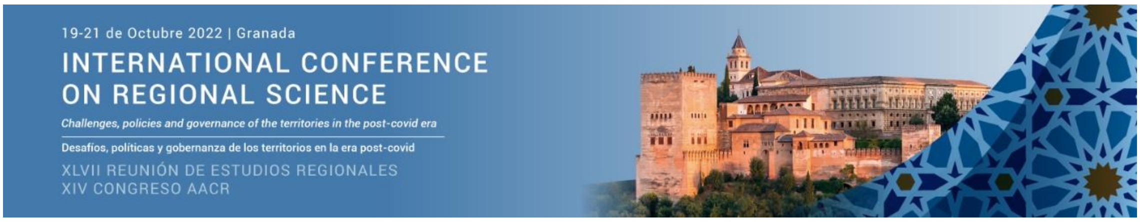

EXTENDED ABSTRACT

Title: Machine learning approaches to construct composite indices: A simulation study and application to measuring subjective well-being

Authors and e-mail of all:

Stephen O'Neill (Stephen.oneill@lshtm.ac.uk)
Zia Sadique (Zia.sadique@lshtm.ac.uk)

Department:
Department of Health Services Research and Policy

University:
London School of Hygiene and Tropical Medicine

Subject area: *(please, indicate the subject area which corresponds to the paper)*
Health Economics

Abstract: *(minimum1500 words)*

Background

Health-related quality of life (HRQoL), one of the key outcome measures of economic evaluation, focuses on important aspects of quality of life related to health. A number of different measures of HRQoL are used in practice. The EuroQoL Five Dimension (EQ 5D) is a multi-attribute utility instrument. The EQ-5D descriptive system considers five dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. It is a preference-based measure and has been widely used for assessing the cost-effectiveness of healthcare interventions. The original EQ-5D questionnaire, introduced in 1990, allows respondents to choose between three options; level 1, representing no problems; level 2, reflecting small or moderate problems; and level 3, indicating extreme problems (or 'unable to'). Since 2011 a newer versión of EQ-5D has emerged with five numerical levels for each dimention of health allowing respondent to choose between five options: level 1, representing no problems; level 2, slight problems; level 3, indicating moderate problems; level 4, indicating severe problems; and level 5, indciating unable to function or extreme problems. Self-ratings on the three levels in the five dimensions are summarized to produce health profile of the respondents. Health profiles are valued using appropriate EQ-5D value sets, according to health state preferences elicited from the general population, to calculate EQ-5D utility scores, anchored on a scale from 0 (death) to 1 (perfect health). The utility score from EQ-5D in principle represents people's preferences for a given health state. These utility scores are combined with survival data to calculate Quality Adjusted Life Years (QALYs), which provides a generic common unit of outcome measure and widely used in economic evaluation. The General Health Questionnaire (GHQ-12) is another commonly used and validated instrument for assessing the mental health aspect of general health. The GHQ-12 concentrates on the broader components of psychological morbidity and consists of 12 items measuring such characteristics as general levels of happiness, depression, anxiety, sleep disturbance and self-confidence. Six questions are positively phrased and six questions negatively so. Each item is rated on a four-point scale (less than usual, no more than usual, rather more than usual, or much more than usual). GHQ-12 provides a description of mental health status across various dimensions and it is reported as a numerical scoring system. It is often used in evaluating outcomes of mental health promotion interventions, but cannot be used in cost-utility analyses to estimate cost per quality adjusted life year (QALY), as it is not preference-based.

However, many healthcare interventions may impact more broadly on quality of life encompassing the broad range of factors that are important to people in living their lives and well-being rather than just health. Well-being integrates mental and physical health, resulting in a more holistic approach to disease prevention and health promotion. It can provide a common metric that can help evaluate the effects of different interventions and policies. There is no sole determinant of individual well-being, but in general, well being is dependent upon good health, positive social relationships, and the availability of, and access to, basic resources. Well-being indicators measure when people feel very healthy and satisfied or content with life. Commonly used HRQoL indicators fail to capture these experiences of people's daily lives, the quality of their relationships, their positive emotions, resilience, and realization of their potential. Positive evaluations of a person's life can include the presence of positive emotions in daily activities, participation in society, satisfying relationships, and overall life satisfaction. These attributes are commonly referred to as well-being and are associated with numerous postive benefits to health, work, family, and economics. However, there is a lack of evidence about the existence, direction and magnitude of association within and between different measures of HRQoL and well-being.

The aim of this article is to compare a range of existing and data-driven approaches to identify the most effective methodological technique for constructing a multidimensional measure of well-being. We apply machine learning approaches to (a) develop composite measures of well-being, (b) assess their association with components capturing health and well-being and (c) assess the performance of the methods using simulated datasets.

Data

The study uses data from the Health Survey for England (HSE), which is a nationwide survey which has been carried out each year since 1991 and provides a random, nationally representative sample with which to monitor trends in the nation's health. It provides information about adults aged 16 and over, and children aged 0 to 15, living in private households in England. The survey consists of an interview, followed by a visit from a nurse who takes some measurements and blood and saliva samples. Each survey in the HSE series includes core questions, and measurements such as blood pressure, height and weight measurements and analysis of blood and saliva samples. In addition there are modules of questions on specific topics that vary from year to year. We have used the HSE 2018 which includes questions on EQ-5D, and well-being. HSE 2018 interviewed a total of 8,178 adults (aged 16 and over) and 2,072 children (aged 0 to 15) in the 2018 survey. Our analysis is based on the adult sample (aged 16 and over). In our analysis we have excluded missing data, and 'not applicable'/'don't know' responses for EQ-5D, GHQ and wellbeing questions.

Methods

We contrast composite indices based on principal component analysis with two novel approaches.

The first method is based on adaptive LASSO (Zou, 2006) and leverages auxiliary data from a variable correlated with the phenomenon of interest (self-reported life satisfaction here). This approach uses an 'outcome' influenced by the phenomenon being measured under an assumption that indicators operate exclusively through that phenomenon, i.e. are independent of the 'outcome' conditional on the true composite index. Throughout we use an 'honest' estimation approach whereby each approach is applied to part of the data (a "training set") and evaluated on data not used to construct the composite measure (the "test set").

LASSO performs regularisation and indicator selection (Tibshirani, 1996), applying a model selection process which penalises the coefficients of the regression variables less correlated with the outcome measure (after accounting for the other indicators), shrinking some of them to zero, thereby eliminating indicators. Indicators that still have non-zero coefficients after the shrinking process are retained and a prediction used as the composite indicator. The approach taken here, adaptive LASSO applies different amounts of shrinkages to different indicators. Adaptive LASSO performs well in the presence of multicollinearity (Luo *et al.*, 2012) and provides good prediction accuracy because shrinking and removing the coefficients can reduce variance without a substantial increase in bias (Fonti & Belitser, 2017). Adaptive LASSO also reduces model over-fitting by eliminating irrelevant indicators that are not associated with the outcome variable (Fonti & Belitser, 2017).

The second approach consists of stacking copies of the original dataset (one for each component of the composite index). The individual components of the composite index are then stacked into a single outcome after normalization. A machine learning model (e.g. random forest (Breiman, 2001) is then used to predict this stacked outcome, using the components as 'explanatory' variables. Since the loss function for the Machine Learning method takes account of prediction errors for each component, the algorithm will aim to form predictions that seek to 'fit' the components as well as posible. The predictions can then be used as a composite index directly. Different loss funcions will give rise to composite índices with different properties. We apply a number of different machien learning methods on the stacked dataset including random forests, boosting, adaptive lasso and elastic nets and contrast their performance.

To assess the conditions under which each of these methods perform well, we will conduct a simulation study where components are simulated under different correlation structures, and with different functional relationships to the 'auxilliary outcome' and true well-being mesures. Each method then be applied to the artificial datasets, and their ability to represent well-being will be assessed.

Each of the methods will then be applied to the HSE case study.

Preliminary results

Initial application of the methods to the dataset suggests the approaches are feasible and explain more of the variation in life satisfaction than composite índices based on Prinicipal Component Analysis, albeit the gains are relatively modest. The importance of using an honest estimation approach is evidenced in the case study análisis where in-sample goodness of fit is far superior to out-of-sample performance, indicating overfitting. This is an important consideration for other work that adopts data driven approaches to the construction of composite índices since the composite index may fail to generalise to other settings if performance is only assessed on the data used to construct it.

Further work will refine the approach to improve performance in light of the results of the simulation study.

Conclusion

The algorithms developed in this study can be used to determine cost-effectiveness of services or interventions that use GHQ as a primary outcome where the utility measures are not collected. The composite measure of well-being provides additional insights into the determinants of well-being which would be useful to assess the wider impact of healthcare interventions.

References

Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp. 5-32.

Fonti, V. and Belitser, E., 2017. Feature selection using LASSO. *Business Analytics*.

Luo, Y., Hawkley, L.C., Waite, L.J. and Cacioppo, J.T., 2012. Loneliness, health, and mortality in old age: a national longitudinal study. *Social Science & Medicine, 74*(6), pp. 907-914.

Tibshirani, R., 1996. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, 58*(1), pp. 267-288.

Zou, H., 2006. The adaptive lasso and its oracle properties. *Journal of the American statistical association, 101*(476), pp. 1418-1429.

Keywords: *Composite Indices, Health Related Quality of Life; Machine Learning*

JEL codes: I10