UNIVERSITY OF CALIFORNIA
RIVERSIDE

Exploring Predictors of Subjective Well-Being Using Machine Learning and Propensity Score Techniques

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Seth Michael Margolis

June 2020

Dissertation Committee:
Dr. Sonja Lyubomirsky, Chairperson
Dr. Daniel J. Ozer
Dr. David C. Funder

The Dissertation of Seth Michael Margolis is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

I am overwhelmed when I think about the quantity and quality of people that made this dissertation possible. Only a few people contributed directly to this particular project. But several people educated me on matters used in this dissertation. Scores of people supported me socially. And many thousands of people developed the technology that made this project possible.

Both software and hardware limitations would have made the project nearly impossible just two decades ago. However, the technological underpinnings necessary for this project go back at least 350 years. In fact, all of the innovations I discuss can be traced back to Aristotle, who lived over 2,300 years ago. I am particularly indebted to his contributions in logic, philosophy, and science.

Probability theory, first developed in the 17^{th} century, contributed greatly to both the statistics and technology used in this project. I am grateful to the following people, among others, for their contributions to probability theory: Gerolamo Cardano, Pierre de Fermat, Blaise Pascal, Pierre-Simon Laplace, Carl Friedrich Gauss, Adrien-Marie Legendre, and Robert Adrain. Following in the footsteps of these mathematicians, several statisticians are responsible for the modern statistics relied upon heavily in this dissertation. I would like to thank the following people, among others, for their contributions to modern statistics: Francis Galton, Karl Pearson, Ronald Fisher, Egon Pearson, Jerzy Neyman, and Donald Rubin. Building further upon probability and statistics, several people pioneered the field of machine learning. I thank these pioneers, who include, but not are limited to: Warren McCulloch, Walter Pitts, D. O. Hebb, Andrey

Tikhonov, Frank Rosenblatt, Vladimir Vapnik, Seppo Linnainmaa, Paul Werbos, Robert Tibshirani, Tin Kam Ho, Corinna Cortes, and Geoffrey Hinton.

Mathematical developments in probability, statistics, and machine learning are only useful if they can be performed in a reasonable amount of time. These tasks would be impossible without the development of computers. I am forever indebted to the following people, among others, for their roles in developing computer science: George Boole, Ada Lovelace, Herman Hollerith, Akira Kakashima, Alan Turing, Kathleen Both, Grace Hopper, John Backus, Sasaki Tadashi, Alan Kay, and Richard Stallman. In using statistics and computer programming, I often turned to the internet for assistance, and it delivered answers nearly every time. I am grateful to key figures in the development of the internet, including, but not limited to: Robert Kahn, Vint Cerf, and Tim Berners-Lee.

All statistical analyses for this project were done in Python. I would like to thank the Python Software Foundation and, specifically, Guido van Rossum. I am also grateful to the teams at Project Jupyter and IPython (in particular, Fernando Pérez) for creating software to enable beautiful markdown notebooks with Python code in read-eval-print loops. The data management for this project was an endeavor, but it was made much easier by Wes McKinney's excellent pandas library and the community of developers that continues to support it. All machine learning models were built using the amazing scikit-learn library. I am grateful to David Cournapeau and the scikit-learn development team for making it relatively easy to train and analyze complex machine learning models. Both pandas and scikit-learn use NumPy extensively. Thus, I would like to acknowledge the work of Travis Olpihant and the rest of the NumPy development team.

Although the software mentioned above makes advanced programming and statistical techniques relatively easy, a few books were quite helpful in learning how to use this software. I am particularly grateful to Mark Lutz, Jake VanderPlas, and Aurélien Géron for their wonderful books on data management and machine learning in Python.

I am indebted to Carol Ryff and the MIDUS investigators for producing such rich datasets. It took an unbelievable amount of work to randomly sample thousands of United States adults and have them complete hours-long surveys. I also want to thank them for including an incredible number of measures relevant to subjective well-being and making their data publicly available.

With this dissertation being a product of the UCR Psychology Department, it would not be possible without the incredible support from the Psychology Business Office. I want to specifically thank Toya Adams, Kirsten Alonso, Faye Harmer, and Renee Young for helping me with the many issues that arise as a graduate student.

I could not have completed this dissertation without the knowledge graciously shared by other people. I first learned from my family. My parents, Vicki and Howard Margolis, supported me in every way possible and I cannot thank them enough. I learned many lessons from my uncles, Richard Margolis and John Norton, as well as my grandparents, Florence Margolis and Bob Auerbach. I have also learned invaluable lessons from my incredible sets of friends and had a great deal of fun doing so. I am grateful to the friends I made throughout primary and secondary education, my undergraduate education, and in graduate school.

I have learned lessons more specific to psychology and statistics through amazing mentorship. Thank you to all the teachers who went above and beyond their job description to help me learn. I want to specifically acknowledge the following mentors: Brett Bentley, Tim Bono, Ryan Bogdan, Caitlin Carey, David Baranger, Randy Larsen, Gillian Russell, Julia Staffel, and Jake Gray.

In addition to all the teaching I received from professors, I also learned a great deal from fellow graduate student colleagues. There is no guide to most of the tasks of graduate school, and I have reached this point because of the guidance I received from other graduate students. I want to thank my current and former lab mates: Christina Armenta, Peter Ruberton, Megan Fritz, Lilian Shin, Julia Revord, Lisa Walsh, Annie Regan, and Karynna Okabe-Miyamoto.

My committee members, Dan Ozer and David Funder, contributed directly to this project. They both had great advice on how to analyze the MIDUS data and present the results. However, I started learning from them years before. I want to thank both for contributing greatly to my methodological and statistical knowledge.

My last committee member also directly contributed to this project, providing advice throughout the process. But being my advisor, Sonja Lyubomirsky did a lot more than help me directly with this dissertation.

Sonja provided a perspective on research that I appreciate more every day. She taught me how to think about research, both critically and with appreciation. Sonja helped me in particular with study design and writing, and I am grateful for that.

Anytime I had a question about a tricky situation, Sonja provided invaluable guidance. She knew how to handle every situation and I am still in awe of that ability. Sonja was hands-off, and she allowed me to spend time learning about statistics and programming, even when it did not directly benefit her. But as soon as I needed her to be hands-on, she was, immediately and fully.

Because of her professional success, Sonja has been presented with incredible opportunities. But more incredible to me is her generosity with these opportunities. I have been able to mingle and collaborate with top-notch scientists, only because Sonja is so generous.

Sonja's effect on my life goes beyond graduate training. When Sonja and I discussed matters beyond our work, I always found her perspective helpful. I hope to carry our connection forward beyond my days as a graduate student. Sonja, thank you for everything.

ABSTRACT OF THE DISSERTATION

Exploring Predictors of Subjective Well-Being Using Machine Learning and Propensity Score Techniques

by

Seth Michael Margolis

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, June 2020
Dr. Sonja Lyubomirsky, Chairperson

The majority of people around the world report wanting "the good life." But how do they achieve it? Most research in well-being science operationalizes "the good life" as subjective well-being, which is comprised of positive affect, negative affect, and life satisfaction. This project uses a nationally representative publicly available dataset from the Midlife in the United States (MIDUS) project (*N* = 4,378) to investigate predictors of subjective well-being. Importantly, this dataset contains measures of most of the previously identified predictors of subjective well-being. In addition to determining which predictors are stronger than others, this project also explores the utility of machine learning models and propensity score methods. Machine learning models are used in this project to determine the extent to which non-linear and interaction effects predict subjective well-being. I also evaluate the value of a propensity score method for identifying causal effects on subjective well-being.

Linear effects accounted for the vast majority of variance in subjective well-being. Machine learning models that could model non-linear and interaction effects predicted subjective well-being approximately as accurately as linear multiple regression models that only allowed for linear effects. Furthermore, linear multiple regression models appeared well-suited to model non-linear and interaction effects via variable transformations. Indeed, these models predicted subjective well-being as accurately or more accurately than machine learning models. Unfortunately, a propensity score method provided little value in identifying causal effects because it failed to eliminate relationships between a predictor of interest and other predictors. The role (or lack thereof) that machine learning and propensity score techniques could play in subjective well-being research is discussed.

Replicating previous research, sociability, physical health, disengagement from goals, sex life quality, wealth, and religious activity were among the strongest predictors of subjective well-being. Consistent with previous research in the U.S., demographic factors appeared to be relatively weak predictors of subjective well-being. Finally, control over one’s life—and financial and work matters in particular—was a strong predictor of subjective well-being, an effect that previous research may have downplayed.

Table of Contents

List of Tables

Exploring Predictors of Subjective Well-Being Using Machine Learning and Propensity Score Techniques

By definition, we should all want “the good life.” Given that the good life is the ultimate goal for a rational individual, an important question is “How does one live the good life?” To answer this question, however, scholars must decide what constitutes the good life. Philosophers have tackled this question for thousands of years (see Aristotle, 4th century B.C.E./2001) and often equate “the good life” to being high in “well-being,” which consists of the things that are ultimately good for every person. Thus, by definition, a person with “the good life” is high in well-being. After millennia of serious thought, debates continue over the elements of well-being (see Haybron, 2013).

Psychologists who wished to study well-being in the early-mid twentieth century could not wait for philosophers and other scholars to agree on the elements of the good life. Psychological research on well-being plunged ahead without worrying about the definitions of the good life or well-being (see Beckham, 1929; Watson 1930). However, to study well-being empirically, Diener (1984) defined “subjective well-being” as comprised of an emotional or affective component (i.e., positive affect and negative affect) and a cognitive component (i.e., life satisfaction). Although many philosophers do not endorse Diener’s subjective well-being as the correct definition of “well-being,” this definition has allowed psychologists to integrate and interpret one another’s research. Even if one’s personal definition of well-being differs from Diener’s definition, several types of well-being (e.g., desire fulfillment, eudaimonia) are highly correlated with subjective well-being (Margolis, Schwitzgebel, Ozer, & Lyubomirsky, 2020a).

Drawing Causal Conclusions

Perhaps the ultimate question motivating research on subjective well-being is simply: What causes subjective well-being? According to conventional methodological wisdom, randomized controlled trials are necessary to draw causal conclusions (see Campbell & Stanley, 1963). In a randomized controlled trial, also known as an experiment, participants are randomly assigned to conditions with different manipulations. With a large enough sample, random assignment ensures that the participants in each condition are quite similar on any (measured or unmeasured) construct. Thus, after the manipulations, any observed differences between individuals in different conditions can be causally attributed to the difference in the manipulations.

Randomized controlled trials are considered the gold standard for drawing casual conclusions and have been used extensively in research on subjective well-being (see Bolier et al., 2013, for a review). However, because randomized controlled trials are usually resource-intensive longitudinal studies, researchers cannot test every plausible cause of an outcome. Instead, investigators typically rely on the results of observational (i.e., correlational) studies to decide which effects to test with randomized controlled trials. Over the last few decades, this step-by-step process has unfolded in the subjective well-being literature (see Diener, Lucas, & Oishi, 2018). Although correlation does not imply causation, causation does imply correlation. Thus, using the logic of modus tollens, if correlational evidence points to a near-zero effect, a randomized controlled trial will likely find a near-zero causal effect. For this reason, correlational research can be used to eliminate some possible causes of an outcome and prioritize future manipulations by the expected size of their impact on the outcome.

The Strongest Predictors of Well-Being

For the reasons stated above, subjective well-being research began with correlational methods (see Wilson, 1967, for a review). Today, several decades of research have identified a long, diverse list of well-being correlates (Diener, Lucas, & Oishi, 2018).

But which correlates of subjective well-being are stronger predictors than others? This question is pivotal to determining which randomized controlled trials to conduct. However, little empirical or theoretical work has been done to answer it. Indeed, to my knowledge, only a couple of reviews have attempted to broadly summarize the relative importance of predictors of subjective well-being. Lyubomirsky, Sheldon, and Schkade (2005; see also Sheldon & Lyubomirsky, in press) estimated that 50% of the variance in happiness can be attributed to genes, 10% can be attributed to life circumstances, and the remaining 40% can be attributed to intentional activity (see Brown & Rohrer, 2018, for a critique of these estimates). In addition, Lyubomirsky, King, & Diener (2005) meta-analyzed cross-sectional correlations with well-being and grouped these correlates into nine categories. They found that aspects of sociability, likability, prosocial behavior, and positive perceptions of oneself and others were most predictive of well-being. However, a more nuanced approach is needed to inform future experimental work. For example, is prosocial behavior a stronger predictor of subjective well-being than physical health?

One approach is simply to turn to meta-analyses to determine which constructs are the strongest predictors of subjective well-being. Table 1 presents a list of constructs sorted by their meta-analytic correlations with well-being. Notably, subjective traits appear to be more predictive of well-being than objective qualities (e.g., household income). Unfortunately, the meta-analytic results do not directly indicate which predictors of well-being are stronger than others. A few realities render this task difficult. First, the predictors of subjective well-being are correlated to a considerable degree with one another. For example, Big Five traits are related to a host of outcomes related to subjective well-being (Ozer & Benet-Martinez, 2006). Due to the correlations among well-being predictors, one cannot simply compare correlations between predictors and subjective well-being. Second, the predictors of subjective well-being may interact with one another. For example, it is possible that that the effect of sociability on subjective well-being depends on levels of agreeableness. With so many correlates of subjective well-being, many interaction effects are possible. Obtaining data to test those interactions is difficult, because the dataset needs to include many measures and a large sample. Third, correlates of subjective well-being may relate to well-being in non-linear ways. For example, income's association with subjective well-being appears to be non-linear (Tay, Zyphur, & Batz, 2018). Fourth, estimates of the effects of well-being correlates often come from different studies that used different methods or populations. Thus, if a set of studies finds a higher correlation between meaning and well-being than another set of studies finds between gratitude and well-being, one cannot conclude that meaning correlates with well-being to a greater degree than does gratitude, as these correlations are impacted by the methods used in each study. Indeed, the meta-analyses presented in Table 1 used a variety of measures of well-being (e.g., life satisfaction vs. happiness) and were conducted on various populations (e.g., older adults vs. college students).

Fortunately, recent methodological advances can remedy these challenges in several ways. First, multiple regression can be used to account for correlations among predictors of subjective well-being. Second, interaction and non-linear effects can be estimated with regression and machine learning techniques. For example, one can examine the increase in predictive accuracy of a model that allows for interaction and non-linear effects (e.g., machine learning models, regression models with product terms) compared to a model that only allows for independent, linear effects (e.g., linear multiple regression models). To the extent that the more complex models are more predictive of subjective well-being than the simpler regression model, the greater extent to which predictors relate to subjective well-being via interaction and non-linear effects. Third, datasets containing multiple measures of predictors of subjective well-being administered to a single large sample have become available to researchers, allowing researchers to compare predictors of well-being in the same population. The availability of these methodological resources provides an opportunity to estimate the relative importance of predictors of subjective well-being with improved confidence.

Establishing Causal Effects

After correlational research has been used to reveal the largest predictors of well-being, targeted randomized controlled trials can then be conducted to establish causal effects. However, randomized controlled trials have significant limitations. Even with insights from correlational research, establishing causal effects with randomized controlled trials can be difficult or impossible. Due to their high cost, typically only one or a few causes can be tested in a single trial. Thus, many randomized controlled trials need to be conducted to study outcomes with many causes. In addition, because such trials are costly, non-representative convenience samples are often used, which can severely limit the generalizability of any causal conclusions. Furthermore, some constructs are difficult or impossible to manipulate. Does income impact well-being? A straightforward randomized controlled trial could be conducted, such that people are randomly assigned to receive either no money or a large sum of money every month. However, the cost of such a study is prohibitive. In addition, it may be unethical to manipulate certain correlates of well-being, such as personal beliefs, relationships, or health. Lastly, the effect size from a randomized controlled trial depends on which manipulation is used (e.g., whether gratitude is induced by prompting participants to write a gratitude letter or to count their blessings), the participants' levels of compliance, and the point at which the outcome is measured (e.g., immediately after the intervention or months later). Thus, many randomized controlled trials need to be conducted for each possible cause. For these reasons, randomized controlled trials only provide limited insights into the magnitude of causal effects.

Despite conventional wisdom, observational data can be well-suited for causal inference. Propensity score methods estimate causal effects from observational data by using techniques to minimize the relationship between the independent variable and possible confounding variables (Rosenbaum & Rubin, 1983). When the independent variable is continuous, this minimization can be done by using inverse probability of treatment weighting (IPTW) to create weights for each participant (Rosenbaum, 1987). The causal effect can then be estimated by predicting the outcome from the independent variable with participants weighted by the IPTW weights.

Propensity score methods provide accurate estimates of causal effects to the extent that 1) all possible confounding variables are unrelated to the independent variable when the IPTW weights are used and 2) the reverse causal path is not present (i.e., the presumed dependent variable does not cause the presumed independent variable). Thus, estimates of causal effects from correlational datasets can be most accurate when many measures are included in the dataset (Rubin, 1997). Estimates from a large correlational dataset typically have much greater external validity (i.e., generalizability) than randomized controlled trials and not depend on the selected manipulation, participant compliance, and the time of measurement. Furthermore, many more causal effects can be estimated from a large representative dataset than from randomized controlled trials, including effects that cannot be tested with randomized controlled trials for practical or ethical reasons.

Current Project

The current project seeks to take advantage of recent methodological advances in machine learning and propensity score methods to 1) establish the relative importance of predictors of subjective well-being, 2) evaluate the extent to which predictors relate to subjective well-being in non-linear and interactive ways, and 3) evaluate the utility of propensity score methods to estimate causal effects on subjective well-being. This study uses the MIDUS datasets to achieve these aims. To my knowledge, no previous study has examined a set of predictors of well-being as relevant to subjective-being and as broad as those in the MIDUS datasets. In addition, no other study to my knowledge has used machine learning and propensity score techniques with such a dataset.

Method

Participants

The first MIDUS data collection project (MIDUS 1) collected data in 1995 and 1996 using the University of Wisconsin Survey Center. A sample of adults living in the United States was recruited using random digit dialing. In 2004 and 2005, this sample was re-recruited (response rate of 69.8%) and assessed again on similar measures (MIDUS 2). In 2011-2014, a new sample (MIDUS R) was collected, with similar measures to MIDUS 2. These United States adults were also sampled through random digit dialing.

Some participants were recruited as siblings or twins of a participant or as part of designed oversampling of urban areas. These participants were excluded from my analyses to achieve a highly representative sample of United States adults. Initially, this resulted in a sample of 2,257 adults from MIDUS 2 and 3,577 adults from MIDUS R. However, participants were excluded if they had missing data on over 200 items (of 855 total items). This resulted in a final sample of 4,378 adults.

These adults ranged in age from 23 to 84 years old (M = 54.1, SD = 13.8). Of these adults, 87% were White, 54% were female, 64% were employed, 66% were married, 76% were Christian, 97% were heterosexual, and 44% had a bachelor's degree, according to self-reports of these demographic characteristics.

Procedure and Measures

Participants first responded to questions via a telephone interview and were then sent $25. One week later (10 days later for MIDUS R), participants were sent a self-administered questionnaire, instructions, a tape measure for body measurements, $10, and a business reply envelope to send back the complete self-administered questionnaire. When the self-administered questionnaires were received, the participant was mailed $25.

For a list of items used, please see the Supplemental Material and the MIDUS documentation available at https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/203/studies?archive=ICPSR.

Statistical Techniques

Hyperparameter tuning.

All the machine learning models used in this project (i.e., all models except correlations and propensity score models) featured hyperparameters that needed to be tuned. Hyperparameters are parameters that determine how the model is fit (i.e., trained or learned) from the data. For example, imagine an individual who is performing multiple regression with numerous predictors (called "features" in the machine learning literature). The data analyst could choose to use only 10 features to predict the outcome (called the "target" in the machine learning literature). The number of features to include—10—would be a hyperparameter.

When one trains a machine learning model, one needs to tune the hyperparameters to maximize fit (i.e., predictive accuracy). However, because machine learning models are often designed with the capacity to model complex effects, it is quite easy to train a model that overfits the data. This issue is analogous to the one faced when using polynomial regression. If one was fitting a polynomial regression model to 100 observations of a target (y) from a single feature (x) one could include 99 predictors of ($x, x^2, x^3 \ldots x^{99}$). If one fit this model to the data, the result would be perfect model fit (i.e., a multiple R of 1.0 or a correlation of 1.0 between predicted target values and actual target values). Despite this perfect fit, the model would be a poor model of the relationship between x and y. If new observations of x and y were collected, the model would almost surely provide predicted values of y that are quite different from the actual values of y. Thus, the coefficients of the polynomial regression model would not accurately reflect the larger population, essentially eliminating external validity (i.e., generalizability).

To prevent overfitting data and creating models without external validity, hyperparameters of machine learning models are often selected using k-fold cross-validation. This process of selecting hyperparameters to avoid overfitting is called regularization. In k-fold cross-validation, the dataset is randomly split into k samples of approximately equal size. Then, k-1 samples are used to train the machine learning model. The sample that was omitted from training (i.e., the validation dataset) is used to assess the fit of the trained model. Thus, the model is trained on a set of data, and then the model's fit is evaluated on a separate set of data. This process is then repeated so that each of the k samples is used as the validation dataset once. This produces k estimates of model fit, and the average of these estimates is often used as fit of the model, given a certain set of hyperparameters. In this project, a k of five was used in all k-fold cross-validation. Thus, models were repeatedly trained on approximately 80% of the data, and the models were evaluated on approximately 20% of the data.

Many sets of hyperparameters need to be tested using k-fold cross-validation to select the hyperparameters that produce the best cross-validated model fit. In this project, I used the common technique of grid search. In a grid search, a finite set of values are tested for each hyperparameter. Each combination of the hyperparameter values is tested. For example, if a model needed two hyperparameters, A and B, to be tuned, one could test A values of 0.1, 1, and 10 and B values of 1, 2, and 3. All nine possible combinations of A and B would be evaluated using k-fold cross validation. One set of values (e.g., 10 and 2) would be selected as the best values. More specific values can be tested in subsequent grid searches (e.g., testing values closer to 10 and 2). In this project, hyperparameters were tuned to two significant digits unless 1) the hyperparameter must be a whole number and 2) single-digit values produced the best model fit. Thus, if hyperparameters of $A = 15$ and $B = 0.024$ were found to result in the best model fit, this indicates that after repeating many grid searches, a grid search with possible values of A being 14, 15, and 16, and possible values of B being 0.023, 0.024, and 0.025 resulted in $A = 15$ and $B = 0.024$ being the set of hyperparameters that maximized cross-validated fit.

The final hyperparameters were used to produce an indication of the importance of each feature (e.g., a regression coefficient), if the particular machine learning model had that capability. However, the fit of this final model was not used as an estimate of the fit of that machine learning model to the data. The fit of a model on a dataset that uses hyperparameters selected through cross-validation on the same dataset produces biased estimates of model fit, and a nested cross-validation procedure is recommended to reduce this bias (Varma & Simon, 2006). In nested cross-validation, the data are randomly split into k samples, as in k-fold cross validation. Each sample is used as a validation dataset once in an "outer loop." The "outer" training data is split again to produce k subsamples, and the optimal hyperparameters for this "inner loop" are selected with grid search and k-fold cross-validation. In this grid search, I included five values for each hyperparameter, which included the optimal hyperparameter from the non-nested cross validation, and values two and four units away on the second significant digit. For example, if the optimal hyperparameter was .024, the inner loop grid search included .020, .022, .024, 026, and .028. These values were chosen to allow for different optimal hyperparameters across inner loops. The optimal hyperparameters of this inner loop are used to evaluate model fit on the outer loop validation sample. This process is then repeated so that each of the k outer loop samples is used as the outer loop validation dataset once. A k of five was used for both the inner and outer loops. This produced five estimates of model fit (R^2), and the square root of the average R^2 was used as the multiple R for a particular model and dataset.

Correlation.

I correlated all features to subjective well-being (i.e., the target). Because correlations are equivalent to a regression model with one predictor, the model is quite simple, and no regularization was necessary. Thus, the above hyperparameter tuning process was not necessary for the correlation analyses.

Lasso regression.

A linear multiple regression models a target as a linear combination (i.e., weighted sum) of features. Thus, a predicted value is calculated by multiplying a score on a feature by a coefficient and then repeating this process for each feature (which have unique coefficients) and summing the products. The coefficients are selected such that the mean squared error (i.e., the average of squared residuals) is minimized, which maximizes the multiple R of the model.

With many features, a linear multiple regression model will overfit data. One method of regularization is to use lasso regression. In lasso regression, rather than minimizing only the mean squared error, coefficients are selected to minimize a quantity calculated in the following way. First, the absolute values of the coefficients are summed and multiplied by a hyperparameter, α. The quantity being minimized is the sum of this product and the mean squared error. Thus, coefficients are selected to minimize the mean squared error while also minimizing the magnitude of the coefficients to regularize the model. The α hyperparameter controls the extent of regularization. A greater α leads to more regularization, as the magnitudes of the coefficients have a higher weight in the quantity to be minimized. The greater α is, the more that coefficients are reduced from what they would be in a linear multiple regression model without lasso regularization.

Importantly, lasso regularization tends to set some coefficients to zero. Because a coefficient of zero is equivalent to having not included that feature in the model, lasso regularization also performs feature selection. As described below, I used lasso regression for this purpose.

Ridge regression.

Ridge regression is quite similar to lasso regression. However, rather than summing the absolute values of the coefficients and multiplying them by the α hyperparameter, the coefficients are squared, summed, divided by two, and then multiplied by α. It is this product that is added to the mean squared error and that sum is minimized. Ridge regression provides regularization by reducing the magnitude of regression coefficients, but, unlike lasso regression, it does so without setting coefficients to zero (i.e., without performing feature selection).

Support vector regression.

Geometrically, the linear multiple regression techniques described above fit a flat hyperplane to the data in n-dimensional space where n equals the number of features plus one. In a simple case with one feature and one target, a line is fit to data points in a two-dimensional space. This is often visualized in a scatterplot. With two features, one can imagine data points in a three-dimensional space with a flat plane fit to the data. This plane is often allowed to be curvy in response surface analysis. Support vector regression, the regression analogue to support vector machines, is similar to this response surface analysis. In support vector regression, a curvy hyperplane is fit to data in n-dimensional space.

Unlike linear multiple regression, the goal of a support vector regression model is not to describe a hyperplane that minimizes squared residuals. Instead, the goal is to describe a hyperplane that has as many datapoints as possible within a certain distance of it. This distance, labeled ϵ is a hyperparameter of the model. To allow the hyperplane to be curvy, I used a radial basis function kernel. By allowing for a curvy hyperplane, there is increased potential for overfitting. Hyperparameters, C and γ, control how curvy the hyperplane can be. A lower value of C makes the hyperplane generally less curvy and a lower γ value makes the hyperplane less sensitive to individual data points. Thus, lower values of C and γ indicate more regularization. Unfortunately, support vector regression with a radial basis function kernel does not provide a metric that indicates how important each feature is in predicting target scores.

Random forests and extremely randomized trees.

Random forests—as well as a highly related machine learning technique, extremely randomized trees—are comprised of many decision trees. A decision tree begins at a root node. All cases being analyzed for that decision tree are used to train this node. At this node, cases are divided into two sets, which will be analyzed at the next layer of nodes. The cases are divided by using a single feature and a threshold. The feature and threshold are selected to reduce the variance of the target in the two resulting datasets. This procedure repeats at the next layer of nodes, and this process repeats until a certain condition (controlled by hyperparameters) is reached. The nodes in the last layer are called leaf nodes, and the prediction for cases meeting the conditions of a leaf node is the average of training cases in that node.

Decision trees tend to overfit data. In addition to regularizing the decision trees (described below), random forests implement bootstrap aggregating, also called bagging, of decision trees to increase predictive accuracy. In a random forest, the training cases are randomly sampled with replacement (i.e., bootstrapped), a decision tree is fit to each random sample, and then the predictions from each decision tree are averaged (i.e., aggregated). The random forests in this project were comprised of 100 decision trees. In a random forest, the best feature and threshold combination for splitting a node is considered among a random subset of features, which makes the individual trees more different from each other than they would be otherwise.

I used four hyperparameters to regularize the random forest models. First, I restricted the maximum depth of the decision trees (i.e., the number of layers of nodes). Second, the minimum number of cases at a leaf node was tuned. Third, I constrained the number of features that could be considered when looking to split a node. Lastly, I tuned the maximum number of leaf nodes that any decision tree could contain.

Extremely randomized trees are quite similar to random forests but are different in two ways. First, each decision tree is trained with the entire training dataset rather than a bootstrapped sample. Second, instead of selecting the optimal feature and threshold for splitting a node, for each feature, a threshold is randomly selected from (a uniform distribution of) the feature's range. The best combination of feature and threshold is then used to split the node. In this project, results with extremely randomized trees were quite similar to those with random forests, so I only report results for random forest models.

Like multiple regression coefficients, random forest models provide feature importance values. Features used earlier in the decision tree (i.e., closer to the root node and further from lead nodes) impact the prediction of a larger proportion of the training samples. The fraction of the samples that a feature contributes to is used to estimate a feature's importance. In addition, the decrease in variance by splitting a node using a certain feature contributes to the feature's importance rating. In a random forest model, all feature importances sum to unity and higher values indicate greater importance.

Artificial neural networks.

An artificial neural network operates much like a biological neural network. Each feature is represented by an input neuron. A "hidden" layer of neurons calculates a linear combination (i.e., a weighted sum) of the input neurons and outputs this sum to an output neuron. I used a rectified linear unit function, which outputs the linear combination if it is greater than zero, but outputs zero if the linear combination is less than zero. Initial testing showed that this approach improved the predictive accuracy of the neural networks over other activation functions. The output neuron calculates a prediction for the target using a linear combination of the hidden layer neurons.

Artificial neural networks can have many hidden layers between the input layer (i.e., features) and the output layer (i.e., predictions). However, in this project, I only used one hidden layer for two reasons. First, a more complex neural network seemed inappropriate for the number of cases in this project. Second, a neural network with one hidden layer can model anything a more complex neural network can, albeit with more neurons.

I regularized the artificial neural networks by tuning two hyperparameters. First, the number of neurons in the hidden layer was limited. Second, the weights between neurons were constrained in the same manner as coefficients in a ridge regression model. Thus, the hyperparameter α was tuned.

Propensity score analyses.

I conducted models where the target is predicted by a feature in a linear regression model where cases are weighted to minimize the relationship between potential confounders (i.e., other features) and the feature of interest. This weighting is achieved by inverse probability of treatment weighting (IPTW). The IPTW weights for each case are a fraction where the numerator is the conditional density of the normal distribution at the person's Z-score on the feature of interest. To calculate the denominators of the IPTW weights, the feature of interest is predicted from all other features using a linear multiple regression model. Then, for each case, the denominator of the IPTW weight is the conditional density of a normal distribution at the person's observed value on the feature of interest, where the normal distribution has a mean of the predicted value on the feature of interest and a standard deviation equal to the standard deviation of the residuals from this linear regression model. To avoid extreme weights, I set all weights above the 99^{th} percentile to the 99^{th} percentile of the weights.

I predicted the target from the feature of interest using a linear regression model where the cases are weighted by the IPTW weights. In one model, only the feature of interest was used to predict the target. However, because the IPTW weights did not completely eliminate the associations between the feature of interest and the other features, I also conducted a linear multiple regression model in which the other features were used as covariates and the same IPTW weights were used.

Data Cleaning

Creating a target and features.

Creating a target: subjective well-being. Positive and negative affect were measured with 13and 14 items, respectively. Participants were asked how much of the time they felt certain emotions during the past 30 days on a 5-point Likert scale. The positive emotions included cheerful, in good spirits, happy, calm, satisfied, full of life, close to others, belonging, enthusiastic, attentive, proud, active, and confident. Negative emotion items included sad, nervous, restless, hopeless, everything being effortful, worthless, lonely, afraid, jittery, irritable, ashamed, upset, angry, and frustrated. Composites of positive and negative affect were created by averaging each set of items.

A life satisfaction composite was formed by standardizing (i.e., Z-scoring) then averaging one item and two composites. The life satisfaction item asked participants the extent to which they were satisfied with life at present on a 4-point Likert scale. A life rating composite was formed by averaging three items asking participants to rate their present life, life 10 years ago, and life 10 years in the future using an 11-point Likert scale. A self-acceptance composite was calculated by averaging seven items from the self-acceptance sub-scale of the Psychological Well-Being Scale (Ryff & Keyes, 1995), which were answered using a 7-point Likert scale. The life satisfaction item, life rating composite, and self-acceptance composite were standardized and then averaged to form a life satisfaction composite. The self-acceptance items were considered measures of life satisfaction because disattenuated correlations between self-acceptance and measures of life satisfaction have been found to be .95 or greater (Margolis, Schwitzgebel, Ozer, & Lyubomirsky, 2019).

A subjective well-being composite was calculated by standardizing and then averaging the positive affect, negative affect (reverse scored), and life satisfaction composites. Using the standardized loadings from a one-factor exploratory factor analysis of the 38 subjective well-being items, ω_t equaled .96. The subjective well-being composite described above correlated with a composite where all items were weighted equally (i.e., simply averaged) at r = .97.

Creating trait features. The other items of the MIDUS surveys were scored into traits (i.e., trait features) using item averages. Some traits consisted of just one item, but most traits consisted of several items (with 748 items forming 189 traits). The items comprising each trait are listed in the Supplemental Material.

Item transformations. All items were transformed before forming trait composites. First, categorical items were dichotomized. Fortunately, all categorical variables could easily be dichotomized (e.g., combining the few homosexual and bisexual participants into one "not heterosexual" group). Some variables featured high positive skew because they were count variables. For example, one item asked about the number of hours per month one spends volunteering at a hospital. These highly positively skewed count variables were log transformed (base 10) after adding 1 to the variable (to prevent values of negative infinity). Next, some items were multiplied by -1 (i.e., reverse coded) so that higher scores on the associated trait indicated higher levels of that construct. Lastly, all items were standardized to have a mean of 0 and variance of 1.

Excluded items. Items were not used to form traits if they met any of six conditions. First, items were removed if they had a high proportion of missingness, which occurred in two situations. If an item was present in one survey but not the other, the item would have a high proportion of missingness. For example, many MIDUS R items pertained to the 2008 financial crisis, which were not present in the MIDUS 2 survey. In addition, some items simply had low response rates, often because they were not applicable. For example, most people could not provide the age at which they had a heart attack, because most people had not had a heart attack. Items that had more than 500 missing cases were removed.

Second, I removed two items that did not have any variance: whether one's half-sister had a heart attack and whether one received herbal cancer treatment.

Third, any variable that was calculated or determined from other variables was removed, as this was redundant with averaging items into traits.

Fourth, items tautologically related to subjective well-being were removed (following imputation, described below). These items either concerned mental health, were labelled as measures of well-being in the MIDUS documentation or asked participants to rate how well the previous day or month went. However, the subscales of the Psychological Well-Being Scale (e.g., environmental mastery, purpose in life) were included as they are not tautologically related to subjective well-being. Indeed, some argue that eudaimonic constructs, such as those assessed with the Psychological Well-Being Scale, are best seen as causes, rather than constituents, of well-being (Sheldon, 2018).

Fifth, some items simply seemed far more specific than other trait-level variables and could not be formed into a composite. For example, an item concerning whether one would prefer a coronary bypass or medication to solve a heart issue was removed. In addition, several items about a respondent's values or beliefs were specific and could not reasonably be averaged into traits.

Sixth, some items were redundant with other items. For example, because three items concerned blood pressure, the two of the three items with the least specific response options were removed.

Trait labels. Traits were labeled as subjective or objective. All traits were assessed via self-report, and thus all were subjectively reported. However, some traits regarded subjective ratings (e.g., rating the strength of one's financial situation using a Likert scale), whereas others regarded more objective characteristics (e.g., estimating income).

Creating domain features. Based on the content of the objective traits, 17 objective domains were formed. These domains were created by averaging traits, but some domains consisted of one trait (e.g., female status). The subjective traits were entered into an exploratory factor analysis using the minimum residual method and oblimin rotation. A scree plot suggested 11 factors (see Figure 1). Furthermore, the loadings of these 11 factors produced interpretable factors. Thus, 11 subjective domain factor scores were extracted, resulting in a total of 28 domains.

Creating weak features. An advanced machine-learning model may not outperform a linear regression model if the linear regression model explains most of the variance in the target. To evaluate the machine learning models in scenarios where this is not the case, I created datasets of weak traits and weak domains. The goal was to select the traits (or domains) that were most weakly associated with subjective well-being. I selected the weakest features, such that a lasso regression model predicting subjective well-being from these features would have a nested cross validation multiple R of approximately .50.

To create a dataset of weak traits, I removed all traits with a correlation above .15 in magnitude with subjective well-being. When the remaining 80 features were used to predict subjective well-being in a lasso regression model, the nested cross-validation multiple R was .50.

To create a dataset of weak domains, I removed all traits with a correlation above .22 in magnitude with subjective well-being. When the remaining 15 features were used to predict subjective well-being in a lasso regression model, the nested cross-validation multiple R was .49.

Feature selection. The performance of machine learning models is often reduced, rather than unchanged, by the inclusion of features that are not predictive of the target when the other features are included in the model (see, for example, John, Kohavi, &

Pfleger, 1994). Thus, for trait, domain, weak trait, and weak domain datasets, I removed features if their regression coefficient in a lasso regression model was zero.

To examine the ability of linear regression models to incorporate non-linear and interaction effects on subjective well-being, I also made datasets with squared and product terms. For all four of the above datasets, I created three additional datasets: one with untransformed (i.e., linear) and squared features, one with untransformed and product features, and one with untransformed, squared, and product features. Thus, I created 16 datasets in total. I used lasso regression to perform feature selection on each dataset. See Table 2 for results of these lasso regression models and see Table 3 for a list of subjective domains with representative items from those domains.

Based on a lasso regression model, a smorgasbord of features was removed from the trait dataset. Previous research does not identify many of these features as important predictors of subjective well-being. However, some important exceptions include objective chronic pain, objective health limitations, subjective generativity, subjective pessimism, and subjective religiosity. Many of these features are likely to be at least somewhat predictive of subjective well-being, but they failed to predict subjective well-being over and above the selected features. The same possibility applies to the features removed from the domain dataset: objective employment status, objective giving and receiving, and subjective resilience.

Missingness and imputation. Of all cells in the item-level data, 1.5% were missing. Because items with more than 500 missing cases were removed, no item had more than 11.4% of its cells missing. However, most items had some missingness; only 39 items had no missing cells.

Before forming trait composites, the cells of the item data were imputed because the analyses for this project could not be performed on data that have missing values (i.e., listwise deletion would be used, removing most of the sample). The imputation took place after items were transformed. In addition, items that had high missingness, had no variance, or were determined by other variables were removed before imputation. However, some items that would be removed before forming traits were included in the imputation process. These items were ones that would later be removed because they were tautologically related to subjective well-being, redundant with other items, or more specific than traits and could not be used in a composite.

Cells were imputed using a k-nearest neighbor approach. First, each missing cell was imputed using the item medians. The values of these cells were modified further. The median imputation was simply necessary to start the k-nearest neighbor imputation. Next, the variable with the fewest missing cells (excluding those with no missing cells) was imputed. This variable was imputed based on data of the 50 features with the strongest correlations to the variable being imputed. Using these data, the five nearest neighbors (i.e., cases) of a missing case were identified. The five nearest neighbors were cases with the five lowest Euclidean distances from the missing case on the 50 used features. Thus, for each possible neighbor (i.e., all cases except the targeted missing case), the differences between the possible neighbor and the targeted missing case on each of the 50 used features was squared, those 50 squares were summed, and then the square root of the sum equaled the distance of that possible neighbor case from the targeted missing case. Of the five nearest neighbors, one case was randomly selected, and the targeted missing cell was replaced with the score of the randomly selected neighbor on the target variable. This was then repeated for each missing case of the target variable. This procedure was repeated for each variable with missing data, conducting imputation on variables with the least missing cells first. The imputation process was repeated over the dataset four more times for a total of five rounds of imputation, which greatly reduced the impact of the initial median imputation.

I used a k-nearest neighbor imputation for several reasons. First, simulation work has found that this method performs well (Batista & Monard, 2002; Chen & Shao, 2000; Jonsson, & Wohlin, 2004). Second, missing values are imputed with values that already exist for that variable. Thus, unlike with many regression-based imputation methods, the imputed values are possible scores on the variable. Third, this method uses what is known about a participant to impute values, and this likely does not inflate model error as much as a simpler technique like median imputation. Lastly, k nearest neighbors avoids overfitting by randomly selecting from the k (in this case, five) most likely cases.

Results

Model Fit

Using a ridge regression model, both the trait and domain features were highly predictive of subjective well-being (see Table 4). For both domain and weak domain features, the inclusion of square and product terms only slightly improved model fit. For each of these feature datasets, models with squared terms featured approximately the same fit as models with product terms. For the traits and weak traits features, the inclusion of product terms substantially increased model fit.

The support vector regression models (see Table 5) fit better than the ridge regression models with only untransformed features. For the domains and weak domains features, the support vector regression models fit approximately as well as the ridge regression models with square and product terms. However, for the traits and weak traits features, the support vector regression models could not rival the fit of the ridge regression models with transformed (i.e., squared and product) features.

The random forest models (see Table 6) generally did not fit as well as the ridge regression models. For all sets of features except weak domains, the random forest models performed worse than the ridge regression models with untransformed features. For the weak domains features, the random forest model fit slightly better than the ridge regression model with untransformed features. However, the random forest model with weak domains features fit slightly worse than ridge regression models with square or product terms.

Like the support vector regression models, the artificial neural network models (see Table 7) fit slightly better than the ridge regression models with only untransformed features**.** For the domains and weak domains features, the artificial neural network models fit approximately as well as the ridge regression models with square and product terms. However, for the traits and weak traits features, the artificial neural network models fit worse than the ridge regression models with transformed features.

Predictors of Subjective Well-Being

See Table 8 for the list of traits and how strongly they predict subjective well-being. Most traits exhibited an adequate degree of internal consistency (as measured by ω_t) to interpret correlations, regression coefficients, and importance values from random forest models. Generally, these metrics agreed on which trait features were most predictive of subjective well-being. However, the ridge regression coefficients and random forest importances were more similar to each other than either were to the correlations, which likely occurred because the former two metrics statistically control for other features whereas correlations do not control for other features. When compared with objective traits, subjective traits tended to be more predictive of subjective well-being. Indeed, of the 30 most predictive traits, only one (aches) was objective.

Unsurprisingly, the components of psychological well-being (e.g., environmental mastery, purpose in life, positive relations with others) were among the traits most predictive of subjective well-being. These traits were selected to comprise psychological well-being because theories posit them as important to human flourishing (Ryff, 1989). Another intuitive result was that rating particular areas of one's life as going well (e.g., physical health, work situation, financial situation) was highly predictive of subjective well-being. Perhaps more interestingly, ratings of control in one's life (e.g., control over life, control over work situation, health control, autonomy) were highly associated with subjective well-being. In addition, paralleling previous findings, subjective well-being was highly related to subjective ratings regarding social interactions (e.g., social integration, extraversion, social contribution). Demographic factors (e.g., pregnancy, White status) were among the weakest associations with subjective well-being. Although a trait labeled “subjective control” was weakly associated with subjective well-being, this trait reflects control in the sense of cautiousness and constraint.

See Table 9 for the list of domains and how strongly they predict subjective well-being. Mirroring the results with traits, subjective domains were generally more predictive of subjective well-being than objective traits were, with objective physical health issues being an exception. In addition, subjective evaluations of control, physical health, and sociability were strongly associated with subjective well-being. However, domain-level analyses led to some insights that were not clearly visible at the trait level. For example, disengagement (e.g., withdrawing from a stressful event), judgments of socioeconomic status, and subjective evaluations of one’s sex life were all important predictors of subjective well-being. Financial situation, religious activity, and human contact (i.e., the frequency and duration of time spent socializing with friends, family, etc.) were among the most predictive objective domains. Once again, demographic factors (education, physical size, and female status) were among the weakest predictors of subjective well-being.

Propensity Score Analyses

As described above, I conducted two models with the IPTW weights for each feature. One model used only the feature of interest to predict subjective well-being and another model used all other features as covariates. If the IPTW weights were successful in eliminating associations between the feature of interest and all other features, then these models would provide the same result. However, as shown in Tables 7 and 8, this was not the case. The coefficients were substantially smaller in models with covariates, indicating that even with IPTW weights, a feature of interest was still related to the other features.

The propensity score analyses generally agreed with the ridge regression and random forest analyses on which features are most predictive of subjective well-being. The propensity score analyses agreed less with the correlation results, presumably because propensity score analyses involve statistical control of other features (like ridge regression and random forest models) and correlations do not control for other features.

Discussion

The primary goal of this project was to examine the utility of using machine learning to model subjective well-being. My results consistently showed that simpler models limited to linear effects (e.g., multiple regression) performed approximately as well as machine learning models. This finding made it relatively easy to determine which predictors of subjective well-being were more than important than others. Elements of psychological well-being, control, and subjective physical health were among the strongest predictors of well-being. However, causal conclusions could not be made from these cross-sectional data, as propensity score methods failed to minimize the relationships between potential causes of well-being and confounding variables.

Non-Linear and Interaction Effects on Subjective Well-Being and Machine Learning's Ability to Detect Them

Based on the results of the ridge regression models, a small proportion of the variance in subjective well-being can be explained by non-linear and interaction effects.

At the domain level, it appeared that more complex machine learning models (i.e., support vector regressions, random forests, and artificial neural networks) could predict subjective well-being approximately as accurately as ridge regression models with squared and product terms. However, when using the traits and weak traits features, these machine learning models did not fit as well as the simpler ridge regression model with squared and product terms. Importantly, the extent of non-linear and interaction effects appeared greatest in the trait-level datasets. Thus, when non-linear and interaction effects were most present, machine learning models did not fit as well as a model that predicted subjective well-being as a linear combination of features, given that those features include squared and product terms.

This finding may be unique to subjective well-being. Perhaps most of the variance in subjective well-being can truly be explained by linear effects. However, more broadly, machine learning models may not be particularly helpful when the target is continuous (or ordinal with enough categories that it is treated as continuous). Some of the more famous uses of machine learning models include identifying spam emails (Guzella & Caminhas, 2009), predicting purchasing behavior (Zuo, Yada, & Ali, 2016), and optical character recognition (Bhatia, 2014). Notably, all of these tasks involve classification (i.e., categorical targets), not regression (i.e., continuous targets). Among the machine learning models used in this project, the random forest models fit particularly poorly, and this may be due to a mismatch between their architecture (dichotomous node splitting in decision trees) and a continuous target.

Predictors of Subjective Well-Being

Most results regarding particular predictors of subjective well-being match those found in previous research. For example, mirroring the results of this project, previous research has found high associations between subjective well-being and components of psychological well-being (Disabato, Goodman, Kashdan, Short, & Jarden, 2016), satisfaction with particular areas of life (Rojas, 2006), sociability (Cooper, Okamura, & Gurka, 1992), physical health (Cross, Hofschneider, Grimm, & Pressman, 2018), disengagement (Wrosch, Scheier, Miller, Schulz, & Carver, 2003), sex life (Margolis, Schwitzgebel, Ozer, & Lyubomirsky, 2020b), wealth (Ng, 2013), and religious activity (Tay, Li, Myers, & Diener, 2014). I also found that most demographic factors were poor predictors of subjective well-being, as previous research in the United States has found (Andrews & Withey, 1976, Diener, Suh, Lucas, & Smith, 1999). The findings of this project do align well with comparisons of meta-analytic effect sizes from previous research. Table 1 suggests that meaning, neuroticism, optimism, and extraversion are among the strongest predictors of well-being, as was found in this project. However, my findings do not perfectly align with prior work. For example, self-esteem and physical health were stronger predictors of subjective well-being than was suggested by previous meta-analyses. However, some previous research has found self-esteem to be a strong predictor of well-being (Lyubomirsky, Tkach, & DiMatteo, 2006).

Perhaps the most surprising result was the strength of the association between control and subjective well-being. Both previous empirical results and theorizing suggest that autonomy (which is synonymous with or highly related to a sense of control) is important for subjective well-being (see self-determination theory; Deci & Ryan, 2012). Previous research has also revealed that another highly related concept, locus of control (i.e., whether events in one's life are determined by one's actions or external factors), is highly predictive of well-being (Lu, 1997; Lu, Shih, Lin, & Ju, 1999). However, I was surprised that a sense of control was the strongest predictor of subjective well-being at the domain level. Indeed, most previous research considers other factors as more important predictors of well-being. For example, autonomy and control are scantly mentioned in the most recent major review of subjective well-being research (Diener, Lucas, & Oishi, 2018).

The traits with the strongest loadings on the subjective control domain include items regarding the extent to which individuals feel they have control over their life, work, health, and finances. Perhaps these factors are particularly important for subjective well-being, and having a sense of control over them is not being measured using common measures of autonomy, like the Balanced Measure of Psychological Needs (Sheldon & Hilpert, 2012) ,which assess control at a general level (e.g., "My choices expressed my 'true self'"). Indeed, items regarding control over health, finances, and work may be relatively less appropriate for university students, the population of most psychological studies. Many university students have health insurance, are supported financially by their parents, and are not employed.

Propensity Score Analyses and Causal Conclusions

In evaluating the extent to which particular features predict subjective well-being, the results of the propensity score analyses were quite similar to those found with ridge regression and random forest models. The supposed benefit of propensity score analyses is that they allow one to draw causal conclusions. However, I do not believe it would be appropriate to make causal claims based on the results of propensity score analyses in this project. For propensity score methods to provide accurate estimates of causal effects, the IPTW weights must eliminate associations between the feature of interest and other features. However, such associations remained, as evidenced by the smaller estimates when other features were included in the regression model with IPTW weights, compared to estimates when only the feature of interest was included in the regression model with IPTW weights. In addition, propensity score methods do not indicate the direction of the causal effect. In this project, many of the potential causes of subjective well-being could be effects, rather than causes, of subjective well-being. Thus, I believe that propensity score analyses did not help identify causal effects in this project.

Limitations and Future Directions

Although the MIDUS datasets are larger and more representative than most datasets used in psychology, they still have limitations. The MIDUS datasets used random digit dialing to collect a highly representative sample of United States adults. However, the representativeness of the sample likely decreased when participants with high rates of missingness were removed. In addition, findings from this project may not generalize to other countries. For example, control may be a weaker predictor of well-being in countries that are more interdependent than the United States.

The MIDUS datasets are large in that they have both many participants and many measured variables. However, a sample size of 4,378 is small for most machine learning tasks. With more cases, complex effects may become more reliable (i.e., persist under cross-validation). Thus, the machine learning models I used may provide more accurate predictions in datasets much larger than the current dataset. However, one would expect that 4,378 cases is enough to detect basic non-linear and interaction effects with machine learning models.

Although the MIDUS datasets contain many variables, certainly some constructs important for well-being were not measured (e.g., gratitude, work productivity). A study with all constructs relevant to subjective well-being, though quite costly, would likely include a longer list of constructs that are important predictors of subjective well-being.

Unfortunately, all measures involved self-report. I labeled some measures as "objective" because they attempted to measure objective quantities (e.g., hours spent helping a family member). However, even these measures are likely impacted by self-report biases (e.g., socially desirable responding, extreme responding, and acquiescence). In addition, because all measures were assessed via self-report, the associations between features and subjective well-being are likely inflated due to the presence of common method variance. Lastly, self-report can be burdensome on participants, especially with the length of the questionnaires used in the MIDUS project. Thus, some measures may be impacted by fatigue participants felt while completing a long questionnaire.

Because this project focused on only one target—subjective well-being— its conclusions may not extend to other targets. For example, wealth could be predicted by many non-linear and interaction effects, and machine learning models could excel in providing accurate predictions of wealth. Similarly, propensity score methods may be of more value with different targets.

Concluding Thoughts

Machine learning techniques have gained much praise recently for performing tasks that were reserved for science fiction decades ago (e.g., optical character recognition, spam filtering). Similarly, interest in propensity score methods has risen as researchers desire methods that allow them to draw valid causal conclusions from cross-sectional data. However, these techniques will not add much value in particular situations. Specifically, machine learning techniques will not provide much benefit when a linear multiple regression model can capture most of the reliable relationships between features and a target. And propensity score methods will not be useful when they fail to eliminate the relationships between potential confounders and a feature of interest. In this project, both of these conditions were present. Thus, machine learning techniques and propensity score methods provided little additional insight over linear multiple regression models.

The relative importance of linear effects over non-linear and interaction effects allowed for easier interpretation of which features were most predictive of subjective well-being. Although all measures were completed via self-report, features consisting of subjective evaluations consistently demonstrated stronger associations with subjective well-being than did features reflecting relatively more objective characteristics. Notably, most of the results were consistent with previous literature on the key predictors of subjective well-being (see Diener, et al., 2018 for a review). However, some results highlighted relatively neglected constructs in the subjective well-being literature. Specifically, disengaging from goal pursuit when facing adversity and perceiving oneself as in control of one's life—and one's financial, health and work domains, in particular—may be more strongly associated with subjective well-being than previously thought.

The results of this project provide future subjective well-being researchers with an ordered list of potential causes of subjective well-being. Because propensity score methods did not provide much information on causal effects in this project, future researchers may focus their attention on conducting randomized controlled trials to test these potential causes of subjective well-being. Of course, much research in this area has already been conducted and is ongoing (see Boiler et al., 2013). The relatively little knowledge gained from propensity score analyses in this project highlights the value of experimental work in subjective well-being research, despite its limitations. Furthermore, the experimental approach is likely to benefit from the finding that non-linear and interaction effects on subjective well-being appeared to be small in magnitude. Propensity score methods and machine learning techniques also provided little additional understanding of subjective well-being over what was derived from linear multiple regression models. In sum, however, my findings are likely to advance psychological scientists' understanding of subjective well-being and inform the precise ways that future research on subjective well-being should proceed.

References

Alarcon, G. M., Bowling, N. A., & Khazon, S. (2013). Great expectations: A meta-analytic examination of optimism and hope. *Personality and Individual Differences, 54,* 821-827.

Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being: Americans' perceptions of life quality.* New York: Plenum.

Anglim, J., Horwood, S., Smillie, L. D., Marrero, R. J., & Wood, J. K. (2020). Predicting psychological and subjective well-being from personality: A meta-analysis. *Psychological Bulletin, 146,* 279-323.

Aristotle. (4th century B.C.E./2001). Nichomachean ethics. In R. McKeon (Ed.), *The basic works of Aristotle* (pp. 928–1112). New York: The Modern Library.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46,* 399-424.

Batista, G. E., & Monard, M. C. (2002). A Study of k-nearest neighbour as an imputation method. *Hybrid Intelligent Systems, 87,* 251-260.

Beckham, A. S. (1929). Is the negro happy? *The Journal of Abnormal and Social Psychology,* 24, 186-190.

Bhatia, E. N. (2014). Optical character recognition techniques: A review. *International Journal of Advanced Research in Computer Science and Software Engineering, 4,* 1219-1223.

Bolier, L., Haverman, M., Westerhof, G. J., Riper, H., Smit, F., & Bohlmeijer, E. (2013). Positive psychology interventions: A meta-analysis of randomized controlled studies. *BMC public health, 13,* 119.

Breiman, L. (2001). Random forests. *Machine Learning, 45,* 5-32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont: Wadsworth.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology, 163,* 1149-1156.

Brown, N. J., & Rohrer, J. M. (2018). Easy as (happiness) pie? A critical evaluation of a popular model of the determinants of well-being. *Journal of Happiness Studies,* 1-17.

Campbell, D.T. & Stanley, J. C. (1963). Experimental and quasi- experimental designs for research. In N.L. Gage (Ed.), *Handbook of Research on Teaching.* Chicago: Rand McNally.

Chen, J., & Shao, J. (2000). Nearest neighbor imputation for survey data. *Journal of Official Statistics, 16,* 113-131.

Cooper, H., Okamura, L., & Gurka, V. (1992). Social activity and subjective well-being. *Personality and Individual Differences, 13,* 573-583.

Cross, M. P., Hofschneider, L., Grimm, M., & Pressman, S. D. (2018). Subjective well-being and physical health. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of well-being.* Salt Lake City: DEF Publishers.

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (p. 416–436). London: SAGE Publications.

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124,* 197-229.

Diener, E., Lucas, R. E., & Oishi, S. (2018). Advances and open questions in the science of subjective well-being. *Collabra: Psychology, 4,* 15.

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125,* 276–302.

Disabato, D. J., Goodman, F. R., Kashdan, T. B., Short, J. L., & Jarden, A. (2016). Different types of well-being? A cross-cultural examination of hedonic and eudaimonic well-being. *Psychological Assessment, 28,* 471-482.

Wrosch, C., Scheier, M. F., Miller, G. E., Schulz, R., & Carver, C. S. (2003). Adaptive self-regulation of unattainable goals: Goal disengagement, goal reengagement, and subjective well-being. *Personality and Social Psychology Bulletin, 29,* 1494-1508.

Giluk, T. L. (2009). Mindfulness, Big Five personality, and affect: A meta-analysis. *Personality and Individual Differences, 47,* 805-811.

Goh, A. T. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering, 9,* 143-151.

Guzella, T. S., & Caminhas, W. M. (2009). A review of machine learning approaches to spam filtering. *Expert Systems with Applications, 36,* 10206-10222.

Hackney, C. H., & Sanders, G. S. (2003). Religiosity and mental health: A meta–analysis of recent studies. *Journal for the Scientific Study of Religion, 42,* 43-55.

Hajage, D., Tubach, F., & De Rycke, Y. (2017). Overfitting in propensity score model: A commentary on "propensity score model overfitting led to inflated variance of estimated odds ratios" by Schuster et al. *Journal of Clinical Epidemiology, 88,* 160-161.

Haybron, D. M. (2013). *Happiness: A very short introduction.* Oxford University Press.

Howell, R. T., & Howell, C. J. (2008). The relation of economic status to subjective well-being in developing countries: A meta-analysis. *Psychological Bulletin, 134,* 536-560.

Howell, R. T., Kern, M. L., & Lyubomirsky, S. (2007). Health benefits: Meta-analytically determining the impact of well-being on objective health outcomes. *Health Psychology Review, 1,* 83–136.

John, G. H., Kohavi, R., & Pfleger, K. (1994). Irrelevant features and the subset selection problem. In W. W. Cohen & H. Hirsh (Eds.), *Machine learning proceedings of the eleventh international conference* (pp. 121-129). Morgan Kaufmann Publishers: San Francisco, CA.

Jonsson, P., & Wohlin, C. (2004). An evaluation of k-nearest neighbour imputation using likert data. *Proceeding of the 10th International Symposium on Software Metrics,* 108-118.

Kuykendall, L., Tay, L., & Ng, V. (2015). Leisure engagement and subjective well-being: A meta-analysis. *Psychological Bulletin, 141*, 364-403.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521,* 436-444.

Lu, L. (1999). Personal or environmental causes of happiness: A longitudinal analysis. *The Journal of Social Psychology, 139,* 79-90.

Lu, L., Shih, J. B., Lin, Y. Y., & Ju, L. S. (1997). Personal and environmental correlates of happiness. *Personality and Individual Differences, 23,* 453-462.

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin, 131,* 803-855.

Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology, 9,* 111-131.

Lyubomirsky, S., Tkach, C., & DiMatteo, M. R. (2006). What are the differences between happiness and self-esteem?. *Social Indicators Research, 78,* 363-404.

Margolis, S., Schwitzgebel, E., Ozer, D. J., & Lyubomirsky, S. (2019). A new measure of life satisfaction: The Riverside Life Satisfaction Scale. *Journal of Personality Assessment, 101,* 621-630.

Margolis, S., Schwitzgebel, E., Ozer, D. J., & Lyubomirsky, S. (2020a). Empirical relationships among five types of well-being. To appear in Lee, M. T., Kubzansky, L. D., & VanderWeele, T. J. (Eds.), *Measuring well-being: Interdisciplinary perspectives from the social sciences and the humanities.* New York: Oxford University Press.

Margolis, S., Schwitzgebel, S. Ozer, D.J., & Lyubomirsky, S. (2020b). *Two new measures of "objective" well-being: The Riverside Eudaimonia scale and the Rich & Sexy Well-Being Scale.* Manuscript in preparation.

Ng, W. (2013). The duality of wealth: Is material wealth good or bad for well-being?. *Journal of Social Research & Policy, 4*, 7-19.

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

Pinquart, M. (2002). Creating and maintaining purpose in life in old age: A meta-analysis. *Ageing International, 27,* 90-114.

Pinquart, M., & Sörensen, S. (2000). Influences of socioeconomic status, social network, and competence on subjective well-being in later life: A meta-analysis. *Psychology and Aging, 15,* 187-224.

Rojas, M. (2006). Life satisfaction and satisfaction in domains of life: Is it a simple relationship?. *Journal of Happiness Studies, 7,* 467-497.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41-55.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *The Journal of the American Statistician, 82,* 387–394

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65,* 386–408.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127,* 757-763.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57,* 1069-1081.

Schuster, T., Lowe, W. K., & Platt, R. W. (2016). Propensity score model overfitting led to inflated variance of estimated odds ratios. *Journal of Clinical Epidemiology, 80,* 97-106.

Sheldon, K. M. (2018). Understanding the good life: Eudaimonic living involves well-doing, not wellbeing. In J. Forgas & R. Baumeister (Eds.), *The social psychology of living well* (pp. 116-136). New York: Routledge.

Sheldon, K. M., & Hilpert, J. C. (2012). The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. *Motivation and Emotion, 36,* 439-451.

Sheldon, K. M., & Lyubomirsky, S. (in press). Revisiting the Sustainable Happiness Model and pie chart: Can happiness be successfully pursued? *The Journal of Positive Psychology.*

Tay, L., Li, M., Myers, D., & Diener, E. (2014). Religiosity and subjective well-being: An international perspective. In *Religion and spirituality across cultures* (pp. 163-175). Springer, Netherlands.

Tay, L., Zyphur, M., & Batz, C. L. (2018). Income and subjective well-being: Review, synthesis, and future research. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of well-being.* Salt Lake City, UT: DEF Publishers.

Thoemmes, F., & Ong, A. D. (2016). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood, 4,* 40-59.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological),* 58, 267-288.

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics, 7,* 91.

Werbos, P. J. (1975). *Beyond regression: New tools for prediction and analysis in the behavioral sciences.* Unpublished doctoral dissertation. Harvard University, Cambridge, MA.

Wilson, W. R. (1967). Correlates of avowed happiness. *Psychological Bulletin, 67,* 294-306.

Zessin, U., Dickhäuser, O., & Garbade, S. (2015). The relationship between self□compassion and well□being: A meta□analysis. *Applied Psychology: Health and Well□Being, 7,* 340-364.

Zuo, Y., Yada, K., & Ali, A. S. (2016). Prediction of consumer purchasing in a grocery store using machine learning techniques. *Third Asia-Pacific World Congress on Computer Science and Engineering,* 18-25.

Table 1

Meta-Analytic Correlations with Well-Being

Construct	Meta-Analytic Correlation	Citation
Meaning in Life	.47	Pinquart (2002)
Self-Compassion	.47	Zessin, Dickhäuser, & Garbade (2015)
Neuroticism	-.46	Anglim, Horwood, Smillie, Marrero, & Wood (2020)
Optimism	.43	Alarcon, Bowling, & Khazon, 2013
Extraversion	.37	Anglim, Horwood, Smillie, Marrero, & Wood (2020)
Sociability	.37	Lyubomirsky, King, & Diener (2005)
Conscientiousness	.36	Anglim, Horwood, Smillie, Marrero, & Wood (2020)
Prosocial Behavior	.35	Lyubomirsky, King, & Diener (2005)
Mindfulness	.34	Giluk (2009)
Self-Esteem	.31	DeNeve & Cooper (1998)
Leisure Engagement	.26	Kuykendall, Tay, & Ng (2015)
Agreeableness	.25	Anglim, Horwood, Smillie, Marrero, & Wood (2020)
Competence	.21	Pinquart, & Sörensen (2000)
Openness	.19	Anglim, Horwood, Smillie, Marrero, & Wood (2020)
Household Income	.18	Howell & Howell (2018)
Physical Health	.14	Howell, Kern, & Lyubomirsky (2007)
Religiosity	.10	Hackney & Sanders (2003)

Table 2

Lasso Regression Results

Features	Squares	Products	Coefficients equal to 0	Coefficients not equal to 0	α
Traits	No	No	82	106	0.0065
Traits	Yes	No	219	129	0.0085
Traits	No	Yes	17469	297	0.0171
Traits	Yes	Yes	17633	293	0.0171
Domains	No	No	4	24	0.0034
Domains	Yes	No	10	43	0.0034
Domains	No	Yes	317	89	0.0119
Domains	Yes	Yes	313	118	0.0096
Weak Traits	No	No	12	68	0.0060
Weak Traits	Yes	No	34	126	0.0058
Weak Traits	No	Yes	2998	242	0.0240
Weak Traits	Yes	Yes	3057	263	0.0232
Weak Domains	No	No	0	15	0.0002
Weak Domains	Yes	No	0	28	0.0027
Weak Domains	No	Yes	45	75	0.0107
Weak Domains	Yes	Yes	55	78	0.0115

Note. All features datasets included untransformed features. α = α hyperparameter.

Table 3

Example Items of Subjective Domains

Domain	Item
Subjective Control	How would you rate the amount of control you have over your life overall these days?
Subjective Physical Health	In general, compared to most men/women your age, would you say your health is…
Subjective Disengagement	When I experience a stressful event, I give up trying to reach my goal.
Subjective Sociability	I usually like to spend my leisure time with friends rather than alone.
Subjective Socioeconomic Status	I feel safe being out alone in my neighborhood at night.
Subjective Sex Life	How would you rate the sexual aspect of your life these days?
Subjective Societal Evaluation	The world is becoming a better place for everyone.
Subjective Social Dominance	Please indicate how well assertive describes you.
Subjective Prosociality	How much thought and effort do you put into your contribution to the welfare and well-being of other people these days?
Subjective Religiosity	How religious are you?

Table 4

Ridge Regression Results

Features	Squares	Products	R	α
Traits	No	No	.88	370
Traits	Yes	No	.89	630
Traits	No	Yes	.92	590
Traits	Yes	Yes	.92	600
Domains	No	No	.86	69
Domains	Yes	No	.87	120
Domains	No	Yes	.87	230
Domains	Yes	Yes	.87	240
Weak Traits	No	No	.50	510
Weak Traits	Yes	No	.52	600
Weak Traits	No	Yes	.60	730
Weak Traits	Yes	Yes	.61	870
Weak Domains	No	No	.49	51
Weak Domains	Yes	No	.51	160
Weak Domains	No	Yes	.51	350
Weak Domains	Yes	Yes	.52	330

Note. All features datasets included untransformed features. R = multiple R. α = α hyperparameter.

Table 5

Support Vector Regression Results

Features	R	C	ϵ	γ
Traits	.89	5.0	0.11	0.00064
Domains	.87	5.2	0.050	0.0024
Weak Traits	.53	6.4	0.51	0.0019
Weak Domains	.52	1.9	0.60	0.0082

R = multiple R. C = C hyperparameter. ϵ = ϵ hyperparameter. γ = γ hyperparameter.

Table 6

Random Forest Results

Features	R	Maximum Depth	Minimum Samples Per Leaf	Maximum Features	Maximum Leaf Nodes
Traits	.86	21	2	16	870
Domains	.85	18	3	10	620
Weak Traits	.47	20	3	19	430
Weak Domains	.50	23	2	5	230

R = multiple R.

Table 7

Artificial Neural Network Results

Features	R	Hidden Layer Size	α
Traits	.89	11	8.0
Domains	.87	10	2.4
Weak Traits	.51	19	7.0
Weak Domains	.52	52	2.7

R = multiple R. α = α hyperparameter.

Table 8

Trait Results

Trait	ω_t	*r*	Ridge Regression Coefficient	Random Forest Importance	Propensity Simple Coefficient	Propensity Multiple Coefficient
Subjective Environmental Mastery	.80	.74	.13	.14	.45	.12
Subjective Self Esteem	.81	.71	.15	.13	.32	.14
Subjective Purpose In Life	.79	.63	.06	.08	.20	.05
Subjective Control Over Life Rating		.57	.11	.06	.19	.06
Subjective Positive Relations With Others	.79	.57	.05	.05	.21	.01
Subjective Stress Reactivity	.74	-.56	-.08	.05	-.23	-.07
Subjective Optimism	.70	.54	.04	.03	.09	.03
Subjective Physical Health	.83	.50	.05	.03	.23	.05
Subjective Work Situation Evaluation	.63	.49	.04	.03	.13	.04
Subjective Financial Situation Evaluation	.79	.49	.04	.03	.13	.03
Objective Aches	.77	-.47	-.07	.03	-.17	-.07
Subjective Health Compared to Others Same Age	.73	.46	.03	.02	.16	.06
Subjective Sleep Issues	.82	-.45	-.05	.02	-.12	-.04
Subjective Standing In Community		.44	.03	.01	.04	.01
Subjective Personal Mastery	.75	.44	-.00	.01	-.01	.02
Subjective Control Over Work Situation		.43	.02	.01	-.01	.02
Subjective Health Control		.42	.02	.01	.05	.01
Subjective Home Quality	.81	.42	.01	.01	.06	-.02
Subjective Social Integration	.76	.42	.01	.01	.13	.03
Subjective Extraversion	.77	.41	.05	.01	.13	.05
Subjective Home Work Rewarding	.71	.41	.03	.01	.06	.03
Subjective Life Control		.41	.04	.01	.04	.02
Subjective Alienation	.62	-.41	.01	.00	.00	.04
Subjective Disappointed By Achievement		-.40	-.06	.01	-.06	-.04
Subjective Self Protection	.75	.39	.02	.01	.15	.03

Table 8 (continued)

Trait	ω_t	*r*	Ridge Regression Coefficient	Random Forest Importance	Propensity Simple Coefficient	Propensity Multiple Coefficient
Subjective Coping Positive Reinterpretation	.81	.37	.02	.00	.04	-.02
Subjective Autonomy	.72	.37	-.01	.01	-.03	-.04
Subjective Social Contribution	.71	.36	-.01	.00	.05	.01
Subjective Coping Venting Emotion	.82	-.35	-.03	.01	-.04	-.02
Subjective Life Effort Rating		.34	.02	.00	-.13	-.00
Subjective Intellectual Aging	.74	-.34	.01	.00	-.06	.01
Subjective Conscientiousness	.70	.33	-.01	.00	.04	-.01
Subjective Family Strain	.80	-.33	-.02	.01	-.01	-.01
Objective Short Breath	.79	-.33	-.02	.00	-.11	.00
Objective Daily Spiritual Experiences	.89	.33	.05	.01	.08	.05
Subjective Coping Active	.75	.33	-.01	.00	.03	-.03
Objective Mental Health Professionals	.55	-.33	-.07	.01	-.12	-.06
Subjective Coping Behavioral Disengagement	.75	-.33	.01	.00	-.06	.01
Subjective Daily Discrimination	.92	-.32	-.02	.00	-.01	-.00
Subjective Family Support	.85	.32	.01	.00	.01	.00
Subjective Coping Planning	.84	.31	-.00	.00	.17	.01
Subjective Social Closeness	.70	.31	.02	.00	.04	.02
Subjective Sex Life Evaluation	.79	.31	.04	.01	.09	.06
Subjective Health Compared To Five Years Ago	.85	.29	.04	.01	.06	.04
Subjective Control Contribution To Others Rating		.29	-.01	.00	-.07	-.02
Subjective Health Locus Of Control Self	.73	.29	-.01	.00	.05	-.01
Subjective Friend Support	.88	.28	.01	.00	.03	.00
Subjective Contribution To Others		.28	.02	.00	.03	.02
Objective Sleep Issues		-.28	-.02	.00	-.03	-.01
Subjective Social Coherence	.50	.27	-.00	.00	.04	.00

Table 8 (continued)

Trait	ω_t	*r*	Ridge Regression Coefficient	Random Forest Importance	Propensity Simple Coefficient	Propensity Multiple Coefficient
Subjective Openness To Experience	.77	.26	-.01	.00	.06	-.03
Subjective Social Potency	.73	.25	-.02	.00	.03	-.01
Objective Could Not Get Healthcare		-.24	-.02	.00	-.05	-.00
Subjective Selective Secondary Control	.63	.24	-.01	.00	.02	.00
Subjective Foresight		.24	-.01	.00	.04	-.01
Subjective Somatic Amplification	.55	-.23	-.03	.01	-.04	-.03
Subjective Compensatory Primary Control	.75	.23	.01	.00	.02	.00
Objective Digestion Issues	.49	-.22	-.01	.00	-.03	.01
Subjective Agreeableness	.80	.22	-.03	.00	.01	-.04
Subjective Friend Strain	.81	-.21	-.01	.00	.02	-.01
Objective Alcohol Issues	.80	-.20	-.02	.00	-.12	-.01
Subjective Religious Coping	.83	.20	.01	.00	.08	.02
Objective Neighborhood Contact	.73	.20	.01	.00	.01	.00
Subjective Live For Today	.65	-.19	.01	.00	-.02	.01
Objective Married		.19	.04	.00	.03	.04
Objective Age		.17	.04	.00	.15	.05
Objective Health Procedures	.69	-.17	-.01	.00	-.05	-.01
Objective Sleep Time	.66	-.17	-.02	.00	.02	-.01
Objective Drug Use	.68	-.16	-.02	.00	-.05	-.01
Objective Others Alcohol Use	.18	-.16	-.01	.00	-.02	-.01
Objective Brain Issues	.47	-.16	.02	.00	-.03	.02
Objective Receiving Money	.39	-.16	-.01	.00	-.04	-.02
Subjective Health Locus Of Control Others	.11	-.15	.02	.00	.05	.03
Objective Religious Activity	.87	.15	.01	.00	.02	.01
Objective Drug Problem		-.14	-.02	.00	-.06	-.00

Table 8 (continued)

Trait	ω_t	*r*	Ridge Regression Coefficient	Random Forest Importance	Propensity Simple Coefficient	Propensity Multiple Coefficient
Objective Heart Issues	.74	-.14	.01	.00	-.02	.00
Subjective Mental Stimulation	.51	.14	-.02	.00	-.00	-.02
Subjective Weight Perception		-.13	.01	.00	.01	.02
Objective Time Volunteering	.40	.13	-.01	.00	.04	-.01
Objective Highest Education Completed		.13	-.01	.00	.00	-.01
Objective Tobacco Use	.58	-.13	-.02	.00	-.02	-.02
Subjective Financial Situation Effort		.13	-.02	.00	-.04	-.01
Objective Healthcare Place Doctors Office		.13	.01	.00	-.01	.01
Objective Others Tobacco Use	.31	-.12	-.01	.00	-.01	-.00
Subjective Important To Help People		.12	-.01	.00	-.02	-.01
Subjective Control	.66	.12	-.01	.00	.00	-.01
Subjective Religious Mindfulness	.95	.12	-.01	.00	.04	-.01
Objective Grandparent		.11	.01	.00	.04	.01
Objective Mental Health Groups	.81	-.10	.01	.00	-.03	.01
Objective Heterosexual		.10	.01	.00	.05	.01
Objective Years In State		.09	.02	.00	.01	.02
Subjective Spirituality	.92	.09	-.02	.00	-.03	-.03
Subjective Insight Into Past	.39	.09	-.02	.00	-.06	-.02
Objective Number Children		.09	.00	.00	.01	-.01
Objective Hips Size		-.08	.01	.00	.06	.02
Objective Medical Exams	.43	.08	.01	.00	.02	-.00
Subjective Explore Different Religions		-.07	-.01	.00	-.01	-.01
Objective Caregiving		-.07	-.02	.00	.00	-.02
Objective Receiving Physical Assistance	.56	-.07	.01	.00	.00	.01
Objective Lived In Institution		-.06	.01	.00	.00	.01

Table 8 (continued)

Trait	ω_t	*r*	Ridge Regression Coefficient	Random Forest Importance	Propensity Simple Coefficient	Propensity Multiple Coefficient
Objective Diabetes		-.05	.02	.00	.03	.02
Objective Healthcare Professional Generalist		.05	.01	.00	.00	.01
Subjective Traditionalism	.60	.04	.01	.00	.01	.00
Subjective Sympathy	.53	.04	-.01	.00	-.02	-.01
Objective White		.03	.01	.00	.00	.01
Objective Currently Pregnant		.03	.02	.00	.04	-.00

Note. ω_t = omega total. *r* = correlation. ω_t values are not displayed for traits that consisted of one feature, as ω_t cannot be calculated in that case. Propensity Simple Coefficient = regression coefficient in a model where only the feature is predicting subjective well-being using inverse probability of treatment weighting. Propensity Multiple Coefficient = regression coefficient in a model where all trait features predict subjective well-being using inverse probability of treatment weighting.

Table 9

Domain Results

Domain	r	Ridge Regression Coefficient	Random Forest Importance	Propensity Simple Coefficient	Propensity Multiple Coefficient
Subjective Control	.68	.36	.33	.43	.35
Subjective Physical Health	.64	.21	.20	.44	.23
Subjective Disengagement	-.55	-.22	.14	-.23	-.19
Subjective Sociability	.47	.19	.09	.26	.19
Objective Physical Health Issues	-.40	-.07	.03	-.18	-.05
Subjective Socioeconomic Status	.32	.03	.02	.12	.02
Subjective Sex Life	.31	.09	.02	.14	.11
Subjective Societal Evaluation	.28	.05	.02	.10	.05
Objective Financial Situation	.26	.01	.01	.07	.01
Objective Religious Activity	.25	.05	.02	.08	.06
Objective Human Contact	.24	.01	.01	.04	.02
Subjective Social Dominance	.22	.07	.01	.10	.07
Objective Drug Use	-.22	-.07	.01	-.09	-.05
Subjective Prosociality	.21	-.08	.01	-.02	-.07
Objective Marital Status	.19	.06	.00	.04	.06
Subjective Religiosity	.18	.04	.01	.03	.02
Objective Age	.17	.08	.01	.14	.08
Objective Others Drug Use	-.17	-.02	.01	-.00	-.01
Objective Sleep	-.17	-.04	.01	-.01	-.03
Objective Health Treatment	-.16	-.05	.01	-.06	-.03
Objective Education	.13	-.01	.00	.01	-.01
Objective Physical Size	-.12	.02	.01	.03	.03
Objective Physical Activity	.10	.01	.01	-.00	.01
Objective Female Status	.05	.00	.00	-.02	-.01

Note. r = Pearson's correlation. Propensity Simple Coefficient = regression coefficient in a model where only the feature is predicting subjective well-being using inverse probability of treatment weighting. Propensity Multiple Coefficient = regression coefficient in a model where all domain features predict subjective well-being using inverse probability of treatment weighting.

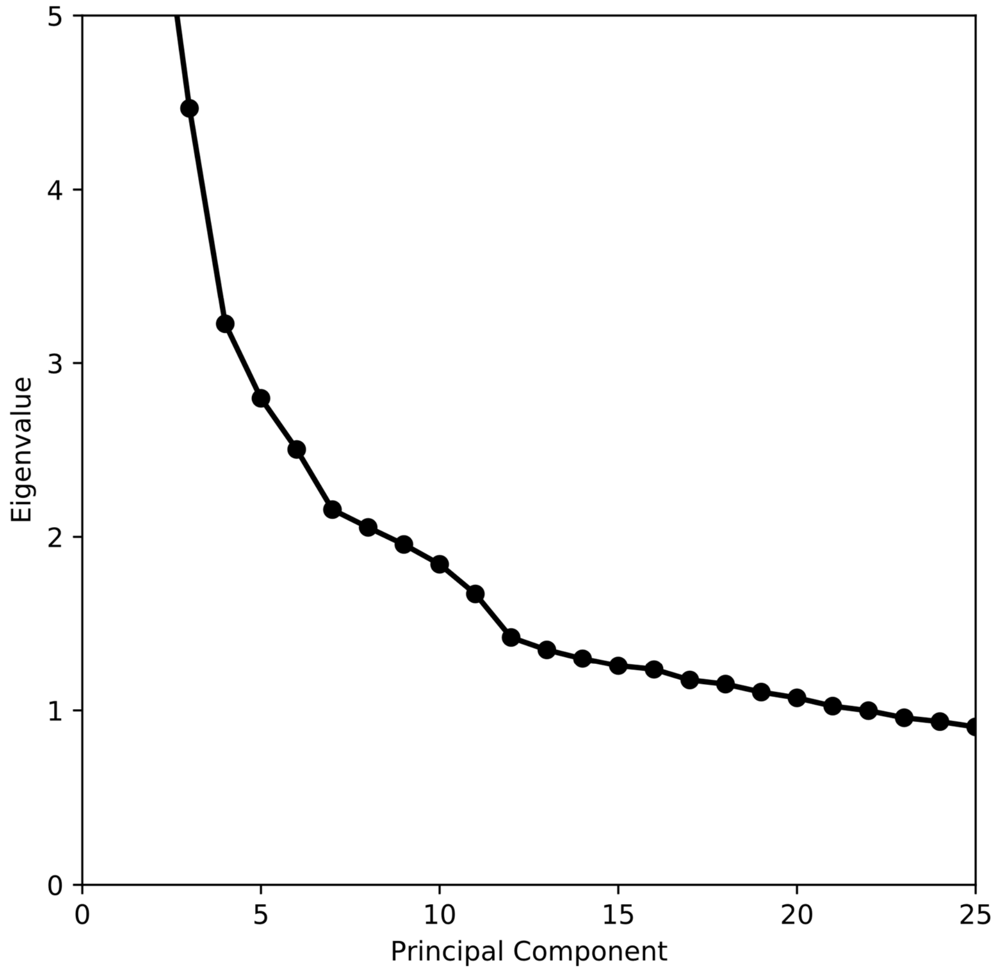

Figure 1. Scree plot of subjective traits. Eigenvalues of components 1 and 2 are greater than 5.0.