Composite measures for assessing multidimensional social exclusion in later life: Conceptual and methodological challenges

Composite Measures for Multidimensional Old-Age Social Exclusion: Conceptual and Methodological Challenges

Sinéad Keogh, Stephen O'Neill and Kieran Walsh

Dr. Sinead Keogh
Irish Centre for Social Gerontology
Institute for Life-course and Society
National University of Ireland Galway
Galway

Dr. Stephen O'Neill
J.E. Cairnes School of Business & Economics
Irish Centre for Social Gerontology
National University of Ireland Galway
Galway
Ireland

and

Department of Health Services Research and Policy
London School of Hygiene and Tropical Medicine
London
UK

Professor Kieran Walsh*(corresponding author)
Irish Centre for Social Gerontology
Institute for Life-course and Society
National University of Ireland Galway
Galway
Tel: +353-91-495460
E-mail: Kieran.walsh@nuigalway.ie

Author contributions

All authors contributed to the study conception and design. Material preparation and data analysis were performed by Sinead Keogh, Stephen O'Neill, with conceptual support and guidance from Kieran Walsh. The first draft of the manuscript was written by Sinead Keogh and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of interest

The authors declare that they have no conflict of interest.

Abstract

Although there are a number of approaches to constructing a measure of multidimensional social exclusion in later life, theoretical and methodological challenges exist around the aggregation and weighting of constituent indicators. This is in addition to a reliance on secondary data sources that were not designed to collect information on social exclusion. In this paper, we address these challenges by comparing a range of existing and novel approaches to constructing a composite measure and assess their performance in explaining social exclusion in later life. We focus on three widely used approaches (sum-of-scores with an applied threshold; principal component analysis; normalisation with linear aggregation), and three novel supervised machine-learning based approaches (least absolute shrinkage and selection operator; classification and regression tree; random forest). Using an older age social exclusion conceptual framework, these approaches are applied empirically with data from Wave 1 of The Irish Longitudinal Study on Ageing (TILDA). The performances of the approaches are assessed using variables that are causally related to social exclusion.

Key words: Social exclusion, older adults, composite measure, machine learning, multidimensionality

1. Introduction

Social exclusion amongst older people is a critical research and policy concern (Van Regenmortel et al. 2016; WHO 2015). In recent years, the increasing prevalence of unequal outcomes amongst older populations (Nazroo 2017), and ‘burden’ discourses linking ageing and structural uncertainty (Walker and Zaidi 2019), has sharpened this concern. It has also reinforced a focus on the social exclusion construct as a flexible, comprehensive means of capturing multifaceted disadvantage in later life. In addition to its other characteristics of agency (or the act of exclusion), relativity, and a dynamic processual form (Atkinson 1998), social exclusion is multidimensional. It can lead to inequities in choice and control, resources and relationships, and power and rights across different domains of life (e.g. economic; social; services; etc.) (Levitas et al. 2007). Through the consideration of these domains, the construct shifts the focus from a binary view of old-age disadvantage to one that incorporates simultaneous exclusions and inclusions. As older adults can be more susceptible to multiple forms of risk (Scharf and Keating 2012; Ferraro and Shippee 2009), understanding this multidimensionality is essential for ageing societies. Therefore, a social exclusion measure that accounts for multidimensionality is crucial (Ward et al. 2014). However, how best to reflect this characteristic in the construction of a composite measure (Van Regenmortel et al. 2017), while maintaining the validity of the measure, remains under explored.

Despite a diverse range of methodological approaches, there is a lack of consensus concerning the most suitable technique for constructing composite measures of complex social phenomenon (Greco et al. 2019). Each approach offers a different capacity to preserve a multidimensional underpinning, and to provide an instrument that is empirically optimized for a particular dataset. Each approach also confronts two fundamental decisions involved in creating a measure. Namely, how to weight and aggregate indicators (OECD 2008). Weighting determines the relative importance of the indicators within a composite measure, while aggregation refers to how these weighted indicators are combined to create the measure (Greco et al. 2019). Within the international literature on old-age exclusion, approaches such as the ‘sum-of-scores’ with an applied threshold (Kneale 2012; Barnes et al. 2006; T. Scharf et al. 2005; Scutella et al. 2009), principal component analysis (Dell'Anno and Amendola 2015), factor analysis and normalisation with linear aggregation (Dahlberg and McKee 2018; MacLeod et al. 2019) have all been used. There is also a question as to whether there are more novel approaches, such as supervised machine learning (Athey 2018) not commonly applied in social indicator construction, that may be more effective in optimising a composite measure.

A rigorous comparison of these different methodologies, and their implications for the assessment of social exclusion of older people, has not been made. Without such an analysis, a society's capacity to monitor changes in the prevalence and construction of exclusion in later life, and to identify risk-factors and outcomes, will always be deficient. So too will its ability to meet the World Health Organisation's priorities for a Decade of Healthy Ageing (2020-2030), and the European Union's stated goal of a strong social Europe for just demographic transitions (European Commission 2020). The need to understand these methodological differences has become more urgent given that old-age exclusion is often measured using secondary datasets. As these surveys are not specifically designed to target social exclusion, the number and quality of relevant indicators available is often limited. Consequently, maximising the robustness and validity of any derived measure is a necessity.

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 social exclusion in later life. Our analysis targets three widely used approaches (sum-of-scores with an applied threshold; principal component analysis; normalisation with linear aggregation), and three more novel supervised machine-learning based approaches (least absolute shrinkage and selection operator (LASSO); classification and regression tree (CART); random forest (RF)). We propose a novel strategy that allows the use of machine learning methods (LASSO; CART; RF) to calculate a composite index where the outcome of interest (social exclusion here) is not observed. This approach uses an outcome influenced by social exclusion under the assumption that all indicators being used in our measure operate exclusively through social exclusion, and are thus independent of the outcome conditional on the true composite index.

We begin by briefly reviewing some of the considerations in relation to weighting and aggregation for the old-age exclusion construct, and for each of the chosen approaches.

2. Social exclusion measures and weighting and aggregation

While there continues to be varied interpretations of what 'multidimensionality' means, existing conceptual frameworks of social exclusion in later life generally demonstrate that a broad range of life domains need to be captured. With reference to Figure 1, Walsh et al. (2017) derived a framework of old-age exclusion from the international literature, identifying six domains: civic participation (CP); services, amenities and mobility (SAM); material and financial resources (MFR); social relations (SR); socio-cultural aspects of society (SA); and neighbourhood and community (NC). Exclusion domains are understood to encapsulate a range of sub-dimensions, which in turn need to be characterised by their own individual sets of indicators. Thus, exclusion in later life is not only multidimensional, but that multidimensionality is layered (MacLeod et al. 2019). As a result, assigning weights within an index is particularly complex given the weights must be allocated at the domain, the sub-dimension and indicator levels to fully determine the nature of the construct's multidimensionality (García et al. 2019). Likewise, decisions related to aggregation are made at these same levels, and dictate the degree to which a multidimensional structure is retained. Thus, weighting and aggregation are fundamental in representing the conceptual make-up of social exclusion. Weighting methods can be categorised as ‘equal’ or ‘unequal’ (Greco et al. 2019) and are typically linked to different forms of aggregation.

Equal weighting, generally the most common strategy used in the development of social exclusion measures, implies that a priori the same weight is assigned to each indicator (Greco et al. 2019; Van Regenmortel et al. 2017). Becker et al. (2017) and Levitas et al. (2007) suggest that equal weighting approaches are unlikely to be empirically or theoretically correct as all variables are unlikely to have the same importance. There are, however, various reasons why equal weights are employed for social exclusion measures, including: the simplicity of construction; issues around identifying exact weights; and a lack of theoretical/empirical evidence justifying a different weighting scheme, which is particularly the case for exclusion in later life, (Decancq and Lugo 2013; Greco et al. 2019). Aggregation approaches that generally assume equal weighting include ‘sum-of-scores’, and normalisation with linear aggregation (Kneale 2012; Van Regenmortel et al. 2017).

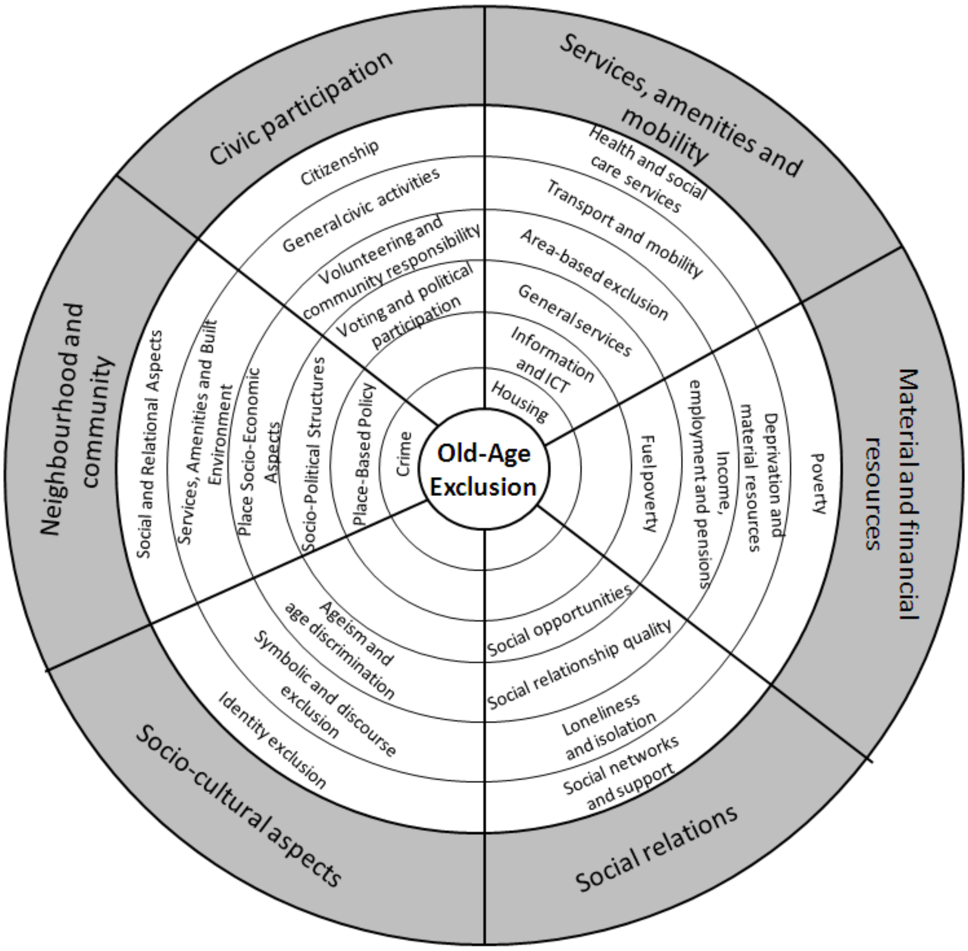

Figure 1 Old-age exclusion framework depicting interconnected domains and sub-dimensions

Source: Walsh et al. 2017

For the sum-of-scores approach, a summed score for each domain is generated based on the number/proportion of sub-dimensions (and therefore the number/proportion of indicators within each sub-dimension) that a respondent is excluded upon (Scutella et al. 2013). Researchers using this approach usually set a minimum threshold that signifies a distinction between exclusion and non-exclusion and therefore produces a binary measure of exclusion. There is little agreement across studies regarding the threshold (Van Regenmortel et al. 2017) and setting the threshold is considered an arbitrary decision. Thresholds can be ‘absolute’ or ‘relative’. Under an absolute threshold, an individual is considered excluded if they are disadvantaged in a set number of domains. For example, Tom Scharf et al. (2005) identified respondents as socially excluded if they were excluded from at least one domain. Relative thresholds determine exclusion relative to either the number of indicators overall (e.g. excluded on half of the indicators) or the number of respondents. Barnes et al. (2006) and Van Regenmortel et al. (2017), for example, identify excluded groups as those who are part of the 10% most excluded in the sample.

Once weighted, indicators are combined using an aggregation function, which can include 'linear' (by summing indicators), 'geometric' (multiplying indicators), and 'multi-criteria' (non-linear techniques) aggregation functions (OECD 2008; Van Regenmortel et al. 2017; Scutella et al. 2013). With linear aggregation, typically the most commonly used, weights express trade-offs between indicators, where deficits in one sub-dimension can be offset by surpluses in another, implying full compensability. Geometric aggregation offers partial (non-constant) compensability while multi-criteria is non-compensatory and is generally the least popular approach to aggregation.

In the normalisation with linear aggregation method, normalisation is carried out to attribute the same measurement units to all indicators prior to aggregation. This enables the direct summation of indicator values and avoids indicators with a larger range of values being attached greater importance. A number of normalisation methods exist such as ranking, standardisation, categorical scales, and cyclical indicators. However, it is scaling using min-max normalisation that is most frequently used within the exclusion literature (Scutella et al. 2009). While outliers may distort the transformed indicator, the advantage of this approach is that it can widen the range of indicators lying within an interval (OECD 2008). For example, Buddelmeyer et al. (2012) normalised the indicators for each domain so that every individual had a ‘score’ that lies between 0 (not excluded in that domain) and 1 (fully excluded). Dahlberg and McKee (2018), MacLeod et al. (2019) and Prattley et al. (2020) have all utilised this approach to construct a measure of social exclusion in later life, moving away from a binary style assessment, and instead opting to assess the extent of exclusion.

The combination of these weighting and aggregation methods are generally critiqued for not optimising data fit. They are however considered to facilitate the retention of a desired multidimensional conceptual structure, whether that is at the sub-dimension or domain level (Mazziotta and Pareto 2016).

Unequal weighting approaches can be divided into three sub-categories, ‘knowledge driven’ (based on the preferences of an expert panel), ‘data driven’ (weights obtained directly from the data structure) and ‘hybrid’ (a mix of knowledge driven and data driven) (OECD 2008). For the purpose of this paper we focus on data-driven weighting techniques, where these methods obtain weights directly from the data, using a mathematical function, to determine a reductionist measure structure (Greco et al. 2019). They are as such considered to generally create measures more directly related to the data and, like the previous approach, are used to capture the extent of exclusion.

Principal component analysis (PCA) is a commonly used example of an unequal weighting technique. The aim is to identify a set of principle components that explain variation in the sample by reducing the dimensionality of the data, while retaining key information. This approach has been used extensively to construct social exclusion measures within the literature, where for example Dell'Anno and Amendola (2015) used PCA to determine the weights and component loadings in an exclusion index for 28 European countries. However, the derivation of weights through PCA is considered to lack transparency (Decancq and Lugo 2013) as it may assign lower weights to a crucial indicator simply because it is weakly correlated with other indicators (Best et al. 2020; Mohnen et al. 2020). The process is also considered arbitrary as the decision to retain components is largely discretionary (OECD 2008).

Increasingly, data-driven machine learning approaches are viewed as an attractive alternative to more familiar methods such as PCA, or regression analysis. These techniques build in model selection that is data-driven rather than motivated by theoretical or statistical assumptions (Athey 2018). That is, variable selection and weighting is performed by the algorithm, reducing the risk of over-fitting, which occurs when the model is too rich relative to the sample size. While to our knowledge this family of techniques, which include LASSO, CART and RF, have not been applied to the creation of a social exclusion measure, researchers are increasingly looking to these approaches to more accurately capture other complex social phenomena (Athey 2018; Radford and Joseph 2020).

3. Developing a Measure of Social Exclusion

The conceptual framework presented by Walsh et al. (2017) provides the basis for the selection of indicators, and is used to guide how these indicators can be weighted and aggregated into a composite measure. Indicators were sought for all six exclusion domains, and the various sub-dimensions of those domains.

This study employs cross-sectional data from Wave 1 of The Irish Longitudinal Study on Ageing (TILDA), a nationally representative cohort study of community dwelling older adults. Variation in available indicators across waves prevented a panel analysis. Despite this, TILDA is the most comprehensive study of ageing in the Irish context, and captures a wide range of variables on health, social and economic aspects of later life. Its structure and question composition, therefore, lends itself to the analysis of complex multidimensional social phenomena, and has been used previously to investigate patterns of multifaceted societal participation (e.g. Donoghue et al. 2019). TILDA is also one of a network of harmonised longitudinal studies on ageing from across the globe and a part of the Gateway to Global Ageing Data initiative, which identifies comparable questions across surveys to facilitate cross-country research (TILDA 2019).

Wave 1 data-collection took place between October 2009 and February 2011 using clustered random sampling of all households in Ireland. Representing a response rate of 62%, 8,175 adults, aged 50 and over, completed a computer assisted personal interview (CAPI) in their own home. Detailed information on all aspects of the respondents' lives was collected in the CAPI, including economic status, health aspects and social participation. An additional self-completion questionnaire, probing on more sensitive topics (e.g. relationship quality; depression; loneliness) was completed by 84.6% (n=6,915) of the sample and returned after the interview (Whelan and Savva 2013; Kearney et al. 2011).

Table 1 presents the relevant TILDA indicators that correspond to the domains and sub-dimensions of the conceptual framework, and which were used to construct the composite measures of exclusion. While the large number of topics included in TILDA is a considerable strength, there are in the context of social exclusion, some absences. The Wave 1 dataset does not include information on ageism and age discrimination, general services, housing and technology, or neighbourhood and community. However, it still provided the best available selection of indicators for our purposes. On this basis, we limit our focus on social exclusion to CP, SAM, MFR, and SR.

Table 1: Domains of old-age social exclusion, their sub-dimensions and their indicators

Domain	Sub-dimension	Indicator
Civic participation (CP)	Citizenship*	
	General civic activities	Regularly attend religious service Participates in a course or educational training Participates in sports or social groups
	Volunteering & community responsibility	Participates in voluntary work at least once a month
	Voting & political participation	Voted in the last general election
Services, amenities & mobility (SAM)	Health & Social care services	Does not receive but has a need for 1) community nurse, 2) chiropody, 3) physiotherapist, 4) home help or 4) day centre
	Transport & mobility	Lack of transport affects lifestyle Not being able to drive affects 1) socialisation with others, 2) ability to attend health & social care appointments, 3) ability to get out about business Ability to drive affected Rating of private & public transport
	Area-based exclusion*	
	General services*	
	Information & ICT*	
	Housing*	
Material & financial resources (MFR)	Poverty	Household equivilised income
	Deprivation & material resources	Shortage of money stops one from doing things one wants to do
	Income, employment & pension*	
	Fuel poverty*	
Social relations (SR)	Social opportunities	Participation in social activities
	Social relationship quality	Interactions with spouse, children, friends or family members
	Loneliness & isolation	UCLA loneliness score
	Social networks & support	Size of social network
Socio-cultural aspects (SA)	Ageism & age discrimination*	
	Symbolic & discourse exclusion*	
	Identity exclusion*	
Neighbourhood & Community (NC)	Social & recreational aspects*	
	Services, amenities & built environment*	
	Socio-political structure*	
	Place-based policy*	
	Crime*	

Note: *No indicator available in Wave 1 of TILDA Survey data.

In the CP domain, three out of the four sub-dimensions are available. Having voted in the last general election is considered an indicator of voting and political participation. Volunteering and community responsibility are captured by participation in voluntary work at least once a month. An indicator for general civic activities is derived from three indicators that relate to frequency of attendance at religious services, participation in courses or education, and participation in groups such as sports or social groups or clubs.

Although Walsh et al. (2017) list six sub-dimensions for the SAM domain, the data set limits our analysis to just two of these sub-dimensions. Exclusion from health and social care services is captured in a question that asks respondents if they have the need for, but do not receive, the following services: community nurse; chiropody; physiotherapy; home help or; a day centre. Social exclusion due to transport and mobility comprise three indicators. The first two ask respondents to rank private and public transport in their neighbourhood and the third asks respondents if the lack of transport facilities in their area affects their lifestyle.

For the MFR domain, two out of the four sub-dimensions could be included: poverty, and deprivation. We use an income variable as a proxy for poverty. This poverty indicator is annual disposable income of the household, adjusted for household composition. The deprivation indicator refers to an inability of the respondent to do the things they want to do because of a shortage of money.

All sub-dimensions of the SR domain are represented to some degree in the data set. The social opportunities indicator captures respondents' participation in social activities such as attending films, plays or concerts; playing cards, bingo or games in general; going to the pub; eating out of the house; visiting or being visited by family or friends. Social relationship quality explores social bonds with a spouse, children, friends or other family members. Loneliness and isolation was measured using the University of California-Los Angeles (UCLA) scale for loneliness.

The social networks and support indicator is derived from responses to three items, concerning the number of (1) children, (2) relatives and (3) friends the respondent feels close to.

Overall, four domains, 11 sub-dimensions within those domains, and 56 indicators describing the sub-dimensions, were included in the analysis.

4. Comparative approach and assumptions

If an appropriate indicator of social exclusion was available within a data set, the performance of alternative measures of social exclusion could be assessed based on their correlation with this 'true' indicator. But like most secondary datasets on ageing, TILDA does not include a validated measure of exclusion, or a similar construct of disadvantage, to facilitate a benchmark comparison across the chosen approaches. We propose a novel way to overcome this limitation. We evaluate the performance of each approach by comparing them in relation to the 19-item Control, Autonomy, Self-realisation, Pleasure scale (CASP-19). CASP-19 is an internationally recognised reliable and valid measure of multifaceted quality of life (Hyde et al. 2015; Wiggins et al. 2004; Hyde et al. 2003). While CASP-19 is not a measure of social exclusion it offers a multidimensional counterpoint to social exclusion. To facilitate our comparison, we assume social exclusion operates through the four domains of CASP-19, with the chosen indicators influencing these domains only through exclusion. This is, admittedly, a strong assumption. However, some clear relationships between the constructs of social exclusion and quality of life, and their constituent domains, have previously been established (Pantazis et al. 2006; Barnes 2019).

In our analysis, first, we assume (A1) that quality of life, as measured by CASP-19, can be decomposed into two parts: a part that correlates with old-age exclusion (OAE); and a part that does not, non-old-age exclusion (NOAE). Second, we assume (A2) that the chosen indicators correlate with OAE, but are independent of NOAE. Under these assumptions, a model using the indicators which provides a better prediction for CASP-19, is likely to do so through predicting the OAE component of CASP-19. Thus, the approach that performs best in terms of their ability to explain quality of life, as captured through CASP-19, is assumed to be the best approach when measuring old-age social exclusion using this particular dataset.

5. Empirical strategy

We construct composite measures of the *extent* of old-age social exclusion using normalisation with linear aggregation, principal component analysis, LASSO, CART, and RF. We also construct a binary measure of exclusion using the sum-of-scores with an applied threshold approach that distinguishes the excluded from the non-excluded. We describe each method before describing our approach to evaluate their performance.

5.1 Normalisation with linear aggregation

There are two primary steps in utilising linear aggregation with normalisation to build a composite measure. First, as the indicators have different measurement units, they must be transformed so that they are comparable. We do this by normalising them (OECD 2008), using a min-max scaling method, specified as:

$$I_{ji} = \frac{X_{ji} - X_{min,i}}{X_{max,i} - X_{min,i}} \tag{1}$$

where $X_{max,i}$ and $X_{min,i}$ are the upper and lower values respectively for the indicator X_i. Thus, (1) represents a linear transformation of the variable, resulting in a normalised indicator score I_{ji} that can assume values between 0 and 1. Second, once the normalisation has been completed, the values between 0 and 1 for each sub-dimension are aggregated and rescaled again to a value between 0 and 1 for the total domain score, with this process being repeated across the domains to obtain a total extent of exclusion score, providing the composite index. This approach assigns equal weight to the sub-dimensions and the domains, assuming the equal importance of all (normalised) indicators.

5.2 Principal component analysis (PCA)

PCA has four primary steps. Following normalisation, the first step is the computation of the correlation matrix of the indicators in (2) where X_{di} and X_{dj} are the *i*th and *j*th indicator in domain d (d=1,…,n) and r_{ij} is the correlation coefficient between the indicators, defined as:

$$r_{ij} = \frac{1}{n-1} \sum_{d=1}^{n} X_{di} X_{dj} \qquad (2)$$

Indicators with correlation values close to +1 (or -1) have a strong correlation and may be largely explained by a common principal component. Next, to determine the number of components in the PCA, eigenvalues and their corresponding eigenvectors are calculated through (3) where R is the indicator correlation matrix, λ represents the eigenvalues and I is the unit matrix:

$$(R - \lambda I) = 0 \qquad (3)$$

Following this, the eigenvector (principal component) F_j is derived as follows:

$$\left(R - \lambda_j I\right) F_j = 0 \qquad (4)$$

where λ_j is the eigenvalue corresponding to F_j. Dividing λ_j by the sum of the eigenvalues gives the proportion of total variation explained by the jth eigenvector. As commonly employed, the first principal component explaining the largest proportion of total variation of all indicators is used as the composite index (Kolenikov and Angeles 2009). A key drawback with this approach is that PCA focuses on correlation between indicators, rather than their correlation with the outcome variable of interest. Therefore, if included indicators are highly correlated, but not relevant to the phenomenon that is ultimately being measured, a composite index based on PCA may perform poorly.

5.3 Adaptive logistic least absolute shrinkage and selection operator (LASSO)

Using the normalisation technique, we apply the adaptive LASSO approach. LASSO performs regularisation and indicator selection (Tibshirani 1996). It applies a model selection process where it 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 included in the model. The approach taken here, adaptive LASSO (Zou 2006), 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 and Belitser 2017). Adaptive LASSO also reduces model over-fitting by eliminating irrelevant indicators that are not associated with the outcome variable (Fonti and Belitser 2017). In our application, we used the command lasso2 included in the lassopack module of Stata (Ahrens et al. 2020) for parameter estimation. The regularisation algorithm as implemented in this procedure minimises the following penalised criterion (Belloni et al. 2012):

$$Q(\tilde{\gamma}) = \frac{1}{n} RSS(\alpha, \delta, \tilde{\gamma}) + \frac{\lambda}{n} J(\tilde{\gamma}) \tag{5}$$

where n is the sample size, $\alpha, \delta,$ and $\tilde{\gamma}$ denote the model parameters, $RSS(\alpha, \delta, \tilde{\gamma})$ is the residual sum of squares, λ is the overall penalty parameter, and $J(\tilde{\gamma})$ is the penalty term. The penalty term λ is chosen by minimising the extended Bayesian Information Criteria (EBIC) proposed by Chen and Chen (2008).

We applied three versions of adaptive LASSO, applying it (i) at the four-domain level (LASSO-4) where we utilise the aggregated value for each domain; (ii) at the 11 sub-dimension level (LASSO-11) where we utilise the aggregate value of the indicators for each sub-dimension; and (iii) the 56-indicator level (LASSO-56). The 11 sub-dimensions were derived using linear aggregation following normalisation of the relevant indicators for each sub-dimension and similarly the four domains were derived by aggregating the sub-dimensions into a single index. If the conceptual framework is correct, we anticipate that the 56 indicators would be retained, and if the weighting approach used to make domain and sub-dimension indices are correct, we anticipate that LASSO-4 and LASSO-11 should perform similarly to LASSO-56. Where this is the case, it may indicate that a data driven approach to determine weights may be superior. However, since the LASSO model is linear, the predicted outcome (our composite index) still utilises linear aggregation. A linear model, such as LASSO may not perform well if the relationship between indicators and social exclusion is non-linear.

5.4 Classification and regression tree (CART)

CART allows for non-linearity by using recursive partitioning. The CART algorithm repeatedly partitions the data into smaller and smaller subsets until those final subsets are as homogeneous as possible in terms of the outcome variable. CART identifies the best split by iteratively considering all possible splits (‘branches’) for all indicators and settling on the split that produces the greatest reduction in impurity within subgroups (Hayes et al. 2015). It then considers each of these subgroups and repeats the process, splitting each subgroup into smaller, more similar, subgroups. CART continues until no further reduction in impurity is possible or where a stopping rule is reached. At this point the final subgroups are commonly referred to as ‘leafs’, forming the basis of a decision tree to enable a prediction. Here we require that any split must have at least ten observations in each subgroup, and that the split must decrease the overall lack of fit by a factor of at least 0.1%.

Since the decision tree is formed by a sequence of binary splits, nonlinear relationships between the outcome and indicators can be accommodated more flexibly than in standard regression models. CART employs cross validation to assist in sizing the tree (Lewis 2000; Loh 2014). Cross-validation involves randomly splitting the sample into K blocks and then estimating the model using data from K-1 of the blocks and assessing performance on the Kth block. Since the excluded block was not used in estimation, the model’s performance which is commonly assessed by R^2 or RMSE, is less likely to be due to over-fitting the data at the estimation stage. Each of the blocks is excluded in turn, and the average performance across the K blocks is used to assess the model’s overall performance. In our analysis, we used 10-fold cross-validation (i.e. K=10). All 56 indicators were included in the CART algorithm. Under assumptions A1 and A2, a tree that predicts CASP-19 should also perform well in predicting social exclusion.

An important limitation of CART is that it can be sensitive to the particular sample used since splitting decisions may be influenced by noise in the data. Relying on a single tree, estimated on a single dataset, may lead to poor out-of-sample predictions, limiting the usefulness of our social exclusion measure (the prediction).

5.5 Random Forest (RF)

To address the potential for poor out-of-sample predictions, we employ the RF technique, a nonparametric approach based on the average predictions of many individual decision trees (Breiman 2001). For each tree, a random bootstrap sample with replacement is used and a random subset of the 56 indicators is chosen (here we used 19 indicators per tree, although results are not sensitive to this choice). Together these reduce the risk of over-fitting. Next the tree is estimated but using only the subset of indicators. This process is repeated for each of the trees giving a 'forest'. Here we use 500 trees, i.e. 500 sets of 19 randomly chosen indicators from our 56 eligible indicators. Next, the predictions from each tree are averaged to reach a forest-based prediction.

5.6 Sum-of-scores with an applied threshold

This binarised approach distinguishes the *excluded* from the *non-excluded* using three steps. First, all item response categories are transformed into a binary indicator of exclusion based on whether the measure exceeds a threshold – for example, the poverty indicator is equal to one if income is below the poverty threshold and zero otherwise. Second, each of these binarised indicators are then summed, with each respondent being given an exclusion 'score' on each sub-dimension with each domain then defined as the sum of binarised indicators within that domain. Third, as there is not a definitive point when a respondent is excluded, a minimum threshold is applied at the domain level to the exclusion score. Similar to Van Regenmortel et al. (2017), we apply four different thresholds – two absolute and two relative – for determining whether a respondent experiences exclusion or not. Using Threshold 1, respondents were considered excluded in a particular domain if they were excluded on at least one of the sub-dimensions. For Threshold 2, respondents were excluded if they were excluded on all sub-dimensions. Threshold 3 is relative to the number of sub-dimensions, and being excluded on at least half of the sub-dimensions within the domain. For Threshold 4, respondents are excluded if they were among the 10% most excluded older adults on that domain.

6. Performance evaluation

The evaluation of the approaches is split into three parts. First, we assess the performance of the five composite measures (extent of exclusion) in predicting the OAE component of CASP-19. A subfield of supervised machine learning is used to split the dataset into a training set (80% of respondents), and a test set (remaining 20%). A regression analysis to build a model of the relationship between the approaches and quality of life measure, CASP-19, is then completed. Importantly, all modelling was conducted on the 80% training set, and all testing on the 20% test sample. A number of performance metrics including $R^2$and mean squared prediction error (MSPE) were calculated to measure the performance of the six measurement approaches. To account for the possibility that some statistical approaches may 'over-fit' the data, our ultimate assessment of performance is based on the R^2 on a portion of the sample not used for estimation (the 'hold-out' sample). Here the 'hold-out' sample was chosen as a random 20% subsample of TILDA Wave 1 that was excluded from the model development stage. We further evaluated which variables were ranked by the best fitting model as most predictive of the part of CASP-19 that explains social exclusion. Empirical evidence based on out-of-sample forecast performance in the hold-out sample is generally considered more trustworthy than evidence based on in-sample performance, which can be more sensitive to outliers and data mining (Clark 2004). Second, to assess variability across approaches for identifying exclusion in individual respondents, we chart the best performing approach against the other four techniques. Third, we then apply a threshold to each of the extent of exclusion approaches to determine whether a respondent is socially excluded and compute the prevalence of social exclusion. We compare these prevalence rates to the prevalence rates derived using the sum-or-scores with applied threshold technique. This gives us a sense of variability across the extent measures, and allows comparisons with the binary approach.

7. Results

The performance metrics, R^2 and MSPE, for both the 'estimation' and 'hold-out' samples for each of the five 'extent of exclusion' approaches are presented in Table 2. These approaches were all applied following normalisation of the specified indicators through min-max rescaling. Looking at the results of the regression analysis on the 'hold-out' dataset, we observe that these five measurement approaches performed in the poor-to-good range for predicting CASP-19 with R^2's ranging from 0.099 to 0.538, with the three adaptive LASSO approaches appearing to outperform all other techniques. It should be noted that since the CASP-19 reflects both OAE and NOAE, we would not anticipate even a perfect measure of social exclusion to lead to an R^2 of 1.

Table 2: Performance evaluations for extent of exclusion approaches

	'Estimation' Sample (80% training set)		'Holdout' Sample (20% testing set)	
	R^2	MSPE	R^2	MSPE
Normalisation through min-max scaling				
Adaptive LASSO 4 Domains	0.442	31.302	0.496	29.067
Adaptive LASSO 11 sub-domains	0.492	28.508	0.536	26.749
Adaptive LASSO 56 Indicators	0.518	27.117	0.538	26.567
Classification & Regression Tree (CART)	0.627	21.365	0.262	43.996
Random Forest (RF)	0.938	5.701	0.416	31.071
Principal Component Analysis (PCA)	0.101	2084.23	0.099	2059.25
Linear aggregation	0.291	39.764	0.307	39.720

Though all three applications of adaptive LASSO perform similarly, LASSO-56 performs marginally better with an R^2 of 0.538, compared to LASSO-11 (R^2 of 0.536) and LASSO-4 (R^2 of 0.496). For all three applications of adaptive LASSO, all indicators, sub-dimensions and domains were retained suggesting that the conceptual framework was an appropriate guide for creating the measure. Furthermore, the similar performance of the three applications implies that the weighting approach used to compute composite measures at the domain and sub-dimension level are fairly close to optimal. Finally, the relatively higher R^2 for LASSO-56 indicators suggests that quality of life (CASP-19) is strongly related to the indicators believed to represent all aspects of social exclusion.

In line with our expectations, RF outperforms CART both in-sample and out-of-sample. We note that its performance deteriorates noticeably in the hold-out sample. Linear aggregation also performs poorly, albeit better than CART, indicating that there are gains from allowing the weights given to each (normalised) indicator to vary. PCA performs the least well of the five extent of exclusion measures. Since here we retain the first principal component, and the indicators are believed to represent different domains, we can view this as suggesting that the domains themselves are not highly correlated so including only one principal component excludes a lot of information that could predict the multidimensionality social exclusion and hence quality of life. This argues against the use of PCA for constructing composite indices where the underlying indicators capture different domains and dimensions of a phenomenon.

Figure 2 presents scatter plots of the relationships between the extent of exclusion measured using the best performing approach (adaptive LASSO-56) against each of the alternative techniques. This allows us to examine the degree of variability across the approaches in relation to capturing exclusion for individual respondents, As would be expected, Figure 2 (a) and (b) suggest strong positive associations between the extent of exclusion using the best performing and the similarly performing adaptive LASSO-4 and LASSO-11 approaches, indicating individuals with higher levels of exclusion using LASSO-56 also have high levels of exclusion using these methods. The linear aggregation (c), random forest (d), CART (e) measures are positively correlated with the LASSO-56, although they have noticeably lower correlation with the best performing measure, highlighting the sensitivity of measurement to the choice of measurement approach used. However, the PCA approach (f) has a negative relationship with the best performing approach, reinforcing the suspicion that PCA is not an appropriate means of constructing a measure of old-age exclusion for this data set.

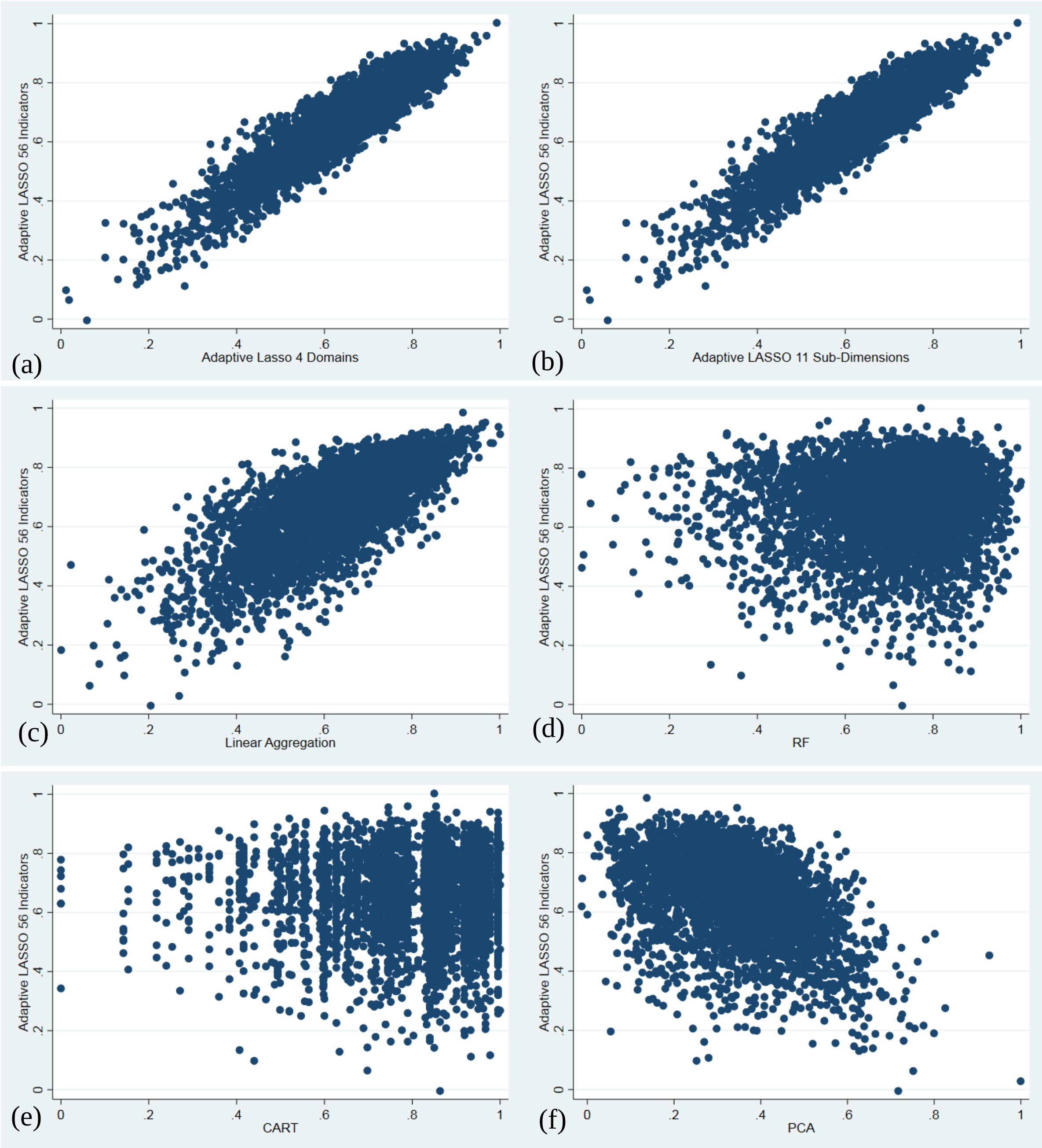

Figure 2: Scatter plots of best performing exclusion measure vs alternative measures

In terms of the sum-of-squares with applied threshold approach, Table 3 presents the prevalence of social exclusion in later life across the four domains for each threshold applied. Largely mirroring the pattern identified by Van Regenmortel et al. (2017), the application of different thresholds yields different prevalence rates across the four domains. In general, for threshold 1 (‘excluded’ if excluded from at least one sub-dimension) high levels of exclusion were found for each domain with the highest prevalence of exclusion (94%) from CP. As would be expected, the prevalence of exclusion using threshold 2 (‘excluded’ if excluded from all sub-dimensions) was much lower, ranging from 0% to only 5% across the domains. Threshold 3 (‘excluded’ if excluded on at least half of the sub-dimensions) produces prevalence rates similar to threshold 1 with the exception of exclusion from SR from which only 2% were excluded. This may be explained by three of the four sub-dimensions of SR having low exclusion rates (0.5%-1.5%). Applying threshold 4 (excluded if included in the 10% most excluded), 49% of older adults experienced exclusion from SAM, followed by exclusion from SR (48%) and CP (44%). The lowest prevalence of exclusion was for MFR with only 4% excluded from this domain. The domain of MRF has only two sub-dimensions (deprivation and poverty) and only 23% and 19% were excluded on each of these sub-dimensions possibly accounting for the low prevalence from exclusion from MRF.

One of the main disadvantages of threshold 1 and 2 (absolute) is that they are dependent on the number of sub-dimensions. As there is a lot of variation in the number of sub-dimensions for different domains, threshold 3 (dependent on the number of sub-dimensions) yields prevalence rates between two extremes (2% - 48%). This makes this absolute approach less desirable. The cut-off applied in threshold 4 is relative to the sample and is the threshold that is typically applied in the case of social exclusion research (Kneale 2012; Van Regenmortel et al. 2017). In this case the average prevalence of social exclusion across the four domains is 36%.

Table 3: Prevalence of exclusion for sum-of-scores with applied threshold approach

	Absolute		**Relative to number of sub-dimensions**	**Relative to sample**
	Threshold 1 (%)	Threshold 2 (%)	Threshold 3 (%)	Threshold 4 (%)
Domains of old-age exclusion				
Social relations	48	0	2	48
Material & financial resources	37	4	37	37
Services, amenities & mobility	49	5	48	49
Civic participation	94	4	44	44

Next, in order to contrast all six approaches, we apply threshold 4 to the composite measures derived using the linear aggregation, PCA, LASSO, CART, and RF approaches. With reference to Table 4, this allows us to present the prevalence of exclusion in the same manner as the sum-of-scores with applied threshold method. The prevalence of exclusion across the five approaches ranges from 8% to 66%. The approaches that performed the best in relation to the extent of exclusion, adaptive LASSO, yield prevalence rates that range from 33% to 39%, similar in range to the average exclusion rate using the same threshold for the sum-of-scores approach. This may suggest that the sum-of-scores approach with a 10% threshold applied performs similar to adaptive LASSO in the case of this data.

Table 4: Prevalence of exclusion for composite indices with applied threshol

	Prevalence of old-age exclusion (%)
Approach (application of threshold 4)	
Adaptive LASSO 4 Domains	37
Adaptive LASSO 11 Sub-domains	39
Adaptive LASSO 56 Indicators	33
Classification & Regression Tree (CART)	54
Random Forest (RF)	66
Principal Component Analysis (PCA)	8
Linear aggregation	24

8. Discussion and conclusion

Using data from a representative sample of older Irish adults, this paper compared a range of existing and novel approaches to constructing a composite measure of social exclusion for older people. In so doing it sought to address knowledge deficits concerning how to account for the conceptual underpinning of the multidimensional construct, while preserving validity in relation to the data set in composite measure development. Using an innovative strategy, we assessed the approaches by employing a quality of life outcome influenced by latent social exclusion, and by incorporating a 'training' sample for measure development and a 'test' sample for measure evaluation. Assessing performance in this way provides a more accurate picture of predictive ability (Clark 2004), and in itself marks a contribution to how complex social indicator measures can be developed.

It is important to acknowledge from the outset, that there are clear restrictions to the scope of our analysis. In focusing on multidimensionality, this article has only considered the measurement aspect of one of the four common characteristics of social exclusion - the others being agency, its relative construction and its dynamic nature (Atkinson 1998). The latter two characteristics are largely outside considerations directly related to a composite measure and are more commonly associated with overall study design – where for example an assessment of exclusion is relative to the given population and context from which the sample is drawn, and where its dynamic elements can be tracked over time if multiple waves of data have been collected. The first characteristic, however, draws attention to a question surrounding the capacity of secondary datasets to assess actual exclusion and to serve as viable sources of exclusion indicators. For the most part, indicators typically measure participation, utilisation and possession levels and, across a variety of survey studies on ageing, are rarely designed and phrased to identify whether a given level represents a form of exclusion. Thus, indicators often fail to capture the actual act or agency of exclusion (Ward et al. 2014). While clearly outside the focus of this paper, it cannot be ignored that these challenges impinge on the validity of any composite measure, and must as such be the focus of future work.

There are also three more direct limitations to our analysis. First, we did not have access to items for the full range of exclusion domains outlined in the conceptual framework (i.e. socio-cultural aspects; neighbourhood and community). This means that the base indicator set was insufficient to fully reflect the multidimensionality of old-age exclusion. Second, contrasting the extent of exclusion approaches against those that present a dichotomy is not ideal, and effectively compromises the core attributes of these techniques in order to facilitate the comparison (e.g. converting extent measures into a binary measure; computing a single 'unidimensional' value from the multi-domain sum-of-scores approach). Therefore, the similarities/differences in identified prevalence rates has to be judged as an approximation. Third, and perhaps most significantly, we make a strong assumption that quality of life measure by CASP-19 can be deconstructed into two parts: a part that correlates with old-age exclusion and a part that does not. A 'true' indicator of social exclusion, if one was identifiable, would provide a more appropriate measurement barometer. Caution is therefore needed when interpreting the performance evaluation presented here.

Nevertheless, the approach taken in this paper has allowed us to go further in the assessment of these different techniques than would have otherwise been possible.

Using the TILDA dataset, our analysis suggests that the optimal approach for measuring the extent of social exclusion is adaptive LASSO applied to normalised indicators. This appeared to be consistent across the different evaluation parts, including the prediction of the CASP-19 exclusion component, the scatter plot sensitivity analysis, and the computation of the exclusion prevalence rate. Moreover, the 56 indicator version of this approach only performs marginally better than the 11 sub-dimension and 4 domain variations. This suggests that the data driven weights for the indicators and those given when directly aggregating indicators to the sub-dimension, or domain levels are similar. It can be argued that aggregation across indicators, e.g. in the way adaptive LASSO synthesises multiple indicators into a single composite, can undermine the very notion of multidimensionality that is core to the concept of social exclusion. It also sacrifices our capacity to examine the relationship and association between different domains of social exclusion as it occurs in later life. However, many researchers recognise the need for some form of aggregation (Vrooman and Hoff 2013; Sacker et al. 2017; van Bergan et al. 2017), and its benefits for creating a tool for policy analysis and public communication. Further, as the adaptive LASSO approach did not eliminate any of the indicators, it does, at least in our case, reinforce the multifaceted nature of social exclusion in later life suggested by the conceptual framework. Arguably then, this approach produces both a measure that is an effective predictor of exclusion for our data, and one that is conceptually valid.

However, our analysis also shows that despite the equal weighting of indicators, and the arbitrary nature of threshold setting, the sum-of-scores technique can produce a similar rate of exclusion to the adaptive LASSO approaches. As found by (Van Regenmortel et al. 2017) it was the relative thresholds that appeared to be most applicable, with the 10% threshold used in this analysis identifying a similar rate of exclusion to that found in studies of other jurisdictions. This reinforces a number of reasons for why this approach is often used. While the binary classification of excluded and non-excluded neglects the multidimensional possibility of simultaneous exclusion and inclusion, the technique allows for the exact specification of a multi-domain structure, and its sub-dimensions. Additionally, and while a number of researchers have focused on the importance of some domains more than others (e.g. neighbourhood and community – MacLeod et al. 2019), employing equal weights provides a conceptual flexibility. A flexibility that may be considered necessary to reflect the lack of knowledge about the precise dynamics across domains in constructing exclusion in later life. Therefore, with careful consideration of the applied threshold, the sum-of-scores approach may offer a reasonably straightforward alternative to more complex processes of constructing composite measures.

In practical terms, our findings regarding the performance of individual approaches cannot be generalised outside of the dataset employed in this analysis, or indeed the underlying social, economic and cultural context of the older Irish population. Ireland, as a post-colonial setting, with what remains, a relatively young demographic structure (13 per cent aged 65 years and over), and a relatively homogenous older population (in terms of race and ethnic and religious backgrounds: CSO (2016)), is likely to be somewhat unique amongst its western developed counterparts. Therefore, patterns and interrelationships within and between the different domains of exclusion may not necessarily translate beyond the country context. As the TILDA data set (including Wave 1) is harmonised with a network of longitudinal studies of ageing from across the globe (TILDA 2019), the potential relevance of our findings to these studies is certainly enhanced. Nevertheless, it is highly unlikely that a ‘one-size-fits-all’ solution will be identifiable (Arrow 2012; Greco et al. 2019). Future research could usefully engage in similar evaluations in relation to representative datasets on older populations in other jurisdictions. What our analysis has very clearly illustrated though is the range of performances of the different approaches. For the extent of exclusion measures, this relates to the diverse R^2’values found when predicting the CASP-19 exclusion component, and the variability across approaches when identifying exclusion for the same individuals. But it also relates to the significant variation in the prevalence of social exclusion, as calculated by the application of the different thresholds in the sum-of-scores method, and by all six approaches when using the 10% threshold.

The variability of these results suggests that there is a critical need to consider how different approaches might impact measurement construction. On the one hand, more rigorous and sustained efforts to evaluate the performance of different approaches is certainly required. Ideally, a comparison across methodological techniques should be performed, with the results presented from alternative approaches to illustrate the sensitivity of findings to the choice of technique. On the other hand, and cognisant that such comparisons may not always be feasible, transparency regarding the implicit assumptions underlying the selection of an approach is needed. The majority of techniques included in this analysis, and their underlying weighting and aggregation processes, possess both advantages and disadvantages. It is likely that there will always be a tension to negotiate between a measure that is technically robust and efficient, and one that preserves the integrity and multidimensionality of the exclusion concept. Giving consideration to a research-informed conceptual framework, and the purpose of the analysis that the composite measure will support, will certainly help to inform this negotiation. Researchers need to state why the attributes and features of one technique were favoured over those of others. It is only by engaging in such practices that the necessary transparency and critical reflection will be embedded into composite measure construction for assessing the social exclusion of older people.

References

Ahrens, A., Hansen, C. B., & Schaffer, M. (2020). Lassopack: Model selection and prediction with regularised regression in Stata. *The Stata Journal, 20*(1), 176-235, doi:https://doi.org/10.1177/1536867X20909697.

Arrow, K. J. (2012). *Social choice and individual values* (Vol. 12): Yale University Press.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (pp. 507-547): University of Chicago Press.

Atkinson, A. B. (1998). Social exclusion, poverty and unemployment. In J. Hills (Ed.), *Exclusion, employment and Opportunity* (pp. 1-20). Centre for Analysis of Social Exclusion (CASE): London School of Economics and Political Science (LSE).

Barnes, M. (2019). *Social Exclusion in Great Britain: An Empirical Investigation and Comparison with the EU*: Routledge.

Barnes, M., Blom, A., Cox, K., & Lessof, C. (2006). The social exclusion of older people: Evidence from the first wave of the English Longitudinal Study of Ageing. http://www.ifs.org.uk/docs/odpm_social_exclusion.pdf. Accessed 17 November 2020.

Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators, 80*, 12-22, doi:https://dx.doi.org/10.1016%2Fj.ecolind.2017.03.056.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica, 80*(6), 2369-2429, doi:https://doi.org/10.3982/ECTA9626.

Best, K. B., Gilligan, J. M., Carrico, A. R., Donato, K. M., Ackerly, B. A., & Mallick, B. (2020). Random forest analysis of two household surveys can identify important predictors of migration in Bangladesh. *Journal of Computational Social Science*, 1-24, doi:https://doi.org/10.1007/s42001-020-00066-9.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32, doi:https://doi.org/10.1023/A:1010933404324.

Buddelmeyer, H., Leung, F., & Scutella, R. (2012). Educating oneself out of social exclusion. https://files.eric.ed.gov/fulltext/ED537490.pdf. Accessed 17 November 2020.

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika, 95*(3), 759-771, doi:https://doi.org/10.1093/biomet/asn034.

Clark, T. E. (2004). Can out-of-sample forecast comparisons help prevent overfitting? *Journal of Forecasting, 23*(2), 115-139, doi:https://doi.org/10.1002/for.904.

CSO (2016). Census of Population 2016. https://www.cso.ie/en/csolatestnews/presspages/2017/census2016summaryresults-part1/. Accessed 17 November 2020.

Dahlberg, L., & McKee, K. J. (2018). Social exclusion and well-being among older adults in rural and urban areas. *Archives of Gerontology and Geriatrics, 79*, 176-184, doi:https://doi.org/10.1016/j.archger.2018.08.007.

Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews, 32*(1), 7-34, doi:https://doi.org/10.1080/07474938.2012.690641.

Dell'Anno, R., & Amendola, A. (2015). Social exclusion and economic growth: An empirical investigation in European economics. *The Review of Income and Wealth, 61*(2), 274-301, doi:https://doi.org/10.1111/roiw.12096.

Donoghue, O. A., McGarrigle, C. A., & Kenny, R. A. (2019). Who's in the driver's seat? Impact on social participation and psychosocial wellbeing in adults aged 50 and over. *Transportation Research Part F: Traffic Psychology and Behaviour, 64*, doi:https://doi.org/10.1016/j.trf.2019.06.010.

European Commission (2020). A strong social Europe for just transitions. https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_20. Accessed 17 November 2020.

Ferraro, K. F., & Shippee, T. P. (2009). Aging and Cumulative Inequality: How Does Inequality Get Under the Skin? *The Gerontologist, 49*(3), 333-343, doi:https://doi.org/10.1093/geront/gnp034.

Fonti, V., & Belitser, E. (2017). Feature selection using LASSO. *Business Analytics*.

García, E. C., Sánchez, A., & Navarro-Pabsdorf, M. (2019). Assessing the performance of the least developed countries in terms of the Millennium Development Goals. *Evaluation and Program Planning, 72*, 54-66, doi:https://doi.org/10.1016/j.evalprogplan.2018.09.009.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research, 141*(1), 61-94, doi:https://doi.org/10.1007/s11205-017-1832-9.

Hayes, T., Usami, S., Jacobucci, R., & McArdle, J. J. (2015). Using Classification and Regression Trees (CART) and random forests to analyze attrition: Results from two simulations. *Psychology and Aging, 30*(4), 911-929, doi:https://doi.org/10.1037/pag0000046

Hyde, M., Higgs, P., Wiggins, R. D., & Blane, D. (2015). A decade of research using the CASP scale: key findings and future directions. *Aging & mental health, 19*(7), 571-575, doi:https://doi.org/10.1080/13607863.2015.1018868.

Hyde, M., Wiggins, R. D., Higgs, P., & Blane, D. B. (2003). A measure of quality of life in early old age: The theory, development and properties of a needs satisfaction model (CASP-19). *Aging & mental health, 7*(3), 186-194, doi:https://psycnet.apa.org/doi/10.1080/1360786031000101157.

Kearney, P. M., Cronin, H., O'Regan, C., Kamiya, Y., Savva, G. M., Whelan, B., et al. (2011). Cohort profile: The Irish Longitudinal Study on Ageing. *International Journal of Epidemiology, 40*(4), 877-884, doi:https://doi.org/10.1093/ije/dyr116.

Kneale, D. (2012). Is social exclusion still Important for older people? https://ilcuk.org.uk/is-social-exclusion-still-important-for-older-people/. Accessed 17 November 2020.

Kolenikov, S., & Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth, 55*(1), 128-165, doi:https://doi.org/10.1111/j.1475-4991.2008.00309.x.

Levitas, R., Pantazis, C., Fahmy, E., Gordon, D., Lloyd, E., & Patsios, D. (2007). The multi-dimensional analysis of social exclusion. https://dera.ioe.ac.uk/6853/1/multidimensional.pdf. Accessed 17 November 2020.
Lewis, R. J. An introduction to classification and regression tree (CART) analysis. In *Annual meeting of the society for academic emergency medicine in San Francisco, California, 2000* (Vol. 14)
Loh, W. Y. (2014). Fifty years of classification and regression trees. *International Statistical Review, 82*(3), 329-348, doi:https://doi.org/10.1111/insr.12016.
Luo, Y., Hawkley, L. C., Waite, L. J., & Cacioppo, J. T. (2012). Loneliness, health, and mortality in old age: a national longitudinal study. *Social Science & Medicine, 74*(6), 907-914, doi:https://doi.org/10.1016/j.socscimed.2011.11.028.
MacLeod, C. A., Ross, A., Sacker, A., Netuveli, G., & Windle, G. (2019). Re-thinking social exclusion in later life: a case for a new framework for measurement. *Ageing and Society, 39*(1), 74-111, doi:https://doi.org/10.1017/S0144686X17000794.
Mazziotta, M., & Pareto, A. (2016). Methods for Constructing Non-Compensatory Composite Indices: A Comparative Study AU. *Forum for Social Economics, 45*(2-3), 213-229, doi:https://doi.org/10.1080/07360932.2014.996912.
Mohnen, S. M., Rotteveel, A. H., Doornbos, G., & Polder, J. J. (2020). Healthcare expenditure prediction with neighbourhood variables – A random forest model. *Statistics, Politics and Policy, 1*(ahead-of-print), doi:https://doi.org/10.1515/spp-2019-0010.
Nazroo, J. (2017). Class and health inequality in later life: patterns, mechanisms and implications for policy. *International Journal of Environmental Health, 14*(12), 1533, doi:https://doi.org/10.3390/ijerph14121533.
OECD (2008). Handbook on constructing composite indicators: Methodology and user guide. Organisation for Economic Co-Operation and Development.
Pantazis, C., Gordon, D., & Levitas, R. (2006). *Poverty and social exclusion in Britain: The millennium survey*: Policy Press.
Prattley, J., Buffel, T., Marshall, A., & Nazroo, J. (2020). Area effects on the level and development of social exclusion in later life. *Social Science & Medicine, 246*, 112722, doi:https://doi.org/10.1016/j.socscimed.2019.112722.
Radford, J., & Joseph, K. (2020). Theory In, theory out: The uses of social theory in machine learning for Social Science. [Hypothesis and Theory]. *Frontiers in Big Data, 3*(18), doi:https://doi.org/10.3389/fdata.2020.00018.
Sacker, A., Ross, A., MacLeod, C. A., Netuveli, G., & Windle, G. (2017). Health and social exclusion in older age: evidence from Understanding Society, the UK household longitudinal study. *Journal of Epidemiology and Community Health, 71*(7), 681-690, doi:https://www.researchgate.net/publication/313968911_Health_and_social_exclusion_in_older_age_Evidence_from_Understanding_Society_the_UK_household_longitudinal_study.
Scharf, T., & Keating, N. (Eds.). (2012). *From exclusion to inclusion in old age: A global challenge*. Bristol: Policy Press at the University of Bristol.
Scharf, T., Phillipson, C., & Smith, A. E. (2005). Multiple exclusion and quality of life amongst excluded older people in disadvantaged neighbourhoods. https://aran.library.nuigalway.ie/bitstream/handle/10379/3209/SEU_Final_Report.pdf?sequence=1&isAllowed=y. Accessed 17 November 2020.
Scharf, T., Phillipson, C., & Smith, A. E. (2005). Social exclusion of older people in deprived urban communities of England. [journal article]. *European journal of ageing, 2*(2), 76-87, doi:https://doi.org/10.1007/s10433-005-0025-6.

Scutella, R., Wilkins, R., & Kostenko, W. (2009). Estimates of poverty and social exclusion in Australia: A multidimensional approach. Melbourne Institute Working Paper No. 26/09.

Scutella, R., Wilkins, R., & Kostenko, W. (2013). Intensity and persistance of individuals social exclusion in Australia. *The Australian Journal of Social Issues, 48*(3), 273-298, doi:https://doi.org/10.1002/j.1839-4655.2013.tb00283.x.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, 58*(1), 267-288, doi:http://www.jstor.org/stable/2346178.

TILDA (2019). Harmonisation: Gateway to global aging data. https://tilda.tcd.ie/data/data-harmonisation/. Accessed 17 November 2020.

van Bergan, A., Hoff, S., van Loon, A., & van Hemert, A. (2017). Social Exclusion Index-for Health Surveys (SEI-HS): A prospective nationwide study to extend and validate a multidimensional social exclusion questionnaire. *BMC public health, 17*(1), doi:https://doi.org/10.1186/s12889-017-4175-1.

Van Regenmortel, S., De Donder, L., Dury, S., Smetcoren, A.-S., De Witte, N., & Verté, D. (2016). Social exclusion in later life: A systematic review of the literature. *Journal of Population Ageing, 9*(4), 315-344, doi:https://doi.org/10.1007/s12062-016-9145-3.

Van Regenmortel, S., De Donder, L., Smetcoren, A.-S., Lambotte, D., De Witte, N., & Verté, D. (2017). Accumulation of disadvantages: Prevalance and categories of old-age social exclusion in Belgium. *Social Indicators Research, 140*(3), 1-22, doi:https://doi.org/10.1007/s11205-017-1817-8.

Vrooman, J., & Hoff, S. (2013). The disadvantage among the Dutch: A survey approach to the multidimensional measurement of social exclusion. *Social Indicators Research, 113*(3), 1261-1287, doi:https://doi.org/10.1007/s11205-012-0138-1.

Walker, A., & Zaidi, A. (2019). Strategies of Active Ageing in Europe. In A. Walker (Ed.), *The Future of Ageing in Europe: Making an Asset of Longevity* (pp. 29-52). Singapore: Springer Singapore.

Walsh, K., Scharf, T., & Keating, N. (2017). Social exclusion of older persons: a scoping review and conceptual framework. [journal article]. *European journal of ageing, 14*(1), 81-98, doi:https://doi.org/10.1007/s10433-016-0398-8.

Ward, P., Walsh, K., & Scharf, T. (2014). Measuring old-age social exclusion in a cross-border context. https://bit.ly/3pEd2nT. Accessed 17 November 2020.

Whelan, B. J., & Savva, G. M. (2013). Design and methodology of The Irish Longitudinal Study on Ageing. *Journal of the American Geriatrics Society, 61*(2), 265-S268, doi:https://doi.org/10.1111/jgs.12199.

WHO (2015). World Report on Ageing and Health. http://apps.who.int/iris/bitstream/10665/186463/1/9789240694811_eng.pdf?ua=1. Accessed 17 November 2020.

WHO (2017). Towards a decade of healthy ageing (2020 - 2030). https://www.who.int/ageing/WHO-ALC-10-priorities.pdf. Accessed 17 November 2020.

Wiggins, R. D., Higgs, P. F. D., Hyde, M., & Blane, D. B. (2004). Quality of life in the third age: key predictors of the CASP-19 measure. *Ageing and Society, 24*(5), 693-708, doi:https://doi.org/10.1017/S0144686X04002284.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association, 101*(476), 1418-1429, doi:https://doi.org/10.1198/016214506000000735.