A transfer to utility study using the R package TTU

Alejandra Scienceace[1,*] Fionn Researchchamp[2]

Abstract

Background: Background text.

Objectives: We aimed to: Add study objectives here.

Methods: Add abstract here.

Results: Add results here.

Conclusions: Add conclusion here

Data: Add details on dataset.

[1] Awesome University, Somewhere, Earth
[2] August Institute, Elsewhere, Earth

[*] Correspondence: Alejandra Scienceace <fake@email.com>

1 Introduction

This article is an automatically generated scientific summary of a study that the authors implemented using the R package TTU [1].

Our study aimed to identify the best TTU regression models to predict Assessment of Quality of Life - Six Dimension (AQoL-6D) utility and evaluate the predictive ability of two candidate measures of depression and psychological distress.

2 Methods

2.1 Sample and setting

2.2 Measures

Data was collected on utility weights, two candidate predictors of utility weights and descriptive population characteristics.

2.2.1 Utility weights

Utility weights were assessed using the AQoL-6D multi-attribute utility instrument.

2.2.2 Candidate predictors

Data from two measures of depression (one measure) and psychological distress (one measure) were used as candidate predictors to construct TTU models.

Depression was measured by Patient Health Questionnaire (PHQ-9 - measured on a scale of 0-27). Psychological distress was measured by Kessler Psychological Distress Scale (6 Dimension) (K6 - measured on a scale of 0-24).

2.2.3 Population characteristics

Population characteristic data were age, relationship status, education and employment status, primary diagnosis and clinical stage.

2.3 Statistical analysis

We implemented the generalised form of the study analysis algorithm developed by Hamilton, Gao and colleagues [2], the key steps of which are described as follows.

2.3.1 Descriptive statistics

Basic descriptive statistics were used to characterise the cohort in terms of baseline population variables. Pearson's Product Moment Correlations (r) were used to determine the relationships between candidate predictors and the AQoL-6D utility score.

2.3.2 TTU regression models

We compared predictive performance of a range of models predicting AQoL-6D utility scores using the candidate predictor that had the highest Pearson correlation coefficient with utility scores. The models compared include ordinary least squares (OLS) regression models and generalised linear models (GLMs). OLS regression models used no transformation, complementary log log transformation, logit transformation, log log transformation and log transformation. GLMs used gaussian distribution and log link, beta distribution and complementary log log link and beta distribution and logit link. Ten-fold cross-validation was used to compare model fitting using training datasets and predictive ability using testing datasets using three indicators including R^2, root mean square error (RMSE) and mean absolute error (MAE) [3,4].

To evaluate whether candidate predictors could independently predict utility scores, we established multivariate prediction models using baseline data with the candidate predictor and demographic, functioning and clinical covariates. Demographic covariates were age, relationship status and education and employment status. The functioning covariate was social and occupational functioning assessment scale. Clinical covariates were clinical stage and primary diagnosis.

2.3.3 Candidate predictor comparison

We compared the usefulness of the candidate predictors by using a random forest model including both candidate predictors and by evaluating the independent predictive ability of different candidate predictors using 10-fold cross-validation.

2.3.4 Longitudinal transfer to utility models

We next established longitudinal models using linear mixed effect models (LMMs) and generalised linear mixed effect models (GLMMs) that included both the baseline and follow-up data. Model fitting was evaluated using Bayesian R^2 [5].

2.3.5 Secondary analyses

One secondary analysis was undertaken. The secondary analysis used Social and Occupational Functioning Assessment Scale as a predictor.

2.3.6 Software

We undertook all our analyses using ***R*** 4.0.5 [6] using the TTU package [1].

Table 1: Participant characteristics

		Baseline		Follow-Up	
		(N =	1068)	(N =	643)
Age	Mean (SD)	17.56	(3.09)	17.77	(3.09)
	Median (Q1 Q3)	17.00	(15.00 20.00)	18.00	(16.00 20.00)
	Min - Max	12.00	25.00	12.00	25.00
	Missing	0.00		0.00	
Relationship Status	In a relationship	317.00	(29.68%)	190.00	(29.55%)
	Not in a relationship	751.00	(70.32%)	453.00	(70.45%)
	Missing	0.00		0.00	
Education and Employment Status	Not studying or working	159.00	(15.35%)	152.00	(24.40%)
	Studying and working	305.00	(29.44%)	146.00	(23.43%)
	Studying only	405.00	(39.09%)	167.00	(26.81%)
	Working only	167.00	(16.12%)	158.00	(25.36%)
	Missing	32.00		20.00	
Primary Diagnosis	Anxiety	264.00	(26.01%)	175.00	(29.02%)
	Depression	182.00	(17.93%)	140.00	(23.22%)
	Depression and Anxiety	332.00	(32.71%)	152.00	(25.21%)
	Other	237.00	(23.35%)	136.00	(22.55%)
	Missing	53.00		40.00	
Clinical Stage	0-1a	625.00	(60.27%)	249.00	(39.78%)
	1b	326.00	(31.44%)	216.00	(34.50%)
	2-4	86.00	(8.29%)	161.00	(25.72%)
	Missing	31.00		17.00	

3 Results

3.1 Cohort characteristics

Participants characteristics at baseline and follow-up are displayed in Table 1. This study included all 1068 participants with complete AQoL-6D data.

There were 643 participants (60.2%) who completed AQoL-6D questions at the follow-up survey three months after baseline assessment.

3.2 AQoL-6D and candidate predictors

Distribution of AQoL-6D total utility score and sub-domain scores are displayed in Figure 1. The mean utility score at baseline is 0.61 (SD = 0.19) and is 0.69 (SD = 0.18) at follow-up. Distribution of candidate predictors, PHQ-9 and K6, are summarised in Table 2. PHQ-9 was found to have the highest correlation with utility score both at baseline and follow-up followed by K6.

Table 2: Candidate predictors distribution parameters and correlations with AQoL-6D utility

		Baseline		Follow-Up		
		(N =	1068)	(N =	643)	p
Kessler Psychological Distress Scale (6 Dimension) (0-24)	Mean (SD)	12.08	(5.60)	10.10	(5.66)	0.00
	Missing	1.00		2.00		0.00
	Correlation with AQOL-6D	-0.67		-0.69		0.00, 0.00
Patient Health Questionnaire (0-27)	Mean (SD)	12.65	(6.23)	9.74	(6.21)	0.00
	Missing	4.00		2.00		0.00
	Correlation with AQOL-6D	-0.78		-0.80		0.00, 0.00

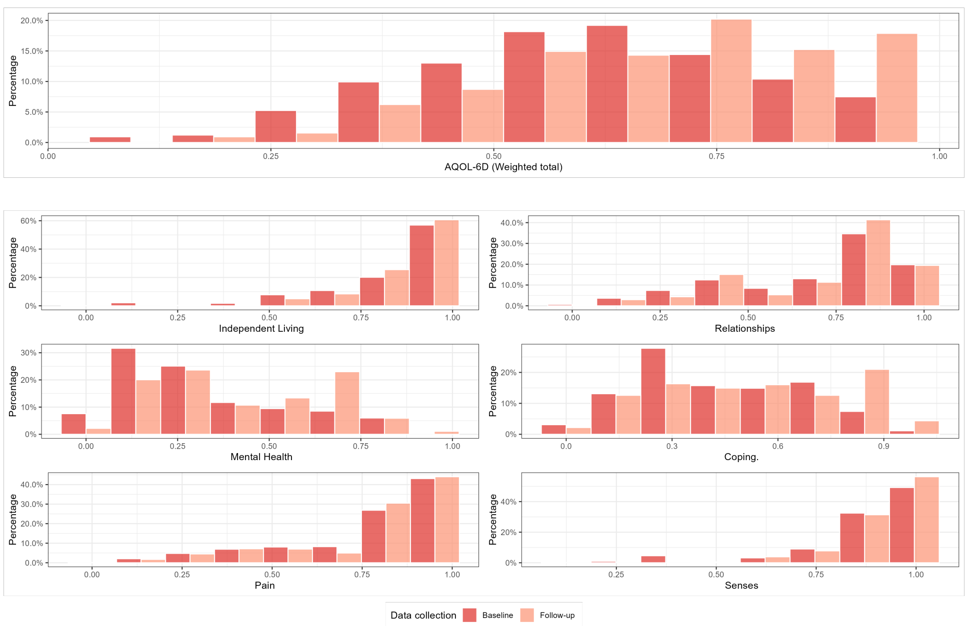

Figure 1: Distribution of AQoL-6D domains

3.3 TTU regression model performance

The 10-fold cross-validated model fitting index from TTU models using PHQ-9 are reported in Table A.1 in the Supplementary Material. The best OLS model was found to be either no transformation, log transformation or clog-log transformation. Both GLM with Gaussian distribution and log link and OLS with clog-log transformation were selected for further evaluation. Predictive ability of each candidate predictor using baseline data were also compared using 10-fold cross-validation.

Table A.2 illustrates that PHQ-9 had the highest predictive ability followed by K6. This is consistent with the random forest model in which PHQ-9 was found to be the most ‘important’ predictor (see Figure A.1). The confounding effect of other participant characteristics when using the candidate predictors in predicting utility score were also evaluated. Using the baseline data, SOFAS was found to independently predict utility scores in models for both candidate predictors *($p<0.01$)*.

3.4 Longitudinal TTU regression models

Regression coefficients of the baseline score and score changes (from baseline to follow-up) estimated in individual GLMM and LLM models are summarised in Table 3. Bayesian R^2 from each model is reported. Modelled residual standard deviations (SDs) were also provided to support simulation studies which need to capture individual level variation. In both GLMM and LLM models, the prediction models using PHQ-9 had the highest R^2 (0.76 and 0.82) and lowest estimated residual SD. R^2 were above 0.8 for all LLM models and above 0.6 for all GLMM models. Variance of the random intercept was comparable with the residual variance.

Table 3: Estimated coefficients from longitudinal TTU models

	GLMM with Gaussian distribution and log link					LMM with clog-log transformation				
Parameter	Estimate	SE	95CI	R2	Sigma	Estimate	SE	95CI	R2	Sigma
PHQ9 model				0.76	0.09				0.82	0.26
SD (Intercept)	0.11	0.01	0.10, 0.13			0.26	0.01	0.24, 0.28		
Intercept	-0.03	0.01	-0.05, -0.01			0.94	0.02	0.89, 0.98		
PHQ9 baseline	-3.81	0.09	-3.98, -3.65			-7.65	0.17	-7.98, -7.32		
PHQ9 change	-2.45	0.12	-2.68, -2.22			-4.74	0.21	-5.16, -4.32		
K6 model				0.74	0.10				0.79	0.28
SD (Intercept)	0.16	0.01	0.15, 0.18			0.34	0.01	0.32, 0.36		
Intercept	-0.05	0.01	-0.08, -0.03			0.89	0.03	0.83, 0.95		
K6 baseline	-3.68	0.12	-3.91, -3.46			-7.45	0.23	-7.90, -6.99		
K6 change	-1.94	0.12	-2.19, -1.70			-3.85	0.22	-4.30, -3.42		

The mean ratio between two coefficients ($\beta_{change}/\beta_{baseline}$) is 0.63 for depression measurements, 0.52 for K6 and between Inf and -Inf for SOFAS.

Distribution of observed and predicted utility scores and their association from GLMM (Gaussian distribution and log link) and LLM (complementary log log transformation) using PHQ-9 are plotted in Figure 2.

We also evaluated models with SOFAS at baseline and SOFAS change from baseline added to depression, psychological distress and functioning predictors (see Tables A.3 and A.4).

Detailed summaries of all models are available in the online data repository (see "Availability of data and materials").

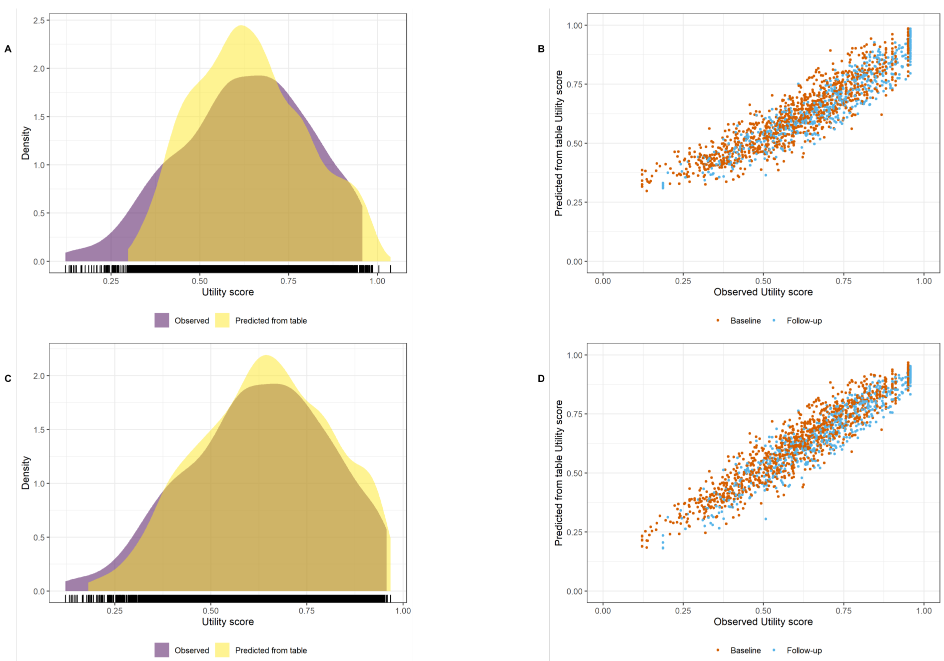

Figure 2: Comparison of observed and predicted AQoL-6D utility score from longitudinal TTU model using PHQ-9 (A) Density plots of observed and predicted utility scores (GLMM (Gaussian distribution and log link)) (B) Scatter plots of observed and predicted utility scores by timepoint (GLMM (Gaussian distribution and log link)) (C) Density plots of observed and predicted utility scores (LLM(complementary log log transformation)) (D) Scatter plots of observed and predicted utility scores by timepoint (LLM (complementary log log transformation))

4 Discussion

Add discussion text here.

5 Conclusions

Conclusion text goes here.

Availability of data and materials

None available

Ethics approval

Add ethics text here.

Funding

Add funding details.

Conflict of Interest

None declared.

References

1. Gao C, Hamilton M. TTU: Transfer to utility mapping algorithm toolkit. 2021.

2. Hamilton MP, Gao CX, Filia KM, Menssink JM, Sharmin S, Telford N, et al. Predicting quality adjusted life years in young people attending primary mental health services. medRxiv. Cold Spring Harbor Laboratory Press; 2021; doi:10.1101/2021.07.07.21260129

3. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction. Springer Science & Business Media; 2009.

4. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai. Montreal, Canada; pp. 1137–1145.

5. Gelman A, Goodrich B, Gabry J, Vehtari A. R-squared for bayesian regression models. The American Statistician. 2019;73: 307–309. doi:10.1080/00031305.2018.1549100

6. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available: https://www.R-project.org/

A Supplementary Material

A.1 Additional tables

Table A.1: 10-fold cross-validated model fitting index for different OLS or GLM models for using PHQ9 total scores as predictor with the baseline data

	Training model fit (averaged over 10 folds)			Testing model fit (averaged over 10 folds)		
Model	R2	RMSE	MAE	R2	RMSE	MAE
OLS						
No transformation	0.60	0.12	0.09	0.60	0.12	0.09
Complementary Log Log transformation	0.60	0.12	0.09	0.60	0.12	0.10
Logit transformation	0.59	0.12	0.10	0.59	0.12	0.10
Log Log transformation	0.57	0.12	0.10	0.57	0.12	0.10
Log transformation	0.57	0.12	0.10	0.57	0.12	0.10
GLM						
Gaussian distribution and log link	0.59	0.12	0.10	0.59	0.12	0.10
Beta distribution and complementary log log link	0.60	0.12	0.09	0.60	0.12	0.10
Beta distribution and logit link	0.60	0.12	0.10	0.60	0.12	0.10

* RMSE: Root Mean Squared Error; MAE: Mean Absolute Error

Table A.2: 10-fold cross-validated model fitting index for different candidate predictors estimated using GLM with Gaussian distribution and log link with the baseline data

	Training model fit (averaged over 10 folds)			Testing model fit (averaged over 10 folds)		
Model	R2	RMSE	MAE	R2	RMSE	MAE
PHQ9	0.59	0.12	0.10	0.59	0.12	0.10
K6	0.44	0.14	0.11	0.44	0.14	0.11

* RMSE: Root Mean Squared Error; MAE: Mean Absolute Error

Table A.3: Estimated coefficients from longitudinal TTU models based on candidate predictors and SOFAS score using LLM (with cloglog transformation)

Parameter*	Estimate	SE	95CI	R2	Sigma
PHQ9 SOFAS model				0.84	0.24
SD (Intercept)	0.24	0.01	0.22, 0.26		
Intercept	-0.16	0.08	-0.32, 0.00		
PHQ9 baseline	-6.72	0.17	-7.05, -6.39		
PHQ9 change	-4.19	0.19	-4.58, -3.82		
SOFAS baseline	1.48	0.11	1.27, 1.70		
SOFAS change	1.05	0.10	0.86, 1.24		
K6 SOFAS model				0.81	0.26
SD (Intercept)	0.30	0.01	0.28, 0.32		
Intercept	-0.61	0.10	-0.80, -0.41		
K6 baseline	-6.24	0.22	-6.66, -5.81		
K6 change	-3.39	0.21	-3.81, -2.97		
SOFAS baseline	2.05	0.13	1.79, 2.31		
SOFAS change	1.24	0.10	1.03, 1.44		

* Calculated as original scores divided by 100

Table A.4: Estimated coefficients from longitudinal TTU models based on individual candidate predictors and SOFAS score using GLM (Gaussian distribution with log link)

Parameter*	Estimate	SE	95CI	R2	Sigma
PHQ9 SOFAS model				0.78	0.09
SD (Intercept)	0.10	0.01	0.09, 0.12		
Intercept	-0.58	0.04	-0.66, -0.49		
PHQ9 baseline	-3.36	0.09	-3.52, -3.19		
PHQ9 change	-2.19	0.11	-2.41, -1.98		
SOFAS baseline	0.74	0.06	0.63, 0.86		
SOFAS change	0.51	0.05	0.41, 0.61		
K6 SOFAS model				0.76	0.09
SD (Intercept)	0.14	0.01	0.12, 0.15		
Intercept	-0.81	0.05	-0.90, -0.71		
K6 baseline	-3.08	0.11	-3.30, -2.87		
K6 change	-1.71	0.12	-1.94, -1.48		
SOFAS baseline	1.03	0.06	0.90, 1.16		
SOFAS change	0.60	0.06	0.48, 0.71		

* Calculated as original scores divided by 100

Table A.5: R Packages used in data analysis and reporting

Package	Version	Citation
arsenal	3.6.3	Ethan Heinzen, Jason Sinnwell, Elizabeth Atkinson, Tina Gunderson and Gregory Dougherty (2021). arsenal: An Arsenal of 'R' Functions for Large-Scale Statistical Summaries. R package version 3.6.3. https://CRAN.R-project.org/package=arsenal
assertthat	0.2.1	Hadley Wickham (2019). assertthat: Easy Pre and Post Assertions. R package version 0.2.1. https://CRAN.R-project.org/package=assertthat
BCEA	2.3-1.1	Baio et al (2017). Bayesian Cost Effectiveness Analysis with the R package BCEA. Springer, New York, NY. doi: 10.1007/978-3-319-55718-2, URL: http://www.springer.com/us/book/9783319557168/.
betareg	3.1-4	Cribari-Neto F, Zeileis A (2010). "Beta Regression in R." _Journalof Statistical Software_, *34*(2), 1-24. doi: 10.18637/jss.v034.i02(URL: https://doi.org/10.18637/jss.v034.i02).
boot	1.3-28	Angelo Canty and Brian Ripley (2021). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-28.
Boruta	7.0.0	Miron B. Kursa, Witold R. Rudnicki (2010). Feature Selection with the Boruta Package. Journal of Statistical Software, 36(11), 1-13. URL http://www.jstatsoft.org/v36/i11/.
brms	2.15.0	Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1-28. doi:10.18637/jss.v080.i01
caret	6.0-88	Max Kuhn (2021). caret: Classification and Regression Training. R package version 6.0-88. https://CRAN.R-project.org/package=caret
cmdstanr	0.4.0.9000	Jonah Gabry and Rok Cesnovar (2021). cmdstanr: R Interface to 'CmdStan'. https://mc-stan.org/cmdstanr, https://discourse.mc-stan.org.
cowplot	1.1.1	Claus O. Wilke (2020). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1. https://CRAN.R-project.org/package=cowplot
dataverse	0.3.8.9000	Will Beasley, Shiro Kuriwaki, Thomas J. Leeper et al. (). dataverse: R Client for Dataverse 4+ Repositories. R package version 0.3.8.9000.
dplyr	1.0.7	Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.7. https://CRAN.R-project.org/package=dplyr

enrichwith	0.3.1	https://github.com/ikosmidis/enrichwith>.
eq5d	0.9.0	Fraser Morton and Jagtar Singh Nijjar (2021). eq5d: Methods for Analysing 'EQ-5D' Data and Calculating 'EQ-5D' Index Scores. R package version 0.9.0. https://CRAN.R-project.org/package=eq5d
faux	1.0.0	Lisa DeBruine, (2021). faux: Simulation for Factorial Designs R package version 1.0.0. Zenodo. http://doi.org/10.5281/zenodo.2669586
ggalt	0.4.0	Bob Rudis, Ben Bolker and Jan Schulz (2017). ggalt: Extra Coordinate Systems, 'Geoms', Statistical Transformations, Scales and Fonts for 'ggplot2'. R package version 0.4.0. https://CRAN.R-project.org/package=ggalt
ggfortify	0.4.12	Yuan Tang, Masaaki Horikoshi, and Wenxuan Li. "ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages." The R Journal 8.2 (2016): 478-489.
ggplot2	3.3.5	H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
ggpubr	0.4.0	Alboukadel Kassambara (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr
gridExtra	2.3	Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra
here	1.0.1	Kirill Müller (2020). here: A Simpler Way to Find Your Files. R package version 1.0.1. https://CRAN.R-project.org/package=here
Hmisc	4.5-0	Frank E Harrell Jr, with contributions from Charles Dupont and many others. (2021). Hmisc: Harrell Miscellaneous. R package version 4.5-0. https://CRAN.R-project.org/package=Hmisc
hutils	1.6.0	Hugh Parsonage (2020). hutils: Miscellaneous R Functions and Aliases. R package version 1.6.0. https://CRAN.R-project.org/package=hutils
knitr	1.33	Yihui Xie (2021). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.33.
knitrBootstrap	1.0.2	Jim Hester (2018). knitrBootstrap: 'knitr' Bootstrap Framework. R package version 1.0.2. https://CRAN.R-project.org/package=knitrBootstrap
lifecycle	1.0.0	Lionel Henry and Hadley Wickham (2021). lifecycle: Manage the Life Cycle of your Package Functions. R package version 1.0.0. https://CRAN.R-project.org/package=lifecycle

lubridate	1.7.10	Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL https://www.jstatsoft.org/v40/i03/.
magrittr	2.0.1	Stefan Milton Bache and Hadley Wickham (2020). magrittr: A Forward-Pipe Operator for R. R package version 2.0.1. https://CRAN.R-project.org/package=magrittr
MASS	7.3-54	Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
MatchIt	4.2.0	Daniel E. Ho, Kosuke Imai, Gary King, Elizabeth A. Stuart (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software, Vol. 42, No. 8, pp. 1-28. URL https://www.jstatsoft.org/v42/i08/
Matrix	1.3-4	Douglas Bates and Martin Maechler (2021). Matrix: Sparse and Dense Matrix Classes and Methods. R package version 1.3-4. https://CRAN.R-project.org/package=Matrix
matrixcalc	1.0-4	Frederick Novomestky (2021). matrixcalc: Collection of Functions for Matrix Calculations. R package version 1.0-4. https://CRAN.R-project.org/package=matrixcalc
methods	4.0.5	R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
mice	3.13.0	Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. URL https://www.jstatsoft.org/v45/i03/.
pacman	0.5.1	Rinker, T. W. & Kurkiewicz, D. (2017). pacman: Package Management for R. version 0.5.0. Buffalo, New York. http://github.com/trinker/pacman
psych	2.1.6	Revelle, W. (2021) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 2.1.6,.
purrr	0.3.4	Lionel Henry and Hadley Wickham (2020). purrr: Functional Programming Tools. R package version 0.3.4. https://CRAN.R-project.org/package=purrr
randomForest	4.6-14	A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22.

readr	1.4.0	Hadley Wickham and Jim Hester (2020). readr: Read Rectangular Text Data. R package version 1.4.0. https://CRAN.R-project.org/package=readr
ready4class	0.0.0.9199	Matthew Hamilton and Glen Wiesner (2021). ready4class: Standardised Developer Tools for Creating and Extending Classes for Use as Part of the Ready4 Suite. https://ready4-dev.github.io/ready4class/, https://github.com/ready4-dev/ready4class, https://www.ready4-dev.com/.
ready4fun	0.0.0.9298	Matthew Hamilton and Glen Wiesner (2021). ready4fun: Standardised Function Authoring and Documentation Tools for Use with the Ready4 Suite. https://ready4-dev.github.io/ready4fun/, https://github.com/ready4-dev/ready4fun, https://www.ready4-dev.com/.
ready4show	0.0.0.9035	Matthew Hamilton and Glen Wiesner (2021). ready4show: Standardised Developer Tools for Sharing Insights from Projects Developed with the Ready4 Suite. https://ready4-dev.github.io/ready4show/, https://github.com/ready4-dev/ready4show, https://www.ready4-dev.com/.
ready4use	0.0.0.9133	Matthew Hamilton and Glen Wiesner (2021). ready4use: Standardised Developer Tools for Retrieving and Managing Data in Projects Developed with the Ready4 Suite. https://ready4-dev.github.io/ready4use/, https://github.com/ready4-dev/ready4use, https://ready4-dev.github.io/ready4/.
rlang	0.4.11	Lionel Henry and Hadley Wickham (2021). rlang: Functions for Base Types and Core R and 'Tidyverse' Features. R package version 0.4.11. https://CRAN.R-project.org/package=rlang
rmarkdown	2.9	JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2021). rmarkdown: Dynamic Documents for R. R package version 2.9. URL https://rmarkdown.rstudio.com.
scales	1.1.1	Hadley Wickham and Dana Seidel (2020). scales: Scale Functions for Visualization. R package version 1.1.1. https://CRAN.R-project.org/package=scales

simstudy	0.2.1	Keith Goldfeld and Jacob Wujciak-Jens (2020). simstudy: Simulation of Study Data. R package version 0.2.1. https://CRAN.R-project.org/package=simstudy
stats	4.0.5	R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
stringi	1.6.2	https://stringi.gagolewski.com/>.
stringr	1.4.0	Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr
Surrogate	1.9	Wim Van der Elst, Paul Meyvisch, Alvaro Florez Poveda, Ariel Alonso, Hannah M. Ensor and Christopher J. Weir & Geert Molenberghs (2021). Surrogate: Evaluation of Surrogate Endpoints in Clinical Trials. R package version 1.9. https://CRAN.R-project.org/package=Surrogate
synthpop	1.6-0	Beata Nowok, Gillian M. Raab, Chris Dibben (2016). synthpop: Bespoke Creation of Synthetic Data in R. Journal of Statistical Software, 74(11), 1-26. doi:10.18637/jss.v074.i11
testthat	3.0.4	Hadley Wickham. testthat: Get Started with Testing. The R Journal, vol. 3, no. 1, pp. 5–10, 2011
tibble	3.1.2	Kirill Müller and Hadley Wickham (2021). tibble: Simple Data Frames. R package version 3.1.2. https://CRAN.R-project.org/package=tibble
tidyr	1.1.3	Hadley Wickham (2021). tidyr: Tidy Messy Data. R package version 1.1.3. https://CRAN.R-project.org/package=tidyr
tidyselect	1.1.1	Lionel Henry and Hadley Wickham (2021). tidyselect: Select from a Set of Strings. R package version 1.1.1. https://CRAN.R-project.org/package=tidyselect
truncnorm	1.0-8	Olaf Mersmann, Heike Trautmann, Detlef Steuer and Björn Bornkamp (2018). truncnorm: Truncated Normal Distribution. R package version 1.0-8. https://CRAN.R-project.org/package=truncnorm
TTU	0.0.0.9285	Caroline Gao and Matthew Hamilton (2021). TTU: Transfer to Utility Mapping Algorithm Toolkit. https://ready4-dev.github.io/TTU/, https://github.com/ready4-dev/TTU, https://ready4-dev.github.io/ready4/.

utils	4.0.5	R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
viridis	0.6.1	Simon Garnier, Noam Ross, Robert Rudis, Antônio P. Camargo, Marco Sciaini, and Cédric Scherer (2021). Rvision - Colorblind-Friendly Color Maps for R. R package version 0.6.1.
youthu	0.0.0.9094	Matthew Hamilton and Caroline Gao (2021). youthu: Youth Outcomes to Health Utility. https://ready4-dev.github.io/youthu/, https://github.com/ready4-dev/youthu, https://www.ready4-dev.com/.
youthvars	0.0.0.9058	Matthew Hamilton and Caroline Gao (2021). youthvars: Youth Mental Health Variables Modelling Toolkit. https://ready4-dev.github.io/youthvars/, https://github.com/ready4-dev/youthvars, https://ready4-dev.github.io/ready4/.

A.2 Additional figures

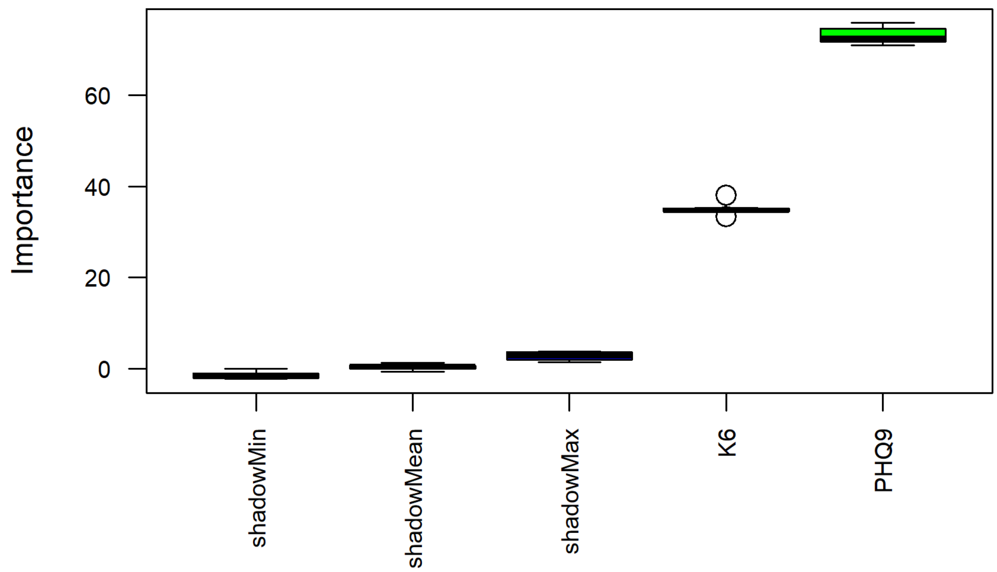

Figure A.1: Variable importance estimated using random forest