex causal pathways linking conscientiousness and its facets to earnings in labor market.

Openness to experience was also associated with higher salaries, $r = .06$, corrected $\rho = .06$. Gender significantly moderated this relation: the effect of openness was positively associated with the proportion of women in the sample, $r = .14$, $p < .001$. Given that the overall meta-analytic sample was disproportionately male, the openness-earnings association might be larger in the population. However, the openness-earnings association could also reflect, at least in part, the third-variable confound of intelligence. General intelligence is both associated with openness to experience ($r = .33$, Ackerman & Heggestad, 1997) and earnings (Heckman, Stixrud, & Urzua, 2006). Unfortunately, most studies of personality and earnings did not report openness-earnings associations that controlled for intelligence. A notable exception is a longitudinal study by Judge et al. (1999) using data from three studies that followed participants from early childhood to retirement. Childhood openness to experience predicted higher earnings in late adulthood ($\beta = .26$, $p < .05$), but this effect diminished to non-significance ($\beta = -.02$, ns) when childhood intelligence was entered as a covariate. Similarly, among 5,000 adults who graduated from Wisconsin high schools 35 years prior, the cross-sectional association between openness to experience and earnings ($\beta = .10$, $p < .001$ for men; $\beta = .12$, $p < .001$ for women) diminished when controlling for intelligence measured in high school ($\beta = .06$, $p < .001$ for men, $\beta = .09$, $p < .001$ for women) (Mueller & Plug, 2006). Since individuals more open to experience tend to have artistic ($r = .48$) and investigative ($r = .28$) interests (Larson et al., 2002), one might expect them, in fact, to select into less lucrative professions (e.g., musician, professor) than individuals of comparable intelligence and opportunity.

Should future studies confirm that openness is correlated with higher earnings, even when controlling for IQ, what might explain this association? Meta-analytic estimates of the effect of openness to experience on job performance are small and non-significant, corrected $\rho = .07$, 90% CI = [-.09, .19] (Barrick et al., 2001). In contrast, openness to experience is more strongly associated with years of education ($r = .31$, $p < .001$) than any other Big Five trait (Goldberg, Sweeney, Merenda, & Hughes, 1998). Thus, to the extent that individuals who are more creative, curious, original, and intellectual earn more than their counterparts, it may be because they enter the labor market with better academic credentials rather than because of superior performance on the job.
Finally, Table 1 shows that more agreeable individuals earn lower salaries ($r = -.04$, corrected $\rho = -.04$), a somewhat surprising finding given that the ability to get along with others would seem advantageous in most occupations. Agreeable individuals do not reliably earn better job performance ratings, corrected $\rho = .13$, 90% CI $[-.01, .22]$ (Barrick et al., 2001). It is possible that agreeable individuals may “get along” better than they “get ahead” (Hogan & Holland, 2003). For instance, kind, trusting, empathic individuals may not be sufficiently aggressive when negotiating salary contracts and otherwise optimizing their own welfare when in conflict with others. Likewise, agreeable individuals may not assert themselves as effect leaders when their decisions conflict with the opinions of colleagues. Finally, agreeable individuals, who have more social vocational interests ($r = .19$, Larson et al., 2002), may also select into helping professions (e.g., social work) with lower salaries.

**Personality and Health**

The research linking personality to health is less nascent than that between personality and wealth. It is now commonly accepted that personality, and personality traits in particular, play a significant role in the health process (see Hampson, 2008 for a brief review). The following review highlights studies that rely on study features described above that diminish the endogeneity problem.

Several studies have examined the prospective relation between childhood personality and adult health outcomes. The first study is the Hawaii Personality and Health cohort, a longitudinal study of 1,054 individuals who were attending elementary school in Hawaii between the years 1959 and 1967 (Hampson, Goldberg, & Dubanoski, 2007). These same individuals were re-contacted in 1999 and 2000 at which time they provided information on their education, health behaviors, and health. A measure of the childhood Big Five traits was used to predict intervening variables of educational attainment and health behaviors, which were in turn used to predict overall health. Four of the Big Five had direct or indirect relations to health status assessed 40 years later in midlife. Conscientiousness was positively related to educational attainment as well as health status, which indicated that conscientiousness had a direct effect on health above and beyond education and health behaviors. Extraversion had a complex relationship to health. It was positively related to physical activity, which was in turn positively related to health status. On the other hand, it was also positively related to smoking behavior, which was negatively related to health status. Agreeableness was negatively related to smoking behavior and positively related to educational achievement, and thus had a positive effect on health in midlife. Openness to experience was only related to higher educational attainment.
In a recent paper (Moffitt et al., 2011), childhood self-control was used to predict adult (age 32) health, wealth, and crime in the Dunedin Multidisciplinary Health and Development study (N = 1037). Self-control was assessed with a composite of ratings done by the participants, their parents, teachers, and trained clinicians. Physical health was assessed through a amalgam of objective indices of cardiovascular functioning, respiratory health, dental quality, sexual health, and inflammatory factors. In all cases, the effect of childhood self-control on adult health was conditioned against and compared to childhood SES and IQ. Low childhood self-control predicted lower physical health. The effect of low self-control was consistently independent of childhood SES and IQ. Moreover, across all outcomes, the effect sizes for childhood low self-control, SES, and IQ were comparable. The effect of low self-control on adult health was explained, in part, by adolescent “snares”, such as dropping out of school and smoking tobacco. These intervening factors highlight the fact that personality traits do not necessarily affect outcomes such as health and wealth directly, but more likely through problematic or protective behaviors taken up on the way to poor or good health.

A number of studies have examined personality and health in middle and older aged samples similar to the HRS. Several studies have used the MIDUS sample for example. In the first study, personality traits were used to predict Body Mass Index independent of parental levels of occupational status and adult levels of occupational status, education, and income (Chapman, Fiscell, Duberstein, Coletta, Kawachi, 2009). Parental occupational status was a protective factor, but its effect was fully explained by the effect of adult socioeconomic indicators. Unfortunately, there were no consistent SES predictors of obesity, such as adult levels of education or income. In contrast, the personality trait of conscientiousness predicted lower BMI scores even when taking into account childhood and adult measures of socioeconomic status.

Several studies have linked conscientiousness to physical health using more objective indices of either conscientiousness or physical health. For example, observer ratings of conscientiousness compiled from friends, family members, and interviewers predicted self-report ratings of physical health (Lodi-Smith et al., 2010). In fact, observer rated conscientiousness predicted above and beyond self-reported ratings of conscientiousness. Alternatively, self-reported ratings of conscientiousness predicted lower medical illness burden as rated by doctors in a clinical setting in a sample of patients between the ages of 65 and 97 (Chapman, Lyness, & Duberstein, 2007). Finally, in a recent paper we used personality ratings gathered from the subsample of HRS couples assessed in 2008 to predict physical health (Roberts, Smith, Jackson, & Edmonds, 2009). Like many studies we found that self-reported conscientiousness and neuroticism predicted self-reported physical health. In addition, we found that if a person was
married to a spouse who was more conscientious, then his or her own ratings of health were higher, which we dubbed the compensatory conscientiousness effect.

Finally, personality traits have been linked to mortality with increasing regularity (Roberts et al., 2007). One excellent example of a long-term prospective study is the Scottish Mental Survey (SMS; Deary, Batty, Pattie, & Gale, 2008). The SMS has been tracked since 1947 at which time measures of IQ were acquired. Three years after, in 1950, personality data were collected from teachers. Despite relatively impoverished measures of personality (3-item, post-hoc scales), ratings of dependability (a facsimile of conscientiousness), predicted mortality over the subsequent 55 years. Moreover, the effect of dependability (hazard ratio = .78) was similar in magnitude to the effect of IQ (hazard ratio = .70), and independent of the effect of family background, BMI, education, and childhood illness. A more recent study of the Edinburgh Artery Study (EAS), which is similar in age and composition to the HRS sample (older, diverse in terms of SES) showed similar effects for conscientiousness and extended the analyses to the remaining Big Five (Michelle et al., 2009). In this study, both conscientiousness and openness emerged as protective factors, in particular for men. When examined in a structural equation-modeling framework, conscientiousness predicted all-cause mortality above and beyond age, social class, smoking, and blood pressure at a similar magnitude ($\beta = -.14$) as social class ($\beta = .14$).

As can be seen by these studies, personality traits play an important role in shaping the health experiences people have and in turn the length and quality of their lifespan. We have highlighted studies that diminish the endogeneity problem, such that we find personality assessed well-before health problems arise can predict important outcomes, as well as the fact that it can predict objective health outcomes.

**General Recommendations for the Use and Enhancement of Personality Assessment in Ongoing Panel Studies**

Personality traits have had a somewhat quixotic existence in large panel studies. For example, in the HRS, there was an initial effort to assess personality in 1996 deemed prematurely to be unsuccessful (see below), and then several different variants incorporated more recently. Specifically, the trait adjectives used in the MIDUS study were assessed in the 2006 HRS assessment along with several additional items intended to expand the assessment of conscientiousness and neuroticism drawn from Lew Goldberg’s IPIP system. Subsequently, a newer measure of a more elaborated model of conscientiousness was incorporated in the 2008 assessment (Hill et al, under review). Going forward, we would propose that personality traits be systematically assessed in panel studies such as the HRS.
Because of the demands of assessment in these large studies (e.g., the cost of increasing the survey length), we make this recommendation with three overarching constraints. First, the assessment needs to be short. Second, specific traits that are more predictive of relevant outcomes should be assessed more thoroughly. And third, the personality measure should be “harmonized” with other studies, which entails using the same or overlapping measures of personality.

In terms of brevity, there are thankfully many good options available. For example, the HRS borrowed the adjectival system from the MIDUS study, and this adjective set works quite well and is short. There are other options, such as the Big Five Inventory (BFI; John & Srivastava, 1999), NEO-FFI (Costa & McCrae, 1992), and the Ten-Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003). In our opinion, any of these four assessment tools would work well, but we would recommend using the BFI because it has been used more widely in large panel studies. For example, some form of the BFI is now being used in the Wisconsin Longitudinal Study, GSEOP, and British Household Panel Study (BHPS). If the goal is to create a body of data that can be translated across cultures and samples, employing the BFI would be a better option than the MIDUS adjectives.

The third constraint on investing more in personality assessment is the option to drill deeper and assess specific traits in more detail. Enhancing the assessment of specific personality traits needs to be weighed against both the costs of asking more questions as well as the incremental utility of the specificity of measures beyond the Big Five level of assessment. For example, a number of researchers associated with large panel studies utilized by economists have expressed the desire to assess “self-control” more directly and in a more expansive way because of the assumption that it should predict economic behavior, such as savings, income, or some form of delay discounting. Whether or not this is a good idea depends on 1) whether self-control is the right construct, and 2) whether it provides anything beyond what one already gets with a generic Big Five measure. Both of these issues come down to a question of validity. Is the outcome one has in mind, such as savings, going to be predicted better by a measure of self-control than some other measure and specifically a measure of conscientiousness? The above review of the links between the Big Five and income would argue for digging deeper in traits other than conscientiousness.

A straightforward way to answer the questions that arise from considering the inclusion of more focused personality assessments is to run pilot research before including new measures of personality. Far too often the decision to include a new measure of personality rests on intuitions and face validity rather than good empirical work. For example, it is surprisingly common to include personality measures in large panel studies without running pilot research. Take
for instance the idea of linking self-control to savings data or delay discounting. According to recent research (Hirsh et al., in press), self-control as typically assessed has little or no relation to delay discounting. However, positive affect—a component of extraversion—does predict delay discounting quite consistently. Alternatively, what one calls self-control may be indistinguishable from a more simple conscientiousness scale. For example, in our ongoing research we have found very little evidence that facets of conscientiousness, including self-control, predict outcomes such as health behaviors or emotions above and beyond the latent trait of conscientiousness (Fayard et al., in press). Studies like the HRS are too large and costly to be haphazard about selecting variables.

With that strongly worded caveat aside, what more specific measures would we recommend? In terms of dimensions related to health and economic outcomes, we would recommend assessing self-control and specific components of conscientiousness in more detail. Candidate systems and measures would be Donald Lynam’s UPPS system (Whiteside & Lynam, 2001) for assessing impulsivity, which breaks self-control down into four facets. Adding some form of Duckworth’s (2009) Grit scale would also be sensible to predict financial outcomes. Joireman’s consideration of future consequences (Joireman et al., 2005) is also a possibility, as it is akin to a general tendency to delay gratification.

Moving beyond the general recommendation to invest more in the assessment of personality, we also recommend that alternative methods of assessment be employed. As many large panel studies inevitably assess several members of a family, the simplest, and most accessible option given is to use spouse ratings or family ratings of the Big Five as observer ratings of personality. These additional ratings can be used to form a composite with self-reports or as a separate variable altogether. It is not uncommon, as we reported above, to find that observer ratings provide incremental validity above and beyond self-reported personality ratings. Certain forms of more “objective” indices of personality could also be considered. For example, there are a variety of experimentally induced inhibitory control measures that could be employed, such as go-no go tasks, flanker tasks, and risk tasks, in addition to self-reports. Like observer ratings of personality, these tasks typically complement self-reports (Edmonds, et al., 2008). As noted above, when combined with self-reports, these additional dimensions could be linked not only to outcomes of interest, but also to GWAS analyses.

We also recommend reassessing personality traits using identical measures in future waves of research and thus use them as both independent and dependent variables of interest. One reason to consider personality traits as dependent variables is that personality-trait change may be quite consequential for people. Mroczek and Spiro (2007) demonstrated that long-term increases in neuroticism were predictive of mortality in an 18-year survival analysis. Those who started
high on neuroticism (above the sample median) and increased over 10 years had higher mortality, controlling for age, depression, and physical health. Similarly, we have recently conducted analyses showing that changes in conscientiousness predict changes in health behaviors and physical health over and above original standing on conscientiousness across two longitudinal studies (Takahashi, et al, in preparation).

Finally, we highly recommend incorporating some system to catch aberrant responding to the survey. In this case, aberrant responding refers to either responding randomly to the survey or, more typically, to employ what we call the “flush-right strategy.” The latter refers to a propensity for fatigued or demotivated participants to simply start circling or bubbling in responses down the extreme right hand column of the survey. We sometimes find this type of responding when we ask undergraduates to complete voluminous surveys for very little in return for their effort. Ironically, we have also found evidence of these forms of malingering in populations that are being well reimbursed (e.g., they are going through the motions just to get the money).

We make this recommendation based on our preliminary work on the 1996 and 2008 HRS personality data. In both cases, the psychometric qualities of the personality scales were puzzling. For example, the negatively keyed items (e.g., shy on an extraversion scale) were uncorrelated with the positive items (e.g., talkative). This pattern of results is common when there is aberrant responding to the personality items. When examined more closely, we found evidence for the flush-right strategy, in which a significant minority of persons were rating all items highly even if they were directly contradictory. When we controlled for this form of aberrancy, the psychometric properties of the personality measures improved markedly. Furthermore, when we examined the correlates of this type of responding we found that it was more often employed by older, less educated, less healthy women, who also happened to be lower on conscientiousness. It appears that a significant minority of the HRS sample became too fatigued to respond validly to the survey questions. Moreover, when we controlled for the aberrant responding the validity of the scales also emerged in ways that one would expect. Specifically, we now show that the measure of conscientiousness assessed in 1996 predicts mortality over the subsequent 10-year period and that this effect is independent of SES and IQ (Hill et al, in press). We recommend incorporating a systematic assessment of aberrant responding, which would not necessitate adding many more items to the survey. We feel this is of critical importance because most of the economic measures in the survey do not have balance of positively and negatively keyed items. Therefore, it is possible that people are completing the survey with a flush-right strategy and this pattern is going undetected as their data is not missing, but clearly invalid.
Conclusion

In conclusion, we recommend that the initial efforts to incorporate personality measures into the HRS and other panel studies be continued and expanded judiciously. Ideally, new constructs would be pilot tested first. We also recommend employing more than self-report methods to assess personality traits, with the least intrusive option being to use observer ratings. Adding validity scales would also be ideal. In terms of the genetic analyses, the most important recommendation would be to think clearly about the “environmental experience” measures. Does any given study really have adequate coverage of the type of environmental insults or palliative experiences that might interact with genetic polymorphisms? The answer to this question rests in a clearly informed theoretical model of adult development in old age, which should be a tractable challenge.

References

References marked with an asterisk indicate that the studies were included in the meta-analysis.


Dick et al., 2010. Unpublished Data.


Krueger et al., 2010 GWAS, unpublished manuscript


