Early investigations of the neuroticism by conscientiousness interaction with regards to health have been promising, but to date, there have been no systematic investigations of this interaction that account for the various personality measurement instruments, varying populations, or aspects of health. The current study – the second of three – uses a coordinated analysis approach to test the impact of the neuroticism by conscientiousness interaction on the prevalence and incidence of chronic conditions. Using 15 pre-existing longitudinal studies (N > 49,375), we found that conscientiousness did not moderate the relationship between neuroticism and having hypertension (OR = 1.00, 95% CI [0.98, 1.02]), diabetes (OR = 1.02 [0.99, 1.04]), or heart disease (OR = 0.99 [0.97, 1.01]). Similarly, we found that conscientiousness did not moderate the prospective relationship between neuroticism and onset of hypertension (OR = 0.98 [0.95, 1.01]), diabetes (OR = 0.99 [0.94, 1.05]), or heart disease (OR = 0.98 [0.94, 1.03]). Heterogeneity of effect sizes was largely nonsignificant, with one exception, indicating that the effects are consistent between datasets. Overall, we conclude that there is no evidence that healthy neuroticism, operationalized as the conscientiousness by neuroticism interaction, buffers against chronic conditions.

Keywords: healthy neuroticism; conscientiousness; chronic condition; health; coordinated analysis; longitudinal

Defined by heightened negative affect and emotional instability, the personality trait neuroticism is associated with nearly all aspects of poor health, from daily health behaviors such as smoking (Hakulinen et al., 2015) and alcohol consumption (Luchetti et al., 2018), to the increased risk of developing chronic conditions (Jokela et al., 2014; Weston et al., 2015), to greater symptomatic complaints (Costa Jr. & McCrae, 1987), and to mortality...
(Graham et al., 2017). Findings such as these have led some to regard neuroticism to be a public health concern (Lahey, 2009). Yet, this characterization of neuroticism is premature because the literature examining the neuroticism-health relationship is quite mixed. While many studies find neuroticism associated with poor health, other studies find no association between these constructs (e.g., Friedman, Kern, & Reynolds, 2010; Jokela et al., 2013). Still others find that neuroticism is linked to better health, most notably greater longevity (Brickman, Yount, Blaney, Rothenberg, & De-Nour, 1996; Ragland & Brand, 1988).

One possible explanation for conflicting findings is that neuroticism may influence health through multiple pathways (Friedman, 2000). If this is true, the heterogeneity of findings represents truly different and conflicting mechanisms through which neuroticism both improves and weakens health. Hypothetical pathways connecting neuroticism to negative health outcomes include taking part in unhealthy behaviors, avoiding medical care, and experiencing poorer outcomes. In contrast, there may be pathways connecting neuroticism to positive health outcomes through vigilance towards changes in health, as well as seeking out and complying with medical advice. It is unclear whether a person high in neuroticism can take multiple pathways simultaneously or switch from one pathway to another during the course of their life or are confined to the path they first walk down. These different possibilities would indicate different types of predictors: situational, developmental/life-stage relevant, or stable, respectively.

To date, one promising candidate predictor is the personality trait of conscientiousness. The notion that another personality trait shapes the path of a person high in neuroticism is consistent with early descriptions of “healthy neuroticism” (Friedman, 2000), which divided people into types: unhealthy and healthy neurotics (pp. 1102). Based solely on description, conscientiousness does appear to be a likely predictor of healthy neurotic behavior. Conscientious individuals avoid unhealthy behaviors (Bogg & Roberts, 2004) and adhere to medication properly (Hill & Roberts, 2011; O’ Cleirigh, Ironson, Weiss, & Costa Jr., 2007). Indeed, the confluence of high conscientiousness and high neuroticism, which some have defined as being “overcontrolled” (Costa Jr. & McCrae, 1998), has been shown to be associated with lower rates of smoking (Terracciano & Costa Jr., 2004; Vollrath & Torgersen, 2002), alcohol use (Turiano, Whiteman, Hampson, Roberts, & Mroczek, 2012), and inflammation (Turiano, Mroczek, Moynihan, & Chapman, 2013).

That being said, this combination of traits is not consistently associated with better health. For example, the combination of high conscientiousness and high neuroticism does not reliably predict smoking (e.g., Turiano et al., 2012) or may only do so for certain populations, such as those diagnosed with chronic conditions (Weston & Jackson, 2015). Finally, given the propensity for null results to go unpublished (Bakker, Dijk, & Wicherts, 2012; Franco, Malhorta, & Simonovits, 2014), it is difficult to determine how many times the combination of neuroticism and conscientiousness has been tested and resulted in non-significance. Thus, it is possible that the interaction between neuroticism and conscientiousness is unrelated to health and appears in the literature merely through sampling error or publication bias.

The current study is the second in a series of three studies investigating how neuroticism, conscientiousness, and their interaction are associated with health outcomes and mortality. The first study in this series examines these traits in relation to health behaviors (i.e., smoking, drinking, and physical activity; Graham et al., this issue); the current and second study examines chronic conditions; and the third study examines mortality (Turiano et al., this issue).

**Chronic Conditions**

There is evidence that personality is associated with chronic condition status. Decreasing optimism is generally linked with greater likelihood of developing a health condition (e.g., stroke, diabetes, hypertension; Chopik, Kim, & Smith, 2015), and neuroticism is prospectively linked to the development of hypertension and heart disease (Weston et al., 2015). Conscientiousness, on the other hand, has been found to be generally protective. For example, conscientiousness is associated with lower rates of hypertension and diabetes (Goodwin & Friedman, 2006; Weston et al., 2015). We build upon this work by addressing whether the combination of high neuroticism and high conscientiousness offsets the risk of developing health conditions.

To narrow the scope of this study, we chose to examine chronic conditions that are both widely prevalent and heavily influenced by behavior. These were hypertension, diabetes, and cardiovascular disease, which are among the 10 most prevalent chronic conditions in American adults, with hypertension being the most prevalent (Gerteis et al., 2015). Further, cardiovascular disease is ranked the
number one leading cause of death worldwide (National Center for Health Statistics, 2017) and was estimated to cost Americans more than $300 billion in 2012–2013 (Benjamin et al., 2017). During the same period, diabetes was estimated to cost more than $245 billion (American Diabetes Association, 2013). These chronic conditions represent some of the most severe costs to Westernized societies, both physically and financially. Moreover, each is linked to smoking, diet, and physical activity (Wingard, Berkman, & Brand, 1982), the health behaviors we investigated in the first project of this series.

Coordinated Analysis

Coordinated analysis, a form of integrative data analysis (Curran & Hussong, 2009), involves the direct comparison of results based on independent analysis of multiple data sets (Hofer & Piccinin, 2009, 2010). This method is sometimes referred to as a two-step individual participant data meta-analysis (Hakulinen et al., 2015; Riley, Lambert, & Abo-Zaid, 2010). Coordinated analyses strengthen the interpretation of results in several ways. By combining the individual results from multiple samples, the statistical power to detect the effect is greatly increased. Differences in key study characteristics (e.g., measurement of variables, the historical era in which the study took place, the age range and other characteristics of the samples, the number and timing of measurement occasions, the types of sampling procedure used, etc.) can provide a challenge to researchers who seek to understand similarities and differences among results from these samples. One solution is to harmonize variables across studies – for example, by creating binary variables that represent the absence and presence of a given condition – to reduce between-study differences. However, coordinating at the lowest possible denominator limits the quality of data available within studies, and may not be suitable for some variables. Further, too much harmonization on the variable level may severely limit the generalizability or even the interpretability of results.

Another solution is to harmonize models rather than measures (Hofer & Piccinin, 2009, 2010). To do so, identical (or nearly identical) statistical models are built using conceptually similar, rather than identically measured, constructs. Harmonizing at the level of the model is one of the strengths of coordinated analysis of existing longitudinal data. Specifically, researchers can examine the heterogeneity of studies not as error but as potential sources of variability. For example, systematic differences between studies collected during different historical eras are evidence of cohort effects. However, it must be acknowledged that including additional variables means adding researcher degrees of freedom (John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011). Thus, it is imperative that studies that use coordinated analysis pre-register their analytic plans, so as to delineate their planned and exploratory comparisons.

The current study used coordinated analysis to examine the interaction between neuroticism and conscientiousness and its relationship to the diagnosis of three major chronic conditions. We ask two main research questions. First, to what extent is the interaction between neuroticism and conscientiousness concurrently associated with having been diagnosed with hypertension, diabetes and heart disease. Second, to what extent is the interaction between neuroticism and conscientiousness prospectively associated with the development of hypertension, diabetes and heart disease. By using coordinated analysis, we can estimate these relationships first within individual studies and then across all studies. These analyses allow us to estimate a population-level effect size (with the population limited to the cultures and countries in which these data were collected), as well as the potential heterogeneity of each effect.

The current study is the second in a series of three studies submitted together. These studies were the result of a single, coordinated project involving multiple co-principal investigators and research labs around the world. The goal of this project was to rigorously analyze the evidence for “healthy neuroticism,” narrowly defined as the significant moderation of the neuroticism-health relationship by conscientiousness. This analysis considered multiple components of health, specifically behaviors, development of (chronic) conditions, and mortality, following a lifespan trajectory. Despite the coordination and similarity of analyses among these three components of health, there were notable differences between the studies, including the datasets which could be used in the analysis, the quantitative models applied, and the theoretical rationale linking personality to health. Moreover, synthesizing this project in a single manuscript would have required over-simplification of critical details and decision points along the way. As a result, the three components were divided into three separate but linked manuscripts, allowing for the requisite detail to be included in each.

Methods

Supplemental material

Raw data could not be made public, due to data sharing agreements by the organizations that collected the data. Results of within-study analyses and meta-analyses are available at osf.io/48fhe. Supplemental files – including R code, additional analyses, and more detailed study information – can be found at IALSAging.github.io/HealthyN.

Studies and Participants

Study information and descriptive statistics can be found in Table 1.

The Berlin Aging Study II (BASE-II) consists of a subsample of younger (20–35 years of age) and older adults (60–84 years of age) who were recruited from the greater metropolitan area of Berlin (for an overview, see Bertram et al., 2013; Gerstorf et al., 2016). Starting in 2009, a total of 1,437 ($M_{age} = 60.16$, $SD_{age} = 15.77$, 50% female) participants were eligible for the current analyses.

The Einstein Aging Study (EAS) is an observational longitudinal cohort study of cognitive aging and dementia, which began in 1993. Older adults who were at least 70 years of age, non-institutionalized, and native English speakers were systematically recruited from an
urban, multi-ethnic, community-dwelling population in Bronx County, New York, USA. Participants receive comprehensive annual medical and neuropsychological evaluations (Katz et al., 2012). A total of 734 (Mean = 78.83, SD = 5.31, 61% female) participants were eligible for the current analyses.

The English Longitudinal Study of Ageing (ELSA) is a longitudinal cohort survey that collects multidisciplinary information on older adults living in England. Data collection began in 2002, and new participants were added at waves 3, 4, 6 and 7 to maintain size and representativeness (Marmot et al., 2017). A total of 6,263 (Mean = 66.39, SD = 8.55, 56% female) participants were eligible for the current analyses.

The Health and Retirement Study (HRS) is a longitudinal panel that tracks retirement-age adults in the United States. Data collection began in 1992, with new cohorts added throughout the 1990s (Sonnega et al., 2012). A total of 18,925 (Mean = 66.28, SD = 11.14, 58% female) participants were eligible for the current analyses.

The Interdisciplinary Longitudinal Study of Adult Development and Aging (ILSE) is a multidisciplinary longitudinal study investigating the aging process of two German birth cohorts born between 1930–1932 and 1950–1952 (see Sattler et al., 2015 for an overview). For the current analysis, only individuals from the older cohort born between 1930–1932 were included (e.g., Aschwanden, Kliegel, & Allemand, 2018). Data collection began in 1993–1996 and a total of 478 participants (Mean = 62.51, SD = 0.96, 52% female) were eligible for the current analyses.

The Lothian Birth Cohort 1936 (LBC1936) consists of individuals who were born in 1936 and completed the Scottish Mental Survey in 1947 (Deary, Gow, Pattie, & Starr, 2012; Taylor, Pattie, & Deary, 2018). The LBC1936 cohort was recruited between 2004 and 2007 by identifying individuals from the original cohort who were residing in Edinburgh and the surrounding areas. In total, 1,091 participants entered the study. A total of 959 participants were eligible for the current analyses (Mean = 69.50, SD = 0.84, 51% female).

The Long Beach Longitudinal Study (LCLS) started in 1978 and was made up of 28–84 year old participants from southern California. This sample was reassessed in 1994–1995, and has since been assessed two additional times (2000–2002 and 2008–2013). Additional cohorts were added in the second two waves of data collections (Zelinski & Kennison, 2001). A total of 935 participants were eligible for the current analyses (Mean = 68.44, SD = 12.91, 55% female).

The Memory and Aging Project (MAP) is a longitudinal, epidemiologic clinical-pathologic cohort study of common chronic conditions of aging with emphasis on decline in cognitive and motor function and risk of Alzheimer’s disease (Bennett et al., 2018; Bennett, Schneider, et al., 2012). Participants are older adults recruited from retirement communities and subsidized senior housing facilities throughout the Chicago metropolitan area and northeastern Illinois. Participants do not have known dementia at baseline. A total of 604 participants (Mean = 79.68, SD = 7.14, 76% female) were eligible for the current analyses.

The Midlife in the United States Study (MIDUS) is an ongoing nationally representative study of 7,108 participants in the United States recruited in 1994/1995. Since then, it has added two waves of data collection,
one in 2004/2005 and one in 2013/2014 (Brim, Ryff, & Kessler, 2004). A total of 5,988 participants were eligible for the current analyses (Mean = 46.85, SD = 12.91, 52% female).

The Veterans Affairs Normative Aging Study (NAS) is a study of the medical and psychosocial aging among men in the United States and is funded by the United States Department of Veterans Affairs (Bosse, Ekerdt, & Silbert, 1984). The sample was originally based in the Greater Boston, Massachusetts metropolitan area and consisted of 2,280 men enrolled from 1961 to 1970. The participants were on average 42 years old at enrollment. A total of 820 participants were eligible for the current analyses (Mean = 64.39, SD = 7.24).

The Older Australian Twins Study (OATS) is a multisite longitudinal study of monzygotic and dizygotic twins aged at least 65 years, with a cohort of 623 participants assessed at baseline, and is the largest, most comprehensive study of older twins in Australia (Sachdev et al., 2009). At present, three waves, each spaced two years apart, have been completed, although due to cohort attrition, only 391 participants from the initial cohort have completed the third wave. A total of 463 participants (Mean = 71.28, SD = 5.37, 66% female) were eligible for the current analyses.

The Religious Orders Study (ROS) is a longitudinal, epidemiologic clinical-pathologic cohort study of aging and Alzheimer’s disease that enrolls older Catholic nuns, priests, and brothers from more than 40 groups across the United States (Bennett et al., 2018, 2012). Participants do not have known dementia at baseline. A total of 1,326 (Mean = 75.94, SD = 7.43, 71% female) participants were eligible for the current analyses.

The Sydney Memory and Ageing Study (MAS) is an ongoing longitudinal cohort study of brain aging and dementia in older individuals, who undertake medical, neuropsychological and psychosocial assessments approximately every two years. Individuals aged 70–90 and living in the Australian community at baseline were randomly recruited through the electoral roll (Sachdev et al., 2010). The baseline MAS cohort comprised 1,037 individuals without dementia, of whom 860 were eligible for the current analyses (Mean = 78.66, SD = 4.76, 46% female).

The Seattle Longitudinal Study (SLS) started in 1956 and has since collected data on close to 6,000 participants in a cohort-sequential design (Schaie, Willis, & Caskie, 2004). Participants were sampled randomly from members of a large health maintenance organization in the Seattle, Washington area. A total of 876 participants (Mean = 68.27, SD = 13.63, 56% female) were eligible for these analyses.

The Wisconsin Longitudinal Study (WLS) follows a cohort of men and women who graduated from Wisconsin high schools in 1957. Data from graduate participants (N = 10,317) span almost 60 years from the baseline assessment in 1957, with follow-up assessments collected up to five times, ending in 2011 (Herd, Carr, & Roan, 2014). In addition to the original cohort, subsequent assessments included randomly selected siblings and spouses of graduate participants. A total of 10,560 participants (Mean = 53.72, SD = 4.45, 54% female) were eligible for these analyses.

**Measures**

Collectively, the measures of personality traits used covered a wide range of narrow constructs that are typically assessed by the broader Big Five model. It is worth noting some systematic differences between the scales. For example, the IPIP-50 measure of conscientiousness included items assessing responsibility, practicality, and thriftiness but not self-discipline or efficiency; in contrast, the BFI measures competency and achievement but not goal-striving. To some extent, these differences in trait coverage across scales is an asset to these analyses: a lack of significant differences in effects would suggest that estimated relationships are robust to choice of scale, while significant differences may point to specific mechanisms (i.e., narrower traits) which may underlie and inspire investigation of causal mechanisms. We provide a table of content measurement by scale online on the page [Personality Scale Content](https://alsaging.github.io/HealthyN). Supplementary files are available at IALS Aging.

**Personality**

Personality traits were assessed using various measures of the Big Five (i.e., neuroticism, conscientiousness, extraversion, agreeableness, and openness to experience). Different measures of the same trait are highly correlated (Lutjeijn, Starren, & Van Dijk, 2000; McCrae & Costa Jr., 1985), which allows for a comparison of the effects of the same construct across studies. All but two of the data sets had measures of all five personality traits. The MAP study had measures of neuroticism, conscientiousness and extraversion, and the MAS had neuroticism, conscientiousness and openness. The most commonly used measures were from the NEO family of instruments (Costa Jr. & McCrae, 2008). The NEO-PI-R was used in the LBLS, OATS and MAS, and the NEO-FFI (Costa Jr. & McCrae, 1989) was used in the ILSE, ROS, and MAP.

Other measures included the Big Five Inventory (John & Srivastava, 1999), used in the BASE-II and WLS; the Midlife Developmental Inventory Personality Scale (Lachman & Weaver, 1997), a short adjective scale developed for panel studies, used in the HRS, MIDUS and ELSA; the IPIP-50 (Goldberg et al., 2006), used in the EAS and the LBC 1936; and Goldberg’s (1992) 50 adjectives, used in the MAS. Cronbach’s alpha was calculated from the samples of participants who were eligible for analyses. Internal reliability estimates for neuroticism ranged from 0.66 (BASE-II) to 0.93 (SLS). Internal reliability estimates for conscientiousness ranged from 0.56 (MIDUS) to 0.91 (NAS). When available, the other personality traits were included as covariates in the models; reliability estimates...
for the personality covariates are included online (Descriptive Statistics).

**Chronic Condition Status**

In 12 of the 15 studies, participants were asked, “Have you ever been diagnosed with” or “has a doctor or nurse ever told you that you have” for each of several chronic conditions. Their answers were coded 1 for yes and 0 for no. This method was used to represent hypertension, diabetes and heart disease in BASE-II, ELSA, HRS, LBC, MIDUS, NAS, ROS, MAP, EAS, OATS, MAS, LBLS and WLS. Some studies asked about hypertension only (BASE-II, EAS, ILSE, NAS), others asked about hypertension or high blood pressure (ELSA, HRS), and others only asked about high blood pressure (LBLS, ROS, MAP, OATS, MAS). Only the protocols in the EAS distinguished between Type I and Type II diabetes. For heart condition, we included any condition related to cardiovascular health. For specific conditions listed in each study, see the online supplemental (Study Information). Finally, ILSE protocols included listening to heart sounds as part of a medical checkup conducted by one to two trained study geriatricians. For all studies, we simplified coding to 0 (not diagnosed with condition) and 1 (diagnosed with condition).

**Covariates**

Models were adjusted for the following covariates: age, sex, education, body mass index, other chronic conditions, extraversion, agreeableness and openness to experience. Age in all studies was age at first personality assessment (i.e., baseline age), standardized within study by subtracting the study’s mean age and dividing by its standard deviation. We included average study age as a between-study variable when examining heterogeneity between studies. Sex was coded as 0 for male and 1 for female. Education in BASE-II, EAS, HRS, LBLS, LBC, ROS, OATS, MAS, and MAP was measured as the number of years of education. In ELSA, ILSE, MIDUS, NAS, and WLS, education was measured using an ordinal scale that referred to the highest degree earned; for the purposes of harmonization, we treated these as interval variables. We standardized education within study by subtracting the study’s mean education value and dividing by its standard deviation. Body mass index (BMI) was measured as kg/m² or lb/in² × 702. Height and weight were self-reported in BASE-II, HRS, LBLS, LBC, MIDUS, and WLS while researchers or medical professionals measured height and weight in ELSA, ILSE, NAS, ROS, EAS, OATS, MAS, and MAP. Again, BMI was standardized within study. Other chronic conditions indicated whether a person had (1) or had not (0) been diagnosed with any of the other chronic conditions assessed by the study. Personality traits other than neuroticism and conscientiousness that were measured were also included as covariates.

**Between-study variables**

Diabetes assessment indicated whether a measure distinguished between Type I and Type II diabetes (0) or specifically measured only Type II diabetes (1; the latter included the ELSA, the NAS, the MAS, the OATS, the LBLS and the prospective measurement in the WLS). Mean study age was the average age of participants in the study, prior to standardization. Average study length and maximum study length were measures of the amount of time between baseline assessment and final chronic condition assessment. The average study length ranged from 3.74 years (OATS) to 18.00 years (WLS) and the maximum study length elapsed from 5 years (BASEII and ELSA) to 26 years (NAS).

**Data analysis**

We used R (Version 3.6.2; R Core Team, 2018) for our meta-analyses and visualization. The inferential models were run on each study's data using binary logistic regression. Functions in the *metafor* package (Version 2.0.0; Viechtbauer, 2010) were used to estimate the overall effects and heterogeneity between studies, as well as to create forest plots. The *sPlot* package (Version 2.5.0; Lüdecke, 2018) was used to calculate predicted values for each study and the *ggplot2* package (Version 3.0.0; Wickham, 2016) was used to visualize effects. Additional package information, including the version used for each individual study analysis, is provided online (R Packages Used).

**Individual Study Analysis**

We ran the same inferential models within each dataset. First, the presence of hypertension, diabetes and a heart condition were each separately regressed onto neuroticism, conscientiousness, and their interaction, using data from each participant's first personality assessment occasion. These models controlled for age, sex, education, the other personality traits in the study, BMI, and whether the participant had been diagnosed with another chronic condition. Second, the presence of hypertension, diabetes, and a heart condition at the last study wave to date for each individual were each separately regressed onto baseline neuroticism, conscientiousness, and their interaction, excluding participants diagnosed with the outcome at baseline. In other words, the models examine the association between personality and the prospective development (i.e., incidence) of the chronic condition. All predictors and covariates were measured at baseline (i.e., the first time the participant had completed the personality assessment).

**Meta-analyses**

The analytic tools of meta-analysis were used to estimate the average weighted effect size of the interaction of neuroticism and conscientiousness in each of the models described above. As part of this estimation, we calculated the heterogeneity between the studies. Finally, we examined the role of three between-study variables in explaining any variability in the effect. Those variables were the personality scale, the method of assessing chronic condition status, and, in the case of diabetes as an outcome, whether the study included Type I diabetes in their assessment. When examining variation in the association of personality with the prospective development of chronic conditions, we also examined the...
average number of years between personality assessment and the most recent assessment, as well as the maximum number of years.

All models used listwise deletion. Sample sizes for each model in each dataset are presented in the relevant figure or table.

**Power analysis (post-hoc)**

Given the anticipated sample size, we did not believe a power analysis was necessary with regards to our interaction coefficient estimate, or the coefficient estimates of neuroticism and conscientiousness. However, our power to detect significant heterogeneity in effect sizes between studies is unclear. We estimate our power using methods described by Hedges and Pigott (2004). Based on the within-study variability and number of studies, we estimate that we are sufficiently able (power > .90) to detect heterogeneity of at least a standard deviation (in odds ratios) of .06 (corresponding r of .10). We note that the majority of psychological meta-analyses find between study variability between tau of 0 and .25 (Van Erp, Verhagen, Grasman, & Wagenmakers, 2017). See our online page, Power Analysis (https://alsaging.github.io/healthyn/chronic_power_analysis.html), for more information about our power analysis.

**Preregistered analyses**

This study was preregistered on OSF (osf.io/m7aen). While data had been collected prior to analysis, the current study constitutes a preregistration because the analytic decisions were made by the first author prior to examination of most of the data. Moreover, the first author only analyzed a subset of one of the fifteen data sets (i.e., the HRS) for the explicit purpose of ensuring the scripts would run smoothly prior to registering the analytic plan and sharing analytic scripts with the other data analysts. The process for analyzing the data was as follows:

The first author decided which chronic conditions to use as outcomes and which variables to use as covariates. The first author wrote code to evaluate the inferential models in R. This R script created an output object containing meta-data, descriptive statistics, statistics from the inferential models, and values predicted from the model. The output object could not contain raw data, as many of the datasets used here are not available in the public domain. As a consequence, we are unable to post the raw data, but we have posted the output objects of each study. To test and refine this script, the first author, who had analyzed the HRS data in prior publications, then cleaned the HRS data and created multiple pseudo-versions of the HRS by randomly sampling rows with replacement. The inferential analysis script was run on each of these pseudo-HRS samples to create pseudo-output objects. Finally, “meta-analysis” scripts were written to extract relevant information from each of the output objects and evaluate the overall effect from the individual analyses. These scripts were tested on the pseudo-output generated when testing the inferential models. At this point, the study was pre-registered and included a template inferential analysis script to be adapted for each individual data set and several meta-analysis scripts that could evaluate the output of the models on the individual data sets. After pre-registration, individual data analysts downloaded the R script and ran this on their respective data set. Output objects were created and uploaded to OSF; the first author then used the meta-analysis scripts on these output objects.

**Deviations**

It was unknown at the time of preregistration that many studies did not have a measure of self-rated health, which was a preregistered covariate. Therefore, the analyses presented here do not include self-rated health as a covariate. However, for data sets that did include this variable, analyses that included self-rated health were also conducted, and the results from those analyses are available at osf.io/48fhe. Additional exploratory analyses, decided upon during the data analysis phase, were models that examined the three-way interaction of neuroticism, conscientiousness and age. These are mentioned briefly in the Results section and are labelled as exploratory. All exploratory analyses controlled for age, sex, education, the other personality traits, BMI, and whether the participant had been diagnosed with another chronic condition. Finally, although we registered that we would examine personality measures as a moderator of results, we did not specify how those scales would be coded. We were unable to preregister this result because at the time of pre-registration, the data analyst did not have the information necessary to decide how to code the scales. Full information regarding the participant sample, recruitment and survey procedures, and the measures used is available in Supplementary File 1 (osf.io/wwfjr), and the data necessary for testing the meta-study models are available on OSF (osf.io/48fhe); we invite, interested readers to recode these data in whatever way they believe is appropriate and test those moderation models for themselves.

**Results**

**Main effects**

Consistent with prior research, the meta-analytic effects showed that greater odds of hypertension were concurrently associated with higher levels of neuroticism \((OR = 1.07, 95\% CI = [1.02, 1.13])\) and lower levels of conscientiousness \((OR = 0.93, 95\% CI = [0.89, 0.98])\). The hypertension-neuroticism effect was only significant in three studies: the HRS \((OR = 1.05, 95\% CI = [1.02, 1.08])\), the LBLS \((OR = 1.20, 95\% CI = [1.05, 1.38])\), and the WLS \((OR = 1.28, 95\% CI = [1.22, 1.34])\). The hypertension-conscientiousness link was only significant in the ELSA \((OR = 0.81, 95\% CI = [0.76, 0.85])\), the HRS \((OR=0.83, 95\% CI = [0.81, 0.86])\), and the MIDUS \((OR = 0.92, 95\% CI = [0.86, 1.00])\).

The meta-analytic effects indicated that neuroticism was prospectively associated with hypertension \((OR = 1.05, 95\% CI = [1.02, 1.08])\), but conscientiousness was not \((OR = 0.97, 95\% CI = [0.91, 1.02])\). In the individual study analyses, the prospective hypertension-neuroticism association was only significant in the HRS \((OR = 1.06, 95\% CI = [1.00, 1.11])\) and the WLS \((OR = 1.06, 95\% CI = [1.01, 1.10])\).
The prospective hypertension-conscientiousness association was only significant in the ELSA (OR = 0.85, 95% CI = [0.76, 0.94]) and the HRS (OR = 0.90, 95% CI = [0.85, 0.94]). Diabetes was concurrently more likely to be seen with higher levels of neuroticism (OR = 1.03, 95% CI = [1.00, 1.06]) and lower levels of conscientiousness (OR = 0.85, 95% CI = [0.80, 0.91]). The diabetes-neuroticism relationship was only significant in the HRS (OR = 1.04, 95% CI = [1.00, 1.08]). The diabetes-conscientiousness link was significant in four studies: the ELSA (OR = 0.78, 95% CI = [0.71, 0.85]), the HRS (OR = 0.78, 95% CI = [0.75, 0.81]), the NAS (OR = 0.79, 95% CI = [0.63, 0.99]), and the ROS (OR = 0.84, 95% CI = [0.71, 0.99]).

Meta-analytically, conscientiousness was prospectively associated with the development of diabetes (OR = 0.88, 95% CI = [0.84, 0.93]), but the meta-analytic relationship between neuroticism and later development of diabetes was nonsignificant (OR = 1.01, 95% CI = [0.97, 1.05]). The prospective diabetes-conscientiousness link was significant in four studies: the ELSA (OR = 0.74, 95% CI = [0.64, 0.85]), the HRS (OR = 0.93, 95% CI = [0.87, 0.98]), the MIDUS (OR = 0.85, 95% CI = [0.77, 0.94]), and the WLS (OR = 0.89, 95% CI = [0.83, 0.95]). The prospective diabetes-neuroticism link was not significant in any of the studies.

Heart disease is concurrently more likely to be seen at lower levels of conscientiousness (OR = 0.88, 95% CI = [0.83, 0.94]) and higher levels of neuroticism (OR = 1.08, 95% CI = [1.03, 1.12]). The heart disease-neuroticism relationship was significantly positive in the BASE-II (OR = 1.22, 95% CI = [1.03, 1.45]), the HRS (OR = 1.06, 95% CI = [1.03, 1.10]), and the WLS (OR = 1.15, 95% CI = [1.06, 1.25]). Heart disease and conscientiousness were significantly negatively associated in many of the studies: the EAS (OR = 0.80, 95% CI = [0.67, 0.94]), the ELSA (OR = 0.77, 95% CI = [0.73, 0.83]), the HRS (OR = 0.83, 95% CI = [0.80, 0.85]), the ILSE (OR = 0.79, 95% CI = [0.65, 0.96]), the MAS (OR = 0.84, 95% CI = [0.73, 0.97]), the MIDUS (OR = 0.89, 95% CI = [0.82, 0.97]) and the NAS (OR = 0.72, 95% CI = [0.60, 0.86]). Notably, conscientiousness was associated with greater rates of heart disease in the BASE-II (OR = 1.22, 95% CI = [1.02, 1.46]).

Conscientiousness was prospectively associated with developing heart disease (b = −0.05, OR = 0.95, 95% CI = [0.92, 0.99], p = 0.014) but neuroticism was not (OR = 1.04, 95% CI = [0.99, 1.10]). The prospective neuroticism-heart disease link was significant for the BASE-II (OR = 1.35, 95% CI = [1.08, 1.70]) and the HRS (OR = 1.08, 95% CI = [1.03, 1.14]). The prospective relationship between conscientiousness and heart disease was only significant in the HRS (OR = 0.91, 95% CI = [0.87, 0.96]). The full results for these main effects can be seen at osf.io/7zmyg.

### N × C Interactions and Prospective Condition Development

Binary logistic regression models were fit within each study to estimate whether the interaction of neuroticism and conscientiousness was prospectively linked with the development of hypertension, diabetes or heart disease. For these analyses, we used the assessment of chronic condition at the last wave of data provided by each participant; a consequence is that time since baseline varies across participants as well as across datasets. Participants were excluded from these analyses if they had been diagnosed with the chronic condition at baseline or earlier. Based on the weighted average effect size, the interaction was not significantly associated with the development of hypertension (OR = 0.98, 95% CI = [0.95, 1.01]), diabetes (OR = 0.98, 95% CI = [0.93, 1.02]), or heart disease (OR = 0.98, 95% CI = [0.94, 1.03]). Moreover, only for one
outcome in one study was the interaction coefficient significant (when estimating incidence of heart disease in the ELSA, $OR = 0.88$, 95% CI = [0.80, 0.96]). See Figure 3 for the interaction coefficient of each study, the weighted average affects and the simple slopes of neuroticism.

**Exploratory analysis: $N \times C \times x$**

After viewing the results of the coordinated analysis, it was suggested by some that there could be a three-way interaction of neuroticism by conscientiousness by age. Despite having not pre-registered these analyses, we do believe they could be informative, as both health and personality change systematically with age (e.g., Wagner, Ram, Smith, & Gerstorf, 2016; Chopik & Kitayama, 2018; Letzring, Edmonds, & Hampson, 2014). We chose to explore these exploratory analyses, and we present the results without significance tests. Weighted interaction effects were small and confidence intervals contained an odds ratio of 1 (hypertension: $OR = 0.99$, 95% CI = [0.95, 1.03]; diabetes: $OR = 1.04$, 95% CI = [0.97, 1.10]; heart
disease: \( OR = 1.00, 95\% \text{ CI} = [0.95, 1.06] \), indicating that the interaction of neuroticism and conscientiousness was not more strongly related to chronic condition status in older participants. This pattern was largely repeated in individual studies. Effect sizes were large and confidence intervals excluded 1 on three occasions: with hypertension in the OATS \( OR = 7.90, 95\% \text{ CI} = [1.10, 56.95] \), with diabetes in the WLS \( OR = 1.11, 95\% \text{ CI} = [1.00, 1.22] \), and with heart disease in the ILSE \( OR = 63.04, 95\% \text{ CI} = [1.28, 3104.32] \). Given that (1) we did not preregister these analyses, (2) the number of statistical tests across all outcomes and all datasets is large, and (3) these confidence intervals either contain the null \( OR = 1.00 \) or have very wide upper boundaries, we do not have faith in the validity of these “significant” results. Therefore, we interpret these findings as more likely noise rather than signal.

**Discussion**

Our coordinated analysis did not find consistent evidence that the relationship between neuroticism and the chronic conditions of hypertension, diabetes, and heart disease varies across levels of conscientiousness. These null associations contribute to a broader theory of healthy neuroticism (Friedman, 2000), which posits that neuroticism may be beneficial for some individuals or under some conditions. Specifically, the current study provides no evidence that high levels of neuroticism may be healthy for individuals who also have high levels of conscientiousness, nor does it find that the negative effects of neuroticism on health are buffered by high levels of conscientiousness. These null results counter previous findings that the neuroticism by conscientiousness interaction was associated with better health (Turiano et al., 2013, 2012; Vollrath & Torgersen, 2002; Weston & Jackson, 2015).

The current study is the second of three studies rigorously testing the healthy neuroticism hypothesis. The conclusions of the current study mirror that of the third study, which found that the relationship between neuroticism and mortality does not differ across levels of conscientiousness. However, the first study found that higher neuroticism was less strongly associated with greater rates of smoking and lower rates of physical activity at higher levels of conscientiousness. Together, the three studies in this coordinated project...
suggest that while it may be true that so-called healthy neurotics engage in slightly better health behaviors, this effect is not substantial enough to impact overall health. This pattern of associations could be explained in several ways. For example, conscientiousness may curb the behavioral tendencies of individuals high in neuroticism but do little to reduce the physiological stress experienced by those high in trait anxiety. (Barlow, 2000). Alternatively, interactions between these traits may only be associated with health for some populations, for example, already-ill samples (as suggested by Weston & Jackson, 2015) that are less likely to participate in longitudinal observational studies. Further, the effects on health behaviors may not be large enough to translate into actual differences in health outcomes.

Figure 3: Forest plots for longitudinal analyses. Code to produce figure available at ialsaging.github.io/healthyN.
The current study (along with the others in this series) implies that personality and health researchers interested in the role of neuroticism should carefully reconsider the potential and limits of conscientiousness to act as a moderator. The current set of studies provides large and somewhat representative tests of this interaction and ultimately does not provide evidence for “healthy neuroticism”. Given these findings (and the likely large number of unpublished null results), we suspect this interaction cannot explain the discrepancies between mortality studies that first inspired the conception of healthy neuroticism (Friedman, 2000). Personality and health researchers may look to other individual differences, such as socioeconomic status (e.g., Hagger-Johnson et al., 2012). We may also look to situations or contexts in which neuroticism may be beneficial, such as after someone has experienced a substantial health threat (Gale et al., 2017; Weston & Jackson, 2015). Consideration must also be given to recent large-scale examinations of neuroticism and mortality, which suggest that neuroticism may not be as strongly linked with health as it was previously believed to be (Graham et al., 2017; Jokela et al., 2013). In other words, perhaps we mistook error for heterogeneity. If the true relationship between neuroticism and health is null, then we should expect to find samples with “positive” effects of neuroticism and other samples with “negative” effects. Instead of searching for a moderator, the simpler explanation may be that no relationship exists.

**Constraints on generality**

The set of studies included in this coordinated analysis, while considerable in size, is not comprehensive. The included studies are largely WEIRD samples (i.e., Westernized, Educated, Industrialized, Rich, and Democratic; Henrich, Heine, & Norenzayan, 2010). Our results may not generalize to Eastern cultures or less-industrialized countries, and disadvantaged populations within the cultures studied here. Results are also likely limited to the chronic conditions that we studied and may not generalize to cognitive (e.g., Alzheimer’s) or non-behavioral conditions (e.g., lymphoma). We believe these results would generalize to measures of personality that were not used in the present study. There was little evidence that the present scales yielded different conclusions from each other, and we have no reason to believe these scales vary in a systematic way from other scales. We also believe that these results are likely to generalize over time, as both the personality and health measurements varied across time within this project; that is, these results would be expected in samples collected in the past and also the foreseeable future.

The current study, along with the others in this series, estimated population-level relationships between constructs through the coordinated analysis approach. One benefit of coordinated analysis is the ability to compare methods of data collection or cohorts to find the boundary conditions of an effect (Hofer & Piccinin, 2010). However, we found limited evidence of heterogeneity of effects across data sets. The observed null effects of neuroticism and conscientiousness on health were replicated across studies in different cultures, with different measures of personality, and across different time spans. This underscores our conclusion that the relationship of neuroticism to concurrent and prospective health conditions is unrelated to levels of conscientiousness.

The current study was limited by the data available. Nearly all measures were self-reported. Ideally, future research will utilize medical diagnoses and physician reports of health or biological indicators of health. Finally, the current study used binary coding to harmonize measures of health status across studies. This choice was made to compare results across countries, time between assessments, and measures of personality; however, we are unable to distinguish between different severities of diagnosis within a condition (e.g., a heart murmur versus a heart attack, or Type I versus Type II diabetes). It is possible that healthy neuroticism may not explain differences in having a chronic condition, but rather in the severity of a chronic condition, and we would be unable to see this effect in the current analysis. Finally, the prospective analyses omitted participants who were diagnosed at baseline, with the explicit purpose of assessing incidence of chronic conditions. Selection effects may have biased the results if those who became diagnosed or those with specific personality profiles dropped out of the study in systematic ways. Model generalizability is constrained by missing data in these models. For any longitudinal study, attrition causes bias in the results. Here, it may be the case that participants who become ill between waves may be missing, thus selecting out the most extremely ill cases. Additionally, certain personality characteristics (notably, agreeableness and openness; Salthouse, 2013) are associated with repeated participation in longitudinal studies, so we may lose participants low in these traits from the sample over time.

**Conclusion**

There is no substantial evidence that healthy neuroticism (as defined by conscientiousness-as-moderator) is associated with chronic condition status. The profile of high-neuroticism and high-conscientiousness may be associated with healthier behaviors, but this does not translate into better overall health, based on the chronic conditions considered here. If we continue to believe that neuroticism may, in some contexts or for some persons, be associated with better health, we should refrain from further testing conscientiousness as the determining factor in this relationship and devote our resources to other theoretical constructs.

**Data Accessibility Statement**

R scripts and meta-data (e.g., coefficients estimated from individual datasets) are available on osf.io/48fhe. Raw data can be obtained through the following websites or emailing the following researchers: – BASE-II: Swantje Müller (swantje.mueller@hu-berlin.de) – EAS: Ruixue Zhaoyang ([rz12@psu.edu][rz12@psu.edu]) – ELSA: https://www.elsa-project.ac.uk/ – HRS: http://hrsonline.
isr.umich.edu/ — ILSE: Damaris Aschwanden ((Damaris.Aschwanden@med.fsu.edu)|Damaris.Aschwanden@med.fsu.edu)) – LBC-1936: Tom Booth ((tbooth1@exseed.ed.ac.uk)|tbooth1@exseed.ed.ac.uk)) – LBLS: Kristi M. Wisniewski ((kwisniew@usc.edu)|kwisniew@usc.edu)) – MAP: Bryan D. James ((Bryan_James@rush.edu)|Bryan_James@rush.edu) – MAS: Fleur Harrison ((f.harrison@unsu.edu.au)|f.harrison@unsu.edu.au) – MIDUS: http://midus.colectica.org/ – NAS: Avron Spiro ([aspiro3@bu.edu]|[aspiro3@bu.edu]) – OATS: Steven R. Makkar ((stevem@unsw.edu.au)|stevem@unsw.edu.au) – ROS: Bryan D. James ((Bryan_James@rush.edu)|Bryan_James@rush.edu) – SLS: Swantje Müller (swantje.mueller@hu-berlin.de) – WLS: https://www.ssc.wisc.edu/wlsresearch/.

Acknowledgements

We would like to thank Alexander Weiss, for his comments on early drafts of this manuscript, and Brenton Wiernik, who consulted on and wrote the R functions to perform power analyses.

Funding Information

This project was made possible by funding from the NIA (P01AG043362, R01-AG1843, R01-AG067622, R01-AG064006, K01AG050823, R01AG017917, P30AG010161, R01AG055653, R01AG056486, R01AG10569, T32AG000307, R01-AG32037, R29-AG07465, and R01-AG002287, R01AG055653, R01AG056486); Dietmar Hopp Stiftung; German Federal Ministry for Family, Senior Citizens, Women, and Youth (Grant Ref. 314-1722/102/16 and 301-1720-295/2); National Health and Medical Research Council (ID: 568969); NHMRC Australia (APP1100579); German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) under grant numbers #16SV5536K, #16SV5537, #16SV5538, #01UW0808; #01GL1716A, #01GL1716B; Age UK; National Health and Medical Research Council (MR/K026992/1).

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Substantial contributions to conception and design: SJW, EKG, NAT, DA, TB, FH, BDJ, NAL, SRM, SM, KMW, TY, RZ, AS, GMT, AMP, DKM. Acquisition of data: SJW, EKG, NAT, DA, TB, FH, BDJ, NAL, SRM, SM, KMW, TY, RZ, AS, JD, GGW, EST, ID, SW, KWS, MS, RAL, MK, IJD, EMZ, DAB, PS, HB, JNT, DA, MJW, DG, MA, SMH. Analysis and interpretation of data: SJW, EKG, NAT, DA, TB, FH, BDJ, NAL, SRM, SM, KMW, TY, RZ, AS, JD, GGW, EST, ID, SW, KWS, MS, RAL, MK, IJD, EMZ, DAB, PS, HB, JNT, DA, MJW, DG, MA, GMT, AMP, SMH, DKM. Final approval of the version to be published: SJW, EKG, NAT, DA, TB, FH, BDJ, NAL, SRM, SM, KMW, TY, RZ, AS, JD, GGW, EST, ID, SW, KWS, MS, RAL, MK, IJD, EMZ, DAB, PS, HB, JNT, DA, MJW, DG, MA, GMT, AMP, SMH, DKM.

References


