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 is entirely fictional and has been created to illustrate TTU package functionality.

The objective of the study was 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

The study sample is fake data that pretends to be young people aged 12 to 25 years who attended Australian primary care services for mental health related needs between November 2019 to August 2020.

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

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 summarised 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 with training datasets and predictive ability with 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 Assessing ability of baseline measures to predict change

We assessed the potential of our evaluated predictors to predict change in AQoL-6D when using only baseline measures. This goal can be accomplished by comparing longitudinal TTU model coefficients of each

predictor's $\beta_{baseline}$ (representing between person variation) and β_{change} (representing within person variation) parameters [2]. We calculated the $\beta_{change}/\beta_{baseline}$ ratio for each evaluated single predictor longitudinal TTU model, reporting the mean ratio across all models for each predictor.

2.3.6 Secondary analyses

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

2.3.7 Software

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

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. The 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.
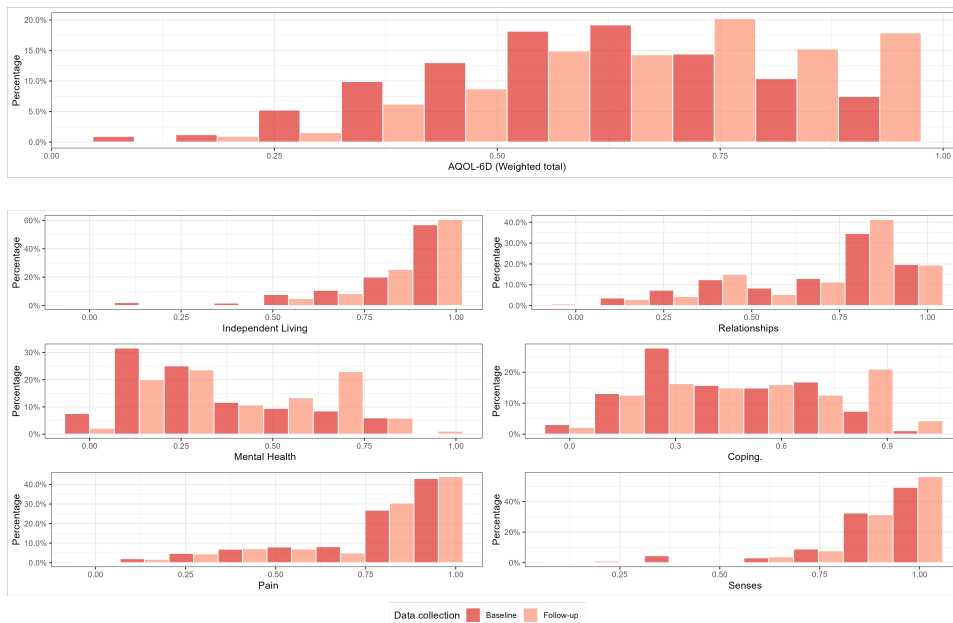

Figure 1: Distribution of AQoL-6D domains

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	

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

		Baseline		Follow-Up		p
		(N =	1068)	(N =	643)	
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

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. Both GLM with Gaussian distribution and log link and OLS with complementary log 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 and modelled residual standard deviations (SDs) from each model are reported. In both GLMM and LLM models, the prediction models using PHQ-9 had the highest R^2 (0.76 and 0.82). R^2 was between 0.74 and 0.76 for all GLMMs and between 0.79 and 0.82 for all LLMs.

Table 3: Estimated coefficients from longitudinal TTU models

| Parameter | GLMM - Gaussian distribution and log link | | | | | LLM - complementary log log transformation | | | | |
	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		

Note: The K6 and PHQ9 parameters were first multiplied by 0.01.

The mean ratio between the within-person and between-person associated coefficients ($\beta_{change}/\beta_{baseline}$) is 0.63 for depression measurements, 0.52 for K6 and 0.45 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").
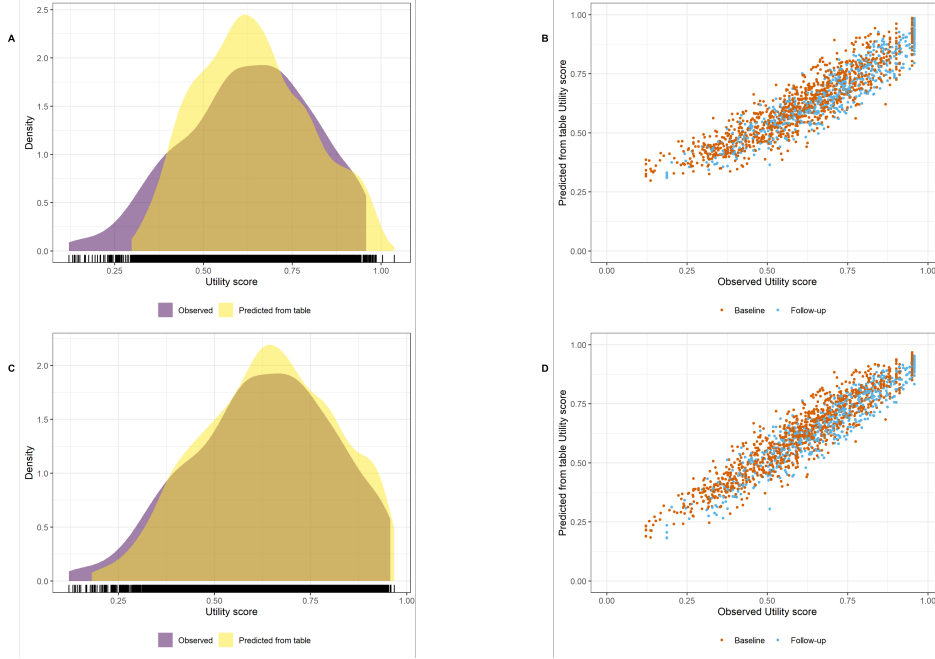

Figure 2: Comparison of observed and predicted AQoL-6D utility score from longitudinal TTU model using PHQ-9 (A) Density plots of observed and predicted utility scores (GLMM (Gaussian distribution and log link)) (B) Scatter plots of observed and predicted utility scores by timepoint (GLMM (Gaussian distribution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

If this study was real, the results would be interesting.

Availability of data and materials

Detailed results in the form of catalogues of the TTU models produced by this study and other supporting information are available in the online repository: https://doi.org/10.7910/DVN/D74QMP

Ethics approval

The study was reviewed and granted approval by Awesome University's Human Research Ethics Committee (1111111.1).

Funding

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 GLM and OLS models using PHQ-9 as predictor with the baseline data

Model	Training model fit (averaged over 10 folds)			Testing model fit (averaged over 10 folds)		
	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

Model	Training model fit (averaged over 10 folds)			Testing model fit (averaged over 10 folds)		
	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 individual candidate predictors with SOFAS using GLMM (Gaussian distribution and 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		

Note: The K6, PHQ9 and SOFAS parameters were first multiplied by 0.01.

Table A.4: Estimated coefficients from longitudinal TTU models based on individual candidate predictors with SOFAS using LLM (complementary log log 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		

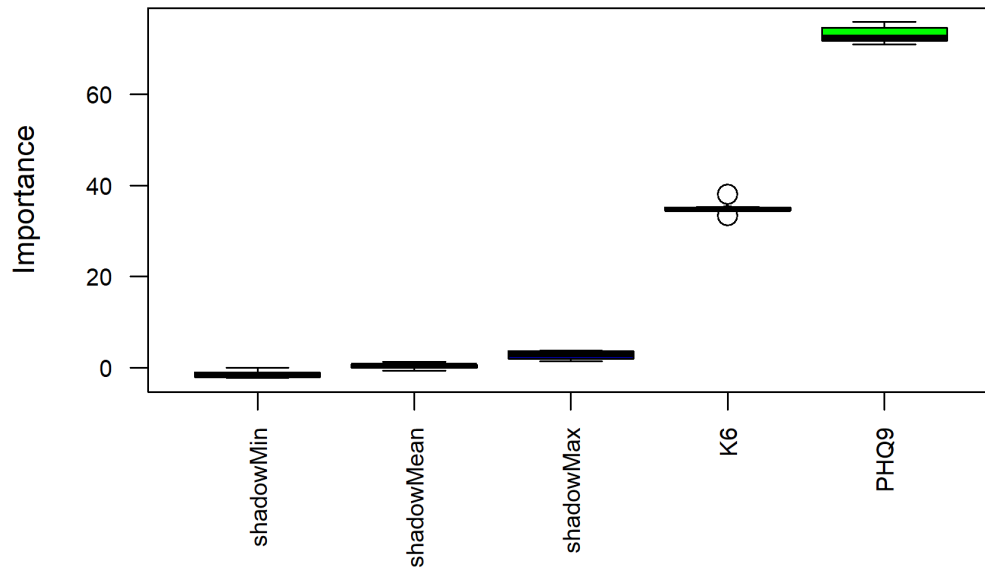Note: The K6, PHQ9 and SOFAS parameters were first multiplied by 0.01.

A.2 Additional figures

Figure A.1: Variable importance estimated using random forest