CURRENT ISSUES IN PERSONALITY PSYCHOLOGY · VOLUME 9(2), 2021
DOI: HTTPS://DOI.ORG/10.5114/CIPP.2021.103050

ORIGINAL ARTICLE

Big Five personality as a predictor of health: shortening the questionnaire through the elastic net

Brian M. Doornenbal

1: Salut, Arnhem, The Netherlands
2: Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

BACKGROUND
The Big Five personality attributes (i.e. openness, conscientiousness, extraversion, agreeableness, and neuroticism) help to predict health. To predict health, researchers may prefer to use a short version of the Big Five Inventory. Although the psychometric properties of the shortened scales can be highly satisfactory, their use can lead researchers to substantially underestimate the role of personality. The aim of this paper is to demonstrate a method appropriate for shortening the Big Five Inventory without losing predictive performance.

PARTICIPANTS AND PROCEDURE
The sample comprised 4,678 panel members. The personality traits were measured in 2017 using the Five Factor Model International Personality Item Pool and subjective health was measured in 2018 using the item "How would you describe your health, generally speaking?" While studying the personality-health relationship, the elastic net was compared to a more conventional regression method.

RESULTS
While predicting health based on personality, using 14 Big Five Inventory items (R^2 = .19) resulted in a similar predictive performance as using 50 Big Five Inventory items (R^2 = .18). Controlled for gender and age, participants experienced lower levels of health when they "often feel blue", are not "relaxed most of the time", and "worry about things." These aspects of neuroticism relate to the lower-order facets anxiety and depression.

CONCLUSIONS
When the primary goal of personality assessment is predictive performance, researchers should consider shortening their questionnaire using the method demonstrated in this paper. Shortening of the questionnaire does not have to result in a lower predictive performance.

KEY WORDS
health; elastic net; machine learning; personality; neuroticism

CORRESPONDING AUTHOR – Brian M. Doornenbal, Ph.D., Salut, Jansbuitensingel 7, 6811 AA Arnhem, The Netherlands, e-mail: brian@gezondmetsalut.nl

AUTHORS' CONTRIBUTION – A: Study design · B: Data collection · C: Statistical analysis · D: Data interpretation · E: Manuscript preparation · F: Literature search · G: Funds collection

TO CITE THIS ARTICLE – Doornenbal, B. M. (2021). Big Five personality as a predictor of health: shortening the questionnaire through the elastic net. *Current Issues in Personality Psychology, 9*(2), 159–164.

RECEIVED 20.10.2020 · REVIEWED 22.11.2020 · ACCEPTED 07.12.2020 · PUBLISHED 27.01.2021

BACKGROUND

Decades of research consistently indicate that the Big Five personality traits are critical to predicting health (Strickhouser et al., 2017). Typically, individuals experience a higher level of subjective health when they score lower on neuroticism and higher on conscientiousness and agreeableness (Strickhouser et al., 2017). Although personality can help to estimate how healthy a population is and will be, researchers often have limited time with their respondents. Fortunately, various short versions of personality trait questionnaires are available (e.g., Bernard et al., 2005; Czerwiński & Atroszko, 2020; Rammstedt & John, 2007; Woods & Hampson, 2005).

The shortened personality scales are highly satisfactory in terms of some psychometric properties. For example, shorter-scale outcomes relate strongly to well-established longer-scale outcomes (i.e. high convergent validity; Wood et al., 2010; Woods & Hampson, 2005) and the test results are relatively stable over time (i.e. high test-retest reliability; Gosling et al., 2003). However, the use of shorter scales can lead researchers to underestimate the role of personality traits (Credé et al., 2012). The hierarchical structure of the Big Five model is a main reason that shorter scales can lead to underestimation (Credé et al., 2012). The Big Five traits are thought to have a hierarchical structure with each trait comprising lower-order facets (Roberts et al., 2004; Soto & John, 2017). For example, neuroticism is thought to comprise the facets *anxiety*, *depression*, and *emotional volatility*. The lower-order facets add value to the predictive performance of personality assessments (Credé et al., 2012). Because smaller sets of items (e.g. 5 or 10) are less capable of capturing all lower-order facets, shorter scales may miss crucial information to predict outcomes.

In this paper, I demonstrate a method that can help balance the needs for predictive performance and short administration time. When limited time is available for assessing personality traits and predictive performance is key, researchers should consider using the regularized regression method the elastic net (Zou & Hastie, 2005). The elastic net extends ordinary least square (OLS) regression by imposing a penalty on highly correlated predictor variables. The penalty constrains the size of the regression coefficients that are empirically less important for predicting the studied outcome. When the penalty of the regularization is set such that coefficients can "shrink" to zero, the elastic net results in a subset of predictors. The elastic net can thus help to reduce the amount of personality trait items needed for prediction.

The elastic net is widely used throughout the literature, among others in the fields of genetics (Barretina et al., 2012), biology (Sunagawa et al., 2015), and econometrics (Bai & Ng, 2008). The elastic net combines two shrinkage methods: RIDGE and LASSO (Hastie et al., 2017; Zou & Hastie, 2005). RIDGE regression shrinks the coefficients of related predictors to each other by averaging them. Empirically unimportant predictors are penalized toward (but not including) zero. Hence, RIDGE does not result in a selection of predictors. LASSO regression shrinks the coefficients of empirically unimportant predictors to zero, which results in a selection of predictors. However, a major disadvantage of LASSO is that it randomly selects predictors among highly correlated predictors, leading to less generalizable findings (Hastie et al., 2017; Zou & Hastie, 2005). The elastic net combines the two shrinkage methods of RIDGE and LASSO, by averaging the coefficient of related variables and shrinking some predictors to zero. The elastic net often outperforms the LASSO, while still being able to create sparsity by selecting predictors (Hastie et al., 2017; Zou & Hastie, 2005).

Simply put, the elastic net can help to select features (e.g. questionnaire items) that are most important in predicting an outcome measure. Zou and Hastie, who introduced the elastic net, compare the technique with a stretchable fishing net that retains "all the big fish" (2005, p. 302). Although the added value of the technique is widely proven for selecting non-questionnaire predictors (Bai & Ng, 2008; Barretina et al., 2012; Sunagawa et al., 2015), the elastic net is also suitable for selecting questionnaire items that are most important for predicting outcome measures such as subjective health.

PARTICIPANTS AND PROCEDURE

PARTICIPANTS

To demonstrate the added value of the elastic net for personality research, this paper uses data of the LISS (Longitudinal Internet Studies for the Social sciences) panel, administered by CentERdata (Tilburg University, the Netherlands). The LISS panel is a representative sample of Dutch individuals (Scherpenzeel & Das, 2010). The sample used in this study consists of 4,678 participants, 53% of whom were women, and the age ranged from 18 to 101 (M = 54.11, SD = 17.46).

PROCEDURE

The personality data (n = 6,010) were collected in 2017 and health data (n = 5,455) were collected in 2018. Collecting predictor variables and outcome variables through different surveys helps to prevent common method bias (Lindell & Whitney, 2001). Participants were asked to read and agree to the LISS informed

consent (see www.lissdata.nl/faq-page#n5512 for information about the ethical approval).

MEASURES

Big Five personality traits. The Big Five personality traits were measured with the 50-item Five Factor Model International Personality Item Pool (IPIP; Goldberg, 1999). The test was administered in Dutch, after being translated by professional translators. The respondents' answers were registered on a 7-point Likert scale ranging from 1 (*totally disagree*) to 7 (*totally agree*). The personality test showed α coefficients between .77 and .89: openness .77; conscientiousness .77; extraversion .88; agreeableness .82; and neuroticism .89.

Subjective health. Subjective health was measured using the item "How would you describe your health, generally speaking?" The respondents' answers were registered on a 7-point Likert scale ranging from 1 (*poor*) to 7 (*excellent*).

ANALYTICAL PROCEDURE

Health was predicted by two models: an ordinary least squares regression model and an elastic net model. The models were fitted based on a training dataset, comprising 3,120 observations (2/3 of the data). Subsequently, the predictive performance of the models was tested on a test dataset, the remaining 1,558 observations (1/3 of the data). Splitting data into a training and test set is necessary to help avoid *overfit* – that is, when a model performs well because it memorizes the data. Repeated 10-fold cross-validation was used to fit the models. Here, the models go through multiple training and test iterations. The training set was randomly split in ten approximately equal sub-samples. After splitting, nine sub-samples were collectively used as the training dataset and the remaining sub-sample was used as the test dataset. Each of the 10 sub-samples was used as a held back test. The cross-validation procedure was repeated 10 times. The model that performed best was used to test the predictive performance.

The ordinary least squares model regressed subjective health to the predictors openness, conscientiousness, extraversion, agreeableness, neuroticism, gender, and age. The elastic net regressed subjective health to the 50 personality traits items and the control variables. Both models were fitted on normalized data (the normalization took place after splitting the data). The hyperparameters of the elastic net (i.e. α and λ) were optimized through random search, which is more efficient than grid search and manual search (Bergstra & Bengio, 2012). During the random search, 100 different combinations of α and λ were explored.

RESULTS

Table 1 presents the correlations among the variables. Two control variables were added to the models: gender and age. The results of the ordinary least squares regression model reported in Table 2 support the relationship between the Big Five personality traits and subjective health (R^2 = .17). The model applied to the test set (i.e., the out-of-sample prediction) yields a similar relationship (R^2 = .18). The elastic net model also supports the relationship between the Big Five personality traits and subjective health both while fitting the model (R^2 = .18) and while testing the model (R^2 = .19). Thus, the predictive performance of the elastic net model is not inferior to the performance of the ordinary least square regression model.

As reported in Table 2, the elastic net selects 15 predictors (α = .19, λ = 0.087), of which 14 are personality trait items. The five most important predictors are, in descending order of importance: age (–0.155), neuroticism item "Often feel blue" (–0.093), neuroticism item "Am relaxed most of the time" (0.065), neuroticism item "Worry about things" (–0.039), extraversion item "Feel comfortable around people" (0.025), and openness item "Have difficulty understanding abstract ideas" (–0.019).

DISCUSSION

This paper demonstrates that shortening the questionnaire through the elastic net does not have to result in a lower predictive performance. The 15 items selected through the elastic net did not perform worse in predicting subjective health compared to the 52-item Big Five Inventory (controlling for gender and age). Therefore, researchers should consider the use of the elastic net for the shortening of personality trait questionnaires.

In addition to reducing administration time, using the elastic net can help to overcome a main disadvantage of short versions of personality trait questionnaires. The Big Five traits are argued to have a hierarchical structure in which traits comprise lower-order facets (Roberts et al., 2004; Soto & John, 2017). Short versions of personality trait questionnaires have smaller sets of items. Smaller sets of items are less capable of capturing the lower-order facets, which may cause a lack of information crucial for prediction. The elastic net selects predictors that empirically are important, even when they are related. Therefore, applying the elastic net might result in the selection of multiple lower-order facets of one personality trait. In this paper, for example, the elastic model indicated that items belonging to lower-order facets of neuroticism, anxiety ("relaxed most of the time" and "worry about things") and depression ("often feel blue"), are most important for the prediction of subjective

Table 1

Means, standard deviations, and correlations among the variables

Variables	*M*	*SD*	1	2	3	4	5	6	7
1. Woman	0.53								
2. Age	54.00	17.00	−0.07***						
3. Openness	3.50	0.51	−0.08***	−0.14***					
4. Conscientiousness	3.77	0.52	0.07***	0.14***	0.24***				
5. Extraversion	3.24	0.67	0.00	−0.01	0.34***	0.15***			
6. Agreeableness	3.89	0.52	0.30***	0.07***	0.26***	0.30***	0.33***		
7. Neuroticism	2.52	0.70	0.18***	−0.17***	−0.19***	−0.25***	−0.24***	−0.06***	
8. Subjective health	3.10	0.78	−0.05***	−0.25***	0.14***	0.08***	0.14***	0.02	−0.28***

Note. $N = 4{,}678$; woman (53%) used as dummy variable for gender; age is reported in years; $^{***}p < .001$.

Table 2

The ordinary least squares regression model

Predictor	β	*p*
Intercept	3.102***	< .001
Age	−0.234***	< .001
Woman	−0.009	.531
Openness	0.027	.060
Conscientiousness	0.021	.138
Extraversion	0.034*	.018
Agreeableness	−0.004	.811
Neuroticism	−0.233***	< .001
$R^2 = .17$		

Note. $N = 3{,}120$; subjective health as target variable; R^2 value is unadjusted; woman used as dummy variable for gender; $^{*}p < .05$, $^{***}p < .001$.

health. Such data-driven insights can be valuable for subsequent theory refinement or new theory building. Previously, researchers argued that anxiety and depression are factors affecting salutogenesis (Schnyder et al., 2000) – that is, a process of moving towards the health end of a health-ease/disease continuum. Possibly, anxiety and depression obstruct people in their movement towards feeling healthy.

Fifteen predictor variables were selected in this study, but the elastic net can be forced to select fewer predictors. During the analysis, the elastic net hyperparameters were chosen based on a random search. Random search is more efficient than grid search and manual search (Bergstra & Bengio, 2012). However, when researchers want to force the elastic net to select fewer predictors, the α value could be set more towards one. When $\alpha = 0$, the elastic net is the same as RIDGE (the coefficients of correlated predictors are similarly shrunk towards zero), when $\alpha = 1$, the elastic net is the same as LASSO (the coefficient of one selected predictor is larger, whilst the others are shrunk to zero). In this study α was set to .19, which resulted in the selection of 15 predictor variables.

The elastic net uses a data-driven approach to select predictors, which warrants important considerations. First, the selected predictors are empirically important in the training data, but might be less important in different datasets. Validating the predictive performance across different datasets is thus recommended. In this paper, repeated cross-validation was used to estimate the predictive performance. Second, because different items might be selected across different samples, the outcomes of the shorter versions are likely to be less comparable across studies. The Big Five Inventory is consistent in the items used.

Table 3

The elastic net model

Predictor	Coefficient	Item
Intercept	3.102	
Age	−0.155	
Neuroticism 1	−0.093	Often feel blue
Neuroticism 2	0.065	Am relaxed most of the time
Neuroticism 3	−0.039	Worry about things
Extraversion 1	0.025	Feel comfortable around people
Openness 1	−0.019	Have difficulty understanding abstract ideas
Neuroticism 4	−0.015	Get upset easily
Neuroticism 5	−0.014	Seldom feel blue
Neuroticism 6	−0.012	Have frequent mood swings
Neuroticism 7	−0.008	Change my mood a lot
Openness 2	−0.005	Do not have a good imagination
Openness 3	0.002	Have a rich vocabulary
Openness 4	0.002	Am quick to understand things
Openness 5	−0.002	Am not interested in abstract ideas
Extraversion 2	0.001	Am the life of the party
R^2 = .18		

Note. N = 3,120; α =.19, λ = 0.087; subjective health as target variable; R^2 value is unadjusted.

This not only allows for better comparison across studies, focusing on personality differences, but also for estimating characteristics that (theoretically) relate to personality traits, such as physical health and health behaviours (Strickhouser et al., 2017). To ensure comparability, practitioners could opt to use a short version of the Big Five Inventory next to the items selected by the elastic net.

LIMITATIONS

Although this study was conducted to demonstrate the added value of the elastic net for personality research, it also yields insights into the relationship between personality and subjective health. While interpreting these insights, it should be considered that the Big Five traits had a somewhat different impact on subjective health than found in previous studies. Meta-analyses revealed higher levels of subjective health among individuals who scored lower on neuroticism and higher on conscientiousness and agreeableness (Strickhouser et al., 2017). In this study (see Table 2), neuroticism was found to have a negative influence on subjective health, but conscientiousness and agreeableness did not have a significant bearing. The impact of openness and extraversion was similar to that found in previous studies. These differences suggest that the elastic net may select different items in a different sample. The current sample is representative of the general population of the Netherlands. Possibly, the items selected by the elastic net in this study are of particular importance to the subjective health of Dutch citizens. Further studies are needed to examine the stability of the results of this study. Applying the elastic net across contexts will reveal which personality items are important across contexts.

CONCLUSIONS

When predictive performance is the primary goal of personality assessment, researchers should consider using the elastic net to shorten their questionnaire. As demonstrated in this paper, shortening of the questionnaire does not have to result in a lower predictive performance. The 15-item elastic net model did not perform worse than the 52-item ordinary least squares regression model.

References

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*, 304–317. https://doi.org/10.1016/j.jeconom.2008.08.010

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M. F., Monahan, J. E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., … Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modeling of anticancer drug sensitivity. *Nature, 483*, 603–607. https://doi.org/10.1038/nature11003

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, *13*, 281–305.

Bernard, L. C., Walsh, R. P., & Mills, M. (2005). Ask once, may tell: Comparative validity of single and multiple item measurement of the Big Five personality factors. *Counseling & Clinical Psychology Journal, 2*, 40–57. https://doi.org/10.1016/j.jengtecman.2006.06.003

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology, 102*, 874–888. https://doi.org/10.1037/a0027403

Czerwiński, S., & Atroszko, P. (2020). Scores of short and free scale for Big Five explain perceived stress at different stages of life: Validity, reliability and measurement invariance of the Polish adaptation of Mini-IPIP. *Current Issues in Personality Psychology, 8*, 73–82. https://doi.org/10.5114/cipp.2020.95149

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg University Press.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big Five personality domains. *Journal of Research in Personality, 37*, 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Hastie, T., Tibshirani, R., & Friedman, J. (2017). Shrinkage methods. In *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed., pp. 61–79). Springer.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology, 86*, 114–121. https://doi.org/10.1037/0021-9010.86.1.114

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*, 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Roberts, B. W., Bogg, T., Walton, K. E., Chernyshenko, O. S., & Stark, S. E. (2004). A lexical investigation of the lower-order structure of conscientiousness. *Journal of Research in Personality, 38*, 164–178. https://doi.org/10.1016/S0092-6566(03)00065-5

Scherpenzeel, A. C., & Das, M. (2010). "True" longitudinal and probability-based internet panels: Evidence from the Netherlands. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet: Advances in applied methods and research strategies* (pp. 77–104). Taylor & Francis.

Schnyder, U., Büchi, S., Sensky, T., & Klaghofer, R. (2000). Antonovsky's sense of coherence: Trait or state? *Psychotherapy and Psychosomatics, 69*, 296–302. https://doi.org/10.1159/000012411

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*, 117–143. https://doi.org/10.1037/pspp0000096

Strickhouser, J. E., Zell, E., & Krizan, Z. (2017). Does personality predict health and well-being? A metasynthesis. *Health Psychology, 36*, 797–810. https://doi.org/10.1037/hea0000475

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., … Bork, P. (2015). Structure and function of the global ocean microbiome. *Science, 348*, 1261359. https://doi.org/10.1126/science.1261359

Wood, D., Nye, C. D., & Saucier, G. (2010). Identification and measurement of a more comprehensive set of person-descriptive trait markers from the English lexicon. *Journal of Research in Personality, 44*, 258–272. https://doi.org/10.1016/j.jrp.2010.02.003

Woods, S. A., & Hampson, S. E. (2005). Measuring the Big Five with single items using a bipolar response scale. *European Journal of Personality, 19*, 373–390. https://doi.org/10.1002/per.542

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B, 67*, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x