

Advancing Extended Redundancy Analysis via Generalized Estimating Equations (ERA-GEE): Integrating Model-Based Recursive Partitioning (MOB) for Subgroup Detection

By

Sam Sichkaruk

A Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
of the University of Manitoba
in partial fulfilment of the requirements of the degree of

Master of Arts

Department of Psychology

University of Manitoba

Winnipeg

Copyright © 2026 by Sam Sichkaruk

Abstract

Extended redundancy analysis (ERA) combined with generalized estimating equations (GEE)—termed ERA-GEE—provides a robust framework for analyzing high-dimensional predictors and multivariate outcomes in large-scale observational studies. ERA reduces high-dimensional predictors into low-dimensional components that maximize variance explained in the outcome variable(s), while GEE accounts for within-subject correlations across repeated or multivariate outcomes. A key limitation of ERA-GEE is its assumption that all parameters are homogeneous across individuals, which may not hold when associations vary across sociodemographic subgroups. To address this, model-based recursive partitioning (MOB) was integrated into the ERA-GEE framework (termed MOB-ERA-GEE), enabling data-driven identification of subgroups with distinct parameter estimates. Simulation studies evaluated the performance of the method across varying conditions, including partitioning covariate type, subgroup balance, sample size, and magnitude of parameter heterogeneity. Performance was evaluated using Type I error, power, classification accuracy, and subgroup size recovery. Results from the simulation demonstrate that MOB-ERA-GEE reliably detects subgroup heterogeneity, and highlighted under which conditions optimal performance occurred. The practical utility of MOB-ERA-GEE was illustrated using data from the Canadian Longitudinal Study on Aging to examine heterogeneity in the social isolation–health relationship. Social isolation was measured using 29 interrelated predictors, summarized into three theoretically defined components, and examined in relation to executive functioning and grip strength. Results from the application demonstrate that MOB-ERA-GEE yields subgroup-specific insights into the social isolation–health relationship that conventional modelling approaches may overlook. These findings highlight MOB-ERA-

GEE as a flexible, data-driven framework for uncovering population heterogeneity in high-dimensional observational data.

Keywords: extended redundancy analysis, model-based recursive partitioning, multivariate methods, data-driven subgroup identification, Canadian Longitudinal Study on Aging

Acknowledgements

I would like to sincerely thank my advisor, Dr. Kim, for her exceptional guidance, encouragement, and dedication throughout this thesis. Her insightful feedback and thoughtful mentorship not only strengthened this work but also profoundly shaped my growth as a researcher. I deeply appreciate the time, patience, and care she invested in both this project and my academic development.

I would also like to thank my committee members, Dr. Lix and Dr. Kruk, for their careful attention, insightful suggestions, and constructive critiques. Their guidance provided invaluable direction throughout this journey and was instrumental in shaping the rigor and depth of this research.

I gratefully acknowledge the University of Manitoba Graduate Fellowship (UMGF) and the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support. This funding made it possible for me to pursue this research with focus and dedication.

Finally, my deepest gratitude belongs to my parents. Their unwavering love, encouragement, and steadfast belief in my abilities have been the foundation of my academic journey. Their patience, guidance, and constant support have carried me through challenges, celebrated my successes, and made this achievement possible.

Table of Contents

Abstract	2
Acknowledgements	4
List of Tables	7
List of Figures	8
Advancing Extended Redundancy Analysis via Generalized Estimating Equations (ERA-GEE): Integrating Model-Based Recursive Partitioning (MOB) for Subgroup Detection	9
Managing Multicollinearity in Observational Data	9
Modelling Correlated Outcomes in Observational Research.....	12
Extended Redundancy Analysis with Generalized Estimating Equations (ERA-GEE).....	14
Model-Based Recursive Partitioning for Subgroup Detection	15
Extending ERA-GEE to Detect Subgroup Heterogeneity	17
Methods.....	18
ERA-GEE	18
Model Specification	18
Parameter Estimation and Significance Testing	20
Model-Based Recursive Partitioning	22
Step 1: Global Model Estimation.....	22
Step 2: Testing for Parameter Instability	23
Step 3: Partitioning the Data Based on Detected Instability.....	24
Step 4: Recursive Partitioning Until Model Stability is Achieved	24
Pruning Strategies to Control Overfitting and Enhance Model Interpretability	25
Integrating MOB into ERA-GEE.....	26
Simulation Study.....	29
Methods.....	29
Data Generation and Model Structure	30
Partitioning Covariates and Subgroup Structure.....	32
Subgroup Balance and Parameter Heterogeneity	33
Performance Metrics	34
Pilot Simulation and Rationale for Final Design	36
Results.....	37
Binary True Covariate Condition.....	37
Continuous True Covariate Condition	39
Binned Continuous True Covariate Condition	41

Empirical Application	42
Methods.....	43
Data and Study Population.....	43
Measures	45
Social Isolation.....	45
Handgrip Strength.....	46
Executive Functioning.....	46
Potential Covariates.....	47
Statistical Analysis.....	48
Results.....	48
Baseline Characteristics	48
Parameter Instability and Subgroup Identification	48
Heterogeneity in the Composition of Social Isolation.....	49
Differential Associations Between Social Isolation Components and Health Outcomes.....	50
Model Complexity and Cross-Validation Results	51
Discussion	51
Simulation	52
Empirical Application.....	55
Strengths and Limitations	60
Future Research Directions.....	62
Conclusion	63
References.....	64

List of Tables

Table 1 Type I Error, Power, Cramér's V, and Subgroup Size Recovery by Sample Size and Parameter Instability (Binary Covariate Condition)	75
Table 2 Type I Error, Power, Cramér's V, and Subgroup Size Recovery by Sample Size and Parameter Instability (Continuous Covariate Condition)	77
Table 3 Type I Error, Power, Cramér's V, and Subgroup Size Recovery by Sample Size and Parameter Instability (Binned Continuous Covariate Condition)	79
Table 4 Variables and Summary Statistics for the CLSA Comprehensive Cohort (Baseline)	81
Table 5 Parameter Instability Test Results from MOB-ERA-GEE for the CLSA Comprehensive Cohort (Baseline)	86
Table 6 Subgroup-Specific Component Weight Estimates from MOB-ERA-GEE for the CLSA Comprehensive Cohort (Baseline)	87
Table 7 Subgroup-Specific Regression Coefficient Estimates from MOB-ERA-GEE for the CLSA Comprehensive Cohort (Baseline)	89

List of Figures

Figure 1 Conceptual Diagram of the ERA-GEE Model Specified in the Simulation Study	90
Figure 2 Computational Performance of MOB-ERA-GEE Across Sample Size and Covariate Conditions	91
Figure 3 Final MOB-ERA-GEE Model Specified for the CLSA Comprehensive Cohort (Baseline) Data.....	92
Figure 4 Comparison of 10-Fold Cross-Validation Loss Across MOB-ERA-GEE Tree Depths	93

Advancing Extended Redundancy Analysis via Generalized Estimating Equations (ERA-GEE): Integrating Model-Based Recursive Partitioning (MOB) for Subgroup Detection

From public health to psychology, observational research plays a central role in understanding human behaviour and societal processes without environmental manipulation. Data derived from such research are often high-dimensional, comprising multiple interrelated variables that would benefit from simultaneous investigation. Although conventional statistical approaches—such as multiple and multivariate regression—remain widely used for this purpose, they face substantial limitations when applied to complex observational data. In particular, these methods are challenged by multicollinearity among predictors, dependence among multiple outcomes, and the identification of subgroup-specific effects when numerous potential moderators—such as age or gender—interact in complex and often unknown ways. To address these limitations, this research develops a data-driven multivariate framework that simultaneously accounts for predictor multicollinearity, models correlated outcomes, and detects subgroup-specific effects. The performance of the proposed framework is evaluated through a series of Monte Carlo simulations, and its practical utility is demonstrated through an empirical application to large-scale observational data from the Canadian Longitudinal Study on Aging (CLSA).

Managing Multicollinearity in Observational Data

One persistent challenge in observational research is isolating the unique contribution of each predictor in the presence of substantial multicollinearity. Multicollinearity occurs when two or more predictors are highly correlated, leading to inflated standard errors, unstable coefficient estimates, and reduced interpretability of each variable's unique contribution to the outcome (Hocking & Pendleton, 1983; Vatcheva et al., 2016). This problem is especially pronounced in

psychological and social science research, where complex constructs are often operationalized using multiple, interrelated predictors rather than a single variable.

A clear example of this arises in the CLSA, a large-scale observational cohort study of more than 50,000 Canadian adults aged 45 years or older (Raina et al., 2009). Within the CLSA, social isolation is conceptualized as a multidimensional construct and assessed using multiple interrelated indicators, including marital status, social network size, and frequency of social interaction (Wister et al., 2019). Although each indicator captures a distinct yet overlapping aspect of social isolation, their joint consideration provides a more comprehensive representation of the underlying construct. However, including such correlated predictors simultaneously in conventional regression frameworks can lead to multicollinearity. This increases the risk of unstable estimates and may obscure how shared information across related variables contributes to outcome prediction.

To address multicollinearity, researchers are often faced with a trade-off between arbitrarily excluding highly correlated predictors—thereby risking the loss of substantively meaningful information—or collapsing them into a single summary measure (e.g., by averaging, summing, or applying fixed weights). An illustrative example is the CLSA Social Isolation Index (CLSA-SII), developed by Wister et al. (2019) and grounded in the multidimensional framework proposed by Valtorta et al. (2016). The CLSA-SII captures three distinct dimensions of social isolation—structural objective, functional objective, and functional subjective—by aggregating multiple indicators within each domain. Although this index has been widely adopted and offers a convenient summary of social isolation, its standard scoring procedure assigns equal weight to all indicators within a domain and equal weight across domains through simple averaging. This approach simplifies subsequent analyses but implicitly assumes that all items and domains

contribute equally to the construct—an assumption that may not reflect their relative importance for predicting specific outcomes.

Rather than relying on summary scores that impose fixed and potentially arbitrary weighting schemes, multicollinearity can be more effectively addressed through data-driven dimension reduction approaches that retain meaningful information while minimizing redundancy. Extended Redundancy Analysis (ERA) offers such a framework by integrating dimension reduction and regression within a unified modelling approach (Takane & Hwang, 2005). Specifically, ERA extracts low-dimensional components from sets of correlated predictors in a supervised manner, such that the resulting components are constructed to maximize explained variance in the outcome variable(s). By estimating weights within predefined sets of conceptually related predictors from the data rather than specifying them a priori, ERA yields components that remain interpretable at the construct level while prioritizing outcome-relevant information over structure driven solely by correlations among predictors (Kim, 2020). By preserving essential predictive information while mitigating multicollinearity, ERA is particularly well-suited for modelling multidimensional constructs in high-dimensional observational research. Figure 1 provides a representation of the ERA framework, illustrating how predefined sets of predictors are reduced into weighted components that are linked to multiple outcomes.

Importantly, although ERA summarizes each set of correlated variables into a weighted composite, it does not constitute a latent variable modelling framework such as factor analysis or structural equation modelling. Unlike approaches that aim to recover an underlying covariance structure among indicators for confirmatory modelling of latent constructs or causal pathways, ERA performs supervised dimension reduction, in which component weights are estimated

conditional on the observed outcome variable(s). As a result, the extracted components are not treated as latent variables: there is no associated measurement error model and no attempt to recover a “true” underlying factor structure. Instead, the components are outcome-oriented summaries of correlated predictors, and their interpretability is driven by predictive relevance rather than shared variance alone. Accordingly, ERA is best understood as a multivariate prediction and association framework, rather than a latent variable or causal pathway modelling approach.

Modelling Correlated Outcomes in Observational Research

Beyond multicollinearity among predictors, many substantive research questions require the simultaneous analysis of multiple, potentially correlated outcomes. Conventional approaches, such as multivariate linear regression, are commonly used for this purpose; however, they rely on strong distributional assumptions that are often difficult to satisfy in practice. In particular, these methods assume multivariate normality, an assumption that is frequently violated in real-world observational data due to skewed distributions, heteroscedasticity, or the presence of categorical or count outcomes (Cain et al., 2017). Moreover, multivariate linear regression requires researchers to correctly specify the within-subject correlation structure among outcomes (e.g., independence, exchangeable, autoregressive, or unstructured). In large-scale observational studies, where relationships among outcomes may be complex or poorly understood, the true covariance structure is typically unknown. Misspecification of this structure can therefore reduce statistical efficiency and compromise inference, particularly when standard errors are estimated under an incorrect dependence structure (Kim et al., 2020).

Generalized Estimating Equations (GEE) provide a flexible alternative for modelling correlated outcomes without requiring strict distributional assumptions regarding the joint

distribution of the data or the full covariance structure among outcomes (Liang & Zeger, 1986).

In contrast to conventional multivariate approaches that typically rely on maximum likelihood estimation and multivariate normality assumptions, GEE is based on quasi-likelihood estimation and can accommodate a wide range of outcome types, including binary, ordinal, skewed continuous, and count data (Kim et al., 2020). This flexibility enables the simultaneous analysis of multiple correlated outcomes within a unified modelling framework, while allowing each outcome to follow an appropriate marginal distribution.

Although GEE provides a flexible framework for modelling correlated outcomes, it is not the only approach available for such data. Generalized linear mixed models (GLMMs), for example, are also widely used, and the choice between these frameworks depends not only on the structure of the data, but also on the substantive inferential objective of the analysis. A key distinction between these approaches lies in the nature of the effects being estimated. Whereas GEE produces population-averaged (marginal) parameter estimates that describe the average association between predictors and outcomes in the target population, GLMMs estimate subject- or cluster-specific (conditional) effects by incorporating random effects to account for within-subject or within-cluster dependence and unobserved heterogeneity (Hu et al., 1998; McCulloch et al., 2008). Because the present research aims to estimate population-averaged associations in multivariate observational data, rather than subject-specific trajectories or cluster-level random effects, the marginal interpretation provided by GEE aligns directly with the substantive and methodological aims of the analysis.

The distinction between GEE and GLMMs also extends to their modelling assumptions and estimation strategies. Whereas GLMMs require specification of a full conditional likelihood and distributional assumptions for the random effects—typically Gaussian—GEE requires only a

correctly specified marginal mean structure and a working correlation matrix to account for within-subject or within-cluster dependence (Hu et al., 1998; Liang & Zeger, 1986). Importantly, even when the working correlation structure is misspecified, GEE yields consistent parameter estimates and valid standard errors through the use of robust sandwich variance estimators (Hardin & Hilbe, 2013; Liang & Zeger, 1986). These properties make GEE particularly advantageous in observational settings where the primary objective is population-level inference, and where outcome dependence may be complex or only partially known.

Extended Redundancy Analysis with Generalized Estimating Equations (ERA-GEE)

The integration of ERA with GEE (Kim et al., 2020; Lee et al., 2019) yields a flexible multivariate regression framework that simultaneously addresses predictor multicollinearity and dependence among multiple outcomes. This approach enables interpretable, population-averaged inference in high-dimensional settings with correlated predictors and outcomes, while avoiding restrictive assumptions about the joint distribution of the data. Despite these advantages, a key limitation of the ERA-GEE framework is its assumption that all model parameters are homogeneous across individuals in the sample. In many applied domains—particularly psychology, public health, and aging research—this assumption is often unrealistic, as the magnitude and direction of associations between variables may vary substantially across population subgroups defined by factors such as age, sex, or socioeconomic status (Braveman & Gottlieb, 2014). Although large-scale observational datasets typically include substantial heterogeneity across such subgroups, ERA-GEE in its current form estimates a single global model that is assumed to apply uniformly across the sample. As a result, meaningful subgroup-specific effects may be obscured—particularly for smaller or historically marginalized populations—potentially limiting both the interpretability and generalizability of the findings.

This limitation motivates the development of an extension to ERA-GEE that allows parameter estimates to vary across empirically identified subgroups.

Model-Based Recursive Partitioning for Subgroup Detection

One promising approach for incorporating subgroup detection into the ERA-GEE framework is model-based recursive partitioning (MOB; Zeileis et al., 2008). MOB is considered a data-driven method in which subgroup identification and splitting decisions are guided directly by empirical patterns in the data, rather than by theoretical assumptions, predefined cut-offs, or researcher-imposed thresholds. Specifically, MOB recursively partitions the data by testing for structural changes in model parameters, thereby identifying subgroups in which these parameters differ significantly across values of candidate partitioning variables (covariates).

To ensure that parameter instability is statistically meaningful, MOB employs score-based fluctuation tests, ensuring that subgroup splits are supported by empirical evidence rather than arbitrary distinctions (Zeileis et al., 2008). This recursive procedure results in a tree structure that visually represents the hierarchy of covariates selected for partitioning, thereby illustrating how covariates interact to explain heterogeneity in model parameters (Kim & Hwang, 2021). These characteristics make MOB particularly well-suited for complex observational settings, where effect heterogeneity is common but difficult to anticipate due to the large number of potential covariates and their interactions.

Although MOB produces a tree structure that is visually similar to classical decision tree methods such as Classification and Regression Trees (CART; Breiman et al., 1984), the two approaches differ fundamentally in both their objectives and splitting criteria. Specifically, classical decision trees partition data to maximize homogeneity of the outcome variable within subgroups, typically by minimizing prediction error or impurity (Breiman et al., 1984). In

contrast, MOB identifies partitions based on structural changes in model parameters, recursively testing whether predictor–outcome relationships vary across values of candidate partitioning variables (Zeileis et al., 2008). Subgroup formation in MOB is therefore driven by heterogeneity in estimated relationships rather than by similarity in observed outcome values alone. This distinction is particularly important in the present context, as the objective of this research is to identify subgroups characterized by differences in multivariate predictor–outcome associations, rather than simply to maximize predictive accuracy of the outcome itself.

Due to its ability to detect structural changes in parameter estimates, MOB has been successfully integrated into a wide range of commonly used statistical frameworks, including linear and logistic regression models (Zeileis et al., 2008), as well as generalized linear mixed models (Fokkema et al., 2025). MOB has also been extended to more complex modelling frameworks, including ERA with a single outcome (Kim & Hwang, 2021). In the univariate ERA setting, the model assumes normally distributed errors, and parameter estimation is carried out by maximizing the log-likelihood function using the Iteratively Reweighted Least Squares algorithm. Following convergence, empirical scores—defined as individual-level contributions to the score function—are obtained as the first derivatives of the log-likelihood with respect to the model parameters, evaluated at their estimated values. These empirical scores form the basis of the MOB parameter instability tests, which assess whether statistically significant fluctuations occur across different levels of candidate partitioning covariates.

When evidence of parameter instability is detected, the covariate exhibiting the strongest instability at the current partitioning stage is selected, and its optimal split point—defined as the value that maximizes score fluctuation—is used to divide the sample into subgroups. This recursive partitioning process is then repeated within each resulting subgroup until no further

instability is detected or a prespecified stopping criterion is met. Upon completion of the partitioning process, separate ERA models are fitted within each terminal subgroup, enabling subgroup-specific estimation of predictor–outcome relationships.

Although Kim and Hwang’s (2021) MOB-ERA framework demonstrates how MOB can identify subgroup heterogeneity in a data-driven manner, its current application remains limited to univariate outcomes. Given that many applied research questions involve multiple, potentially correlated outcomes, extending this framework to a multivariate setting addresses an important methodological gap.

Extending ERA-GEE to Detect Subgroup Heterogeneity

The present thesis addresses this gap by integrating MOB into the ERA-GEE framework, yielding the proposed MOB-ERA-GEE approach. This extension allows for the data-driven identification of subgroups with distinct multivariate predictor–outcome associations, while preserving the core strengths of ERA-GEE—namely, its ability to accommodate high-dimensional predictors and correlated outcomes without strict distributional assumptions. By combining these frameworks, MOB-ERA-GEE enables a more nuanced examination of heterogeneity in complex observational data and produces interpretable, tree-based visualizations that reveal how predictor–outcome relationships vary across subgroups.

Accordingly, the present thesis is guided by the following methodological and applied research questions: (1) Can ERA-GEE be extended to incorporate data-driven subgroup detection while preserving its multivariate, population-averaged interpretations? (2) Under what conditions—such as sample size, subgroup balance, and magnitude of parameter heterogeneity—does MOB-ERA-GEE reliably detect parameter instability and recover true subgroup structures? (3) How does the performance of MOB-ERA-GEE vary across different types of partitioning

covariates (binary, continuous, and binned continuous) in terms of statistical accuracy and computational feasibility? (4) In applied settings, can MOB-ERA-GEE reveal meaningful heterogeneity in predictor–outcome relationships that a homogeneous ERA-GEE model would overlook?

The remainder of this thesis is organized to reflect its primary methodological focus rather than the conventional Introduction–Methods–Results–Discussion format commonly used in empirical research. The first major section establishes the methodological motivation and theoretical foundations of the proposed approach, beginning with a detailed specification of the ERA-GEE framework and its integration with MOB to enable data-driven subgroup detection. A comprehensive simulation study then evaluates the performance of MOB-ERA-GEE under controlled data-generating conditions, serving as methodological validation rather than substantive inference. An empirical application using data from the CLSA subsequently demonstrates the practical utility of the framework in a real-world setting. The final section synthesizes the findings, discusses methodological implications and limitations, and outlines directions for future methodological development and applied research.

Methods

ERA-GEE

Model Specification

In ERA-GEE (Kim et al., 2020; Lee et al., 2019), we assume that there are Q outcome variables and K different predictor sets, each consisting of P_k predictors, where $k = 1, \dots, K$ and $P = \sum_{k=1}^K P_k$. Let y_{iq} denote the value of the q th outcome variable for the i th subject, where $i = 1, \dots, N$ and $q = 1, \dots, Q$. We assume that y_{iq} follows an exponential family distribution with a mean μ_i and variance $\phi\sigma_{iq}^2$, where ϕ is a dispersion parameter. Let x_{ikp} denote the value of the

p th predictor within the k th predictor set for the i th subject, where $p = 1, \dots, P_k$ and $k = 1, \dots, K$. Let w_{kp} represent the component weight assigned to x_{ikp} , and let $f_{ik} = \sum_{p=1}^{P_k} x_{ikp} w_{kp}$ denote the i th component score for the k th predictor set, which is the sum of the weighted predictors for the i th observation in the k th predictor set. Let β_{kq} denote the regression coefficient relating the k th component to the q th outcome variable. Let η_{iq} denote the linear predictor for the i th subject and the q th outcome, and $g(\cdot)$ represent the link function. The ERA-GEE model is then specified as:

$$g(\mu_{iq}) = \eta_{iq} = \sum_{k=1}^K \left[\sum_{p=1}^{P_k} x_{ikp} w_{kp} \right] \beta_{kq} = \sum_{k=1}^K f_{ik} \beta_{kq}$$

indicating that the marginal expectation of the outcome η_{iq} is related to a linear combination of the extracted component scores through the known link function.

Let $\tilde{y}_i = [y_{i1}, \dots, y_{iQ}]'$ denote the Q -dimensional vector of outcome variables for the i th subject. Let Σ_i be the $Q \times Q$ within-subject covariance matrix of $\tilde{y}_i$. To accommodate the unknown and potentially complex correlation structure among multiple outcome variables, the ERA-GEE model incorporates GEE (Liang & Zeger, 1986), which allows for the estimation of model parameters without requiring a fully specified joint distribution. Specifically, the covariance structure of $\tilde{y}_i$ is left unspecified and instead approximated using a working covariance matrix, as follows

$$\text{Cov}(\tilde{y}_i) = \Sigma_i = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2}$$

where ϕ is the common dispersion parameter, A_i is a diagonal matrix containing the marginal variances, with σ_{iq}^2 as its q th diagonal element, and $R(\alpha)$ is a $Q \times Q$ working correlation matrix, fully specified by a vector of nuisance parameters α . The working correlation matrix is assumed to remain constant across all subjects, although the marginal variances may vary across outcomes and individuals through A_i .

The interpretation of $R(\alpha)$ depends on both the structure and the measurement scale of the correlated outcomes. In a multivariate outcome setting, where outcomes represent distinct but potentially related constructs measured within the same individual, $R(\alpha)$ reflects within-subject association among outcomes. In a longitudinal repeated-measures setting, where outcomes represent repeated observations of the same construct over time, it reflects temporal dependence across measurement occasions. For continuous outcomes, the working correlation can be understood as the association among outcome values after accounting for the marginal mean structure. For binary or other non-Gaussian outcomes, it is better viewed as a working approximation to dependence among marginal responses, rather than as a full description of an underlying multivariate distribution. Thus, although the same ERA-GEE formulation can accommodate both multivariate and longitudinal data structures, the choice of working correlation structure (e.g., independence, exchangeable, autoregressive, or unstructured) should be guided by the substantive objectives of the study and the anticipated pattern of dependence among the outcomes. This working correlation structure provides a flexible and computationally efficient way to account for within-subject dependencies, while treating the true covariance structure as a nuisance component that need not be explicitly modelled for consistent estimation of the marginal mean parameters, provided that the mean model is correctly specified (Kim et al., 2020; Lee et al., 2019; Liang & Zeger, 1986).

Parameter Estimation and Significance Testing

Let $\tilde{x}_i$ denote the $P \times 1$ stacked vector of predictors for subject i , and let $\tilde{z}_i$ be the $Q \times 1$ vector of adjusted outcome variables. The component weights are collected in the matrix $\mathbf{W} \in \mathbb{R}^{P \times K}$, and the regression coefficients in the matrix $\mathbf{B} \in \mathbb{R}^{K \times Q}$. The parameter estimation process in ERA-GEE follows an iterative approach to estimate the ERA parameters (both $\mathbf{W}$ and $\mathbf{B}$) and

the nuisance covariance parameters (α and ϕ). This is accomplished by minimizing the penalized least squares criterion given by

$$\varphi_{\alpha, \mathbf{W}, \mathbf{B}} = \sum_{i=1}^N [(\tilde{z}_i - \mathbf{B}'\mathbf{W}\tilde{x}_i)' \Sigma_i^{-1} (\tilde{z}_i - \mathbf{B}'\mathbf{W}\tilde{x}_i)] + \lambda_W \text{trace}(\mathbf{W}'\mathbf{W}) + \lambda_B \text{trace}(\mathbf{B}'\mathbf{B})$$

where λ_W and λ_B represent the tuning parameters controlling the ridge penalties on the component weights and regression coefficients, respectively. The optimization is carried out using a regularized alternating least squares algorithm, where the estimates for $\mathbf{W}$, $\mathbf{B}$, and Σ_i are updated iteratively, each time holding the other parameters fixed. This iterative procedure continues until convergence to a stable solution. More detailed steps of the parameter estimation procedure can be found in Lee et al. (2019).

The statistical significance of the parameter estimates in ERA-GEE can be assessed using bootstrapping (Efron & Tibshirani, 1986), a resampling method that helps construct confidence intervals. In this approach, the dataset is repeatedly resampled with replacement to generate an empirical distribution of the parameter estimates. Percentile-based bootstrap confidence intervals (e.g., the 5th and 95th percentiles of the bootstrap distribution) are then used to quantify the uncertainty around each estimate. This method provides robust statistical inference by directly estimating the variability of the parameter estimates from the data, without requiring strong parametric assumptions about the underlying distributions.

In addition to bootstrapping, component-wise inference can also be conducted using a Wald test (Lee et al., 2019). To assess whether a given component k has a significant effect across all outcome variables, the following test statistic is used

$$T_k = \hat{\beta}_k^T \widehat{\text{Cov}}^{-1}(\hat{\beta}_k) \hat{\beta}_k,$$

where $\hat{\beta}_k$ is the $Q \times 1$ vector corresponding to the k th row of the regression coefficient matrix $\hat{\mathbf{B}} \in \mathbb{R}^{K \times Q}$, which links the K components to the Q outcomes. The covariance matrix of $\hat{\beta}_k$ can be estimated analytically or via permutation. Further details on the derivation and implementation of this test are provided in Lee et al. (2019).

Model-Based Recursive Partitioning

While ERA-GEE effectively reduces predictor dimensionality and accounts for correlated outcomes, it assumes that predictor–outcome relationships are homogeneous across all individuals. However, this assumption may not hold in practice, as these relationships can vary across subgroups due to underlying covariate-dependent heterogeneity. To account for such variability, MOB provides a data-driven approach that systematically detects and models subgroup-specific variations (Zeileis et al., 2008).

Step 1: Global Model Estimation

The MOB procedure begins by fitting a global parametric model to the entire dataset. Let Y_i represent the outcome variable for the i th subject, where $i = 1, \dots, N$, and let $\theta \in \Theta$ be a k -dimensional parameter vector. The model parameters are estimated by minimizing an objective function $\Psi(Y_i, \theta)$, which may be based on least squares (for regression) or the negative log-likelihood (for likelihood-based estimation). The estimated parameter vector, denoted as $\hat{\theta}$, is obtained as

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \Psi(Y_i, \theta).$$

At this stage, the model assumes that a single set of parameters applies to all observations in the dataset. To assess potential heterogeneity, MOB tests for parameter instability to determine whether subgroup-specific parameter estimates are necessary.

Step 2: Testing for Parameter Instability

To assess whether a single set of model parameters is appropriate for all observations, the MOB procedure evaluates parameter instability with respect to a set of partitioning variables Z_j , where $j = 1, \dots, \ell$, and ℓ is the number of partitioning variables under consideration. These variables—such as age, gender, or income level—are potential sources of subgroup heterogeneity. If the relationship between predictors and the outcome differs systematically across values of a partitioning variable, this suggests that subgroup-specific parameter estimates may be warranted.

To formally test for such heterogeneity, MOB applies score-based parameter instability tests. The score function, which measures how sensitive the objective function $\psi(Y_i, \theta)$ is to changes in the parameter vector θ is defined as

$$\psi(Y_i, \theta) = \frac{\partial \Psi(Y_i, \theta)}{\partial \theta}.$$

Under the assumption that the global estimate $\hat{\theta}$ holds uniformly across all observations, the score function values should exhibit random fluctuations around zero. However, if the score values display systematic patterns when ordered by a partitioning variable Z_j , this indicates potential subgroup differences and suggests that separate parameter estimates may be more appropriate.

To detect such deviations, MOB uses an empirical fluctuation process, which aggregates the score function values in an order determined by the partitioning variable Z_j . The empirical fluctuation process is given by

$$W_j(t) = \hat{f}^{-1/2} n^{-1/2} \sum_{i=1}^{[nt]} \hat{\psi} \sigma(z_{ij}),$$

where $\sigma(z_{ij})$ is the ordering permutation, which ranks observations by the partitioning variable, $\hat{f}$ is an estimate of the covariance matrix $COV(\psi(Y_i, \hat{\theta}))$, which standardizes the score process,

and $[nt]$ indexes the cumulative sum of score contributions up to fraction t of the sample. This process evaluates score contributions in an order defined by the partitioning variable Z_j rather than their original sequence in the dataset.

Under the null hypothesis of parameter stability, the fluctuation process $W_j(t)$ follows a Brownian bridge process, meaning that the score fluctuations are random and do not indicate subgroup differences (Zeileis et al., 2008). However, if significant deviations from this expected pattern occur, MOB detects parameter instability and selects the partitioning variable with the strongest instability for the next step in the procedure.

Step 3: Partitioning the Data Based on Detected Instability

Once a partitioning variable is selected, MOB determines the optimal split point that optimizes the split criterion associated with the detected parameter instability. The dataset is then divided into two subgroups, and the parametric model is re-estimated separately within each subgroup to account for subgroup-specific parameter differences. The new parameter estimates for each subgroup b , where $b = 1, \dots, B$ are given by

$$\hat{\theta}_b = \arg \min_{\theta \in \Theta} \sum_{i \in I_b} \Psi(Y_i, \theta),$$

where I_b represents the set of observations assigned to subgroup b . This segmentation process ensures that the model parameters remain homogeneous within each subgroup, allowing MOB to capture meaningful subgroup-specific variations in predictor–outcome relationships.

Step 4: Recursive Partitioning Until Model Stability is Achieved

The partitioning procedure outlined above proceeds recursively within each newly formed subgroup. At each stage of partitioning, the model is fitted once under the null hypothesis of parameter stability within each subgroup, and parameter instability is assessed using scores derived from this fit. If further heterogeneity is detected, the sample is split again based on the

partitioning variable showing the next strongest instability. This process continues until no additional significant instability is observed, or a predefined stopping criterion is met. Once the procedure terminates, the final result is a tree-like structure, where each terminal node represents a distinct population segment with its own parameter estimates.

Pruning Strategies to Control Overfitting and Enhance Model Interpretability

Although recursive partitioning enables MOB to detect meaningful heterogeneity in predictor–outcome relationships, unconstrained tree growth may cause the model to fit noise or sample-specific variation rather than stable, generalizable population patterns. To mitigate this risk, pruning strategies are used to regulate tree complexity by restricting unnecessary splits and removing branches that do not provide meaningful improvement in model fit.

In MOB, pre-pruning is implemented as an early stopping strategy that restricts tree growth during construction. At each candidate split, parameter instability tests are conducted to evaluate whether model parameters vary systematically across levels of a partitioning variable. These tests yield p-values assessing the null hypothesis of parameter stability. A split is permitted only if the p-value falls below a pre-specified significance threshold, which functions as a tuning parameter: smaller thresholds result in more conservative trees with fewer splits, whereas larger thresholds permit additional splits, capturing subtler heterogeneity but increasing the risk of overfitting. However, when analyzing large observational datasets, even small departures from parameter stability may produce statistically significant p-values due to increased statistical power. This highlights the importance of combining pre-pruning with post-pruning strategies to achieve a balance between sensitivity to heterogeneity and model interpretability (Kim & Hwang, 2021; Zeileis et al., 2008; Zeileis & Hothorn, 2022).

Post-pruning is applied after the tree has been fully grown and aims to simplify the tree by removing branches that do not meaningfully improve model performance. In MOB, two commonly used post-pruning approaches are information-criterion-based pruning and cross-validation-based pruning. Information-criterion-based pruning evaluates a fully grown tree using in-sample fit, assessing whether additional splits meaningfully improve the model while penalizing for increased complexity (Zeileis & Hothorn, 2022). In generalized estimating equations, the Quasi-Information Criterion (QIC) extends the Akaike Information Criterion to quasi-likelihood settings and is commonly used for model comparison (Pan, 2001). Cross-validation-based pruning, by contrast, evaluates candidate subtree sizes using out-of-sample prediction error on held-out folds, selecting the tree size that best generalizes to new data (Stone, 1974; Zeileis et al., 2008). Although these pruning approaches differ in implementation, they share a common objective: identifying the point at which additional splits no longer yield meaningful improvement in model fit or predictive performance. Employing multiple pruning strategies in combination therefore enhances the robustness and interpretability of the final tree structure.

Integrating MOB into ERA-GEE

The proposed MOB-ERA-GEE framework was developed by formally integrating ERA-GEE within a model-based recursive partitioning structure, such that ERA-GEE serves as the node-specific estimating model while instability tests are conducted on its empirical score contributions. This integration required adapting the ERA-GEE estimation procedure to satisfy the theoretical requirements of MOB, including the explicit derivation of observation-level score contributions and compatibility with parameter instability testing.

Within this unified framework, ERA-GEE defines the marginal mean structure and corresponding estimating equations for multivariate outcomes within each node, while MOB provides a statistically principled mechanism to detect covariate-dependent heterogeneity in these parameters. Importantly, this construction preserves the population-averaged interpretation of ERA-GEE parameters, as subgroup formation is driven by instability in the estimating equations. Consequently, model parameters are allowed to vary across empirically identified subgroups.

To satisfy the theoretical requirements of MOB, each node-specific ERA-GEE model must return three quantities: (1) the vector of estimated parameters, (2) a scalar objective function value, and, most importantly, (3) observation-level empirical score contributions with respect to those parameters (Zeileis et al., 2008). Accordingly, the ERA-GEE estimation procedure was adapted to explicitly compute and return subject-level score contributions after convergence. The recursive partitioning algorithm was operationalized using the “mob” function in the R package “partykit” (version 1.2-24; Zeileis & Hothorn, 2022), while the underlying estimating equations and score derivations were developed specifically for the ERA-GEE context.

Because parameter instability testing in MOB is grounded in empirical estimating functions, particular attention was given to deriving score contributions consistent with the ERA-GEE formulation. In general, ERA-GEE estimates parameters through (penalized) estimating equations, in which empirical scores correspond to the gradient of the estimating objective (including any penalty terms) with respect to the model parameters, evaluated at the final (converged) estimates, $\widehat{W}$ and $\widehat{B}$. However, to simplify the recursive partitioning procedure and facilitate a more transparent derivation of empirical scores, the present thesis employed an

unpenalized ERA-GEE specification, omitting ridge-type penalties on $\mathbf{W}$ and $\mathbf{B}$. Under this formulation, empirical score contributions are derived directly from the working loss associated with the marginal estimating equations.

For subject i , the individual contribution to the score function with respect to a parameter $\theta \in \{\mathbf{W}, \mathbf{B}\}$ is defined as:

$$u_i = \frac{\partial}{\partial \theta} [(\tilde{z}_i - \hat{\mathbf{B}}^T \hat{\mathbf{W}} \tilde{x}_i)' \boldsymbol{\Sigma}_i^{-1} (\tilde{z}_i - \hat{\mathbf{B}}^T \hat{\mathbf{W}} \tilde{x}_i)].$$

This empirical score vector quantifies the local sensitivity of the model fit for subject i . Within the recursive partitioning procedure, systematic fluctuations in these scores across values of candidate partitioning variables indicate parameter instability. Importantly, the score is computed after convergence, using $\hat{\mathbf{W}}$, $\hat{\mathbf{B}}$, and $\boldsymbol{\Sigma}_i$, and therefore does not depend on the penalization terms. Instead, it reflects each subject's contribution to the core estimating equations that govern the marginal fit of the model.

To compute the empirical score vector for each subject i , the working loss is defined as:

$$\mathcal{L}_i = (\tilde{z}_i - \hat{\mathbf{B}}^T \hat{\mathbf{W}} \tilde{x}_i)' \boldsymbol{\Sigma}_i^{-1} (\tilde{z}_i - \hat{\mathbf{B}}^T \hat{\mathbf{W}} \tilde{x}_i).$$

Let

$$r_i = \tilde{z}_i - \hat{\mathbf{B}}^T \hat{\mathbf{W}} \tilde{x}_i$$

be the residual vector, and define the estimated component scores

$$f_i = \hat{\mathbf{W}} \tilde{x}_i.$$

The predicted marginal mean is then

$$\hat{\mu}_i = \hat{\mathbf{B}}^T f_i,$$

yielding residual

$$r_i = \tilde{z}_i - \hat{\mu}_i.$$

The empirical score contribution for subject i with respect to $\text{vec}(\mathbf{B})$, the vectorized form of $\mathbf{B}$, is

$$u_i^{(B)} = \frac{\partial \mathcal{L}_i}{\partial \text{vec}(\mathbf{B})} = -2(\hat{\mathbf{f}}_i \otimes \mathbf{I}_Q)' \boldsymbol{\Sigma}_i^{-1} r_i,$$

where $\otimes$ denotes the Kronecker product and $\mathbf{I}_Q$ is the identity matrix of dimension Q . These subject-specific score contributions represent each observation's contribution to the marginal estimating equations. Within the recursive partitioning procedure, systematic variation in these score processes across values of candidate partitioning variables signals instability in the regression parameters.

Although score contributions with respect to $\text{vec}(\mathbf{W})$ can in principle be derived via the chain rule,

$$\frac{\partial \mathcal{L}_i}{\partial \text{vec}(\mathbf{W})} = -2(\tilde{x}_i \otimes \hat{\mathbf{B}})' \boldsymbol{\Sigma}_i^{-1} r_i,$$

both $\mathbf{W}$ and $\mathbf{B}$ are estimated jointly under a single objective function and converge simultaneously within the ERA-GEE framework. Consequently, instability in either set of parameters ultimately reflects instability in the same underlying estimating structure. For computational clarity and to avoid redundant instability testing, recursive partitioning is therefore conducted using score contributions with respect to $\text{vec}(\mathbf{B})$, which sufficiently captures variation in the fitted marginal mean structure across subgroups. In support of transparency and reproducibility, the code used to implement the MOB-ERA-GEE methodology developed in this thesis has been made publicly available on GitHub: <https://github.com/QuantMM/MOB-of-ERA-GEE>.

Simulation Study

Methods

The present simulation study evaluated the performance and practical utility of the MOB-ERA-GEE framework for detecting parameter instability and recovering subgroup structures in multivariate outcome settings. The study had two primary objectives. First, it assessed

methodological performance by examining whether MOB-ERA-GEE appropriately controlled Type I error in the absence of true subgroup differences, demonstrated adequate empirical power when subgroup-specific heterogeneity was present, accurately classified observations into their true subgroups, and correctly recovered the intended subgroup proportions. Second, the study examined the practical utility of MOB-ERA-GEE by identifying the conditions under which the method performed optimally, including the type of partitioning covariate, subgroup balance, magnitude of parameter heterogeneity, and sample size. Together, these objectives provided a comprehensive evaluation of both the inferential accuracy and applied usefulness of MOB-ERA-GEE for researchers analyzing multivariate outcomes in behavioural and health research.

Data Generation and Model Structure

The present simulation builds on the work of Kim and Hwang (2021), who applied MOB within a univariate ERA framework. Unlike conventional regression models, which relates outcomes directly to observed predictors, ERA summarizes predictors into weighted components, resulting in a unique parameter structure that differs from standard regression formulations. As a result, Kim and Hwang's data-generation procedure and overall simulation design provided an important foundation for extending and evaluating MOB to the multivariate ERA-GEE context.

For each simulated dataset, N observations were generated with eight predictor variables, denoted $x_i = (x_{i1}, \dots, x_{i8})'$, where each predictor was independently drawn from a standard normal distribution. Two component scores ($K = 2$) were then constructed as weighted linear combinations of these predictors. The first component was defined using predictors x_{i1} through x_{i4} , and the second using predictors x_{i5} through x_{i8} . The corresponding weight vectors were

fixed at $\boldsymbol{w}_1 = (.7, .6, .5, .4)'$ and $\boldsymbol{w}_2 = (.6, .5, .4, .3)'$, with no correlation specified between the two components, consistent with Kim and Hwang (2021).

Two continuous outcomes, Y_1 and Y_2 , were generated as linear functions of the two component scores. Although the ERA-GEE framework can accommodate non-Gaussian outcomes through appropriate link functions, continuous outcomes were selected to evaluate MOB-ERA-GEE under the most stable and interpretable estimation conditions. This design choice ensured that observed differences in performance could be attributed to the partitioning and estimation procedure itself, rather than to additional complexity introduced by nonlinear link functions or outcome-specific variance structures. The baseline regression coefficient matrix was specified as

$$\beta_0 = \begin{pmatrix} 0.70 & 0.25 \\ 0.30 & 0.20 \end{pmatrix},$$

where rows correspond to components and columns correspond to outcomes. This specification produced a relatively large overall effect for the first outcome and small-to-moderate effect for the second, reflecting the broad range of effect sizes commonly observed in applied behavioural and health research (Ambike et al., 2025; Huang et al., 2024; Lewis et al., 2016; Walsh et al., 2018). A constant residual correlation of 0.3 was assumed between the two outcomes for each subject, introducing a moderate level of association between multivariate responses. For the GEE estimation, an exchangeable working correlation structure was specified, which assumes equal correlation among repeated observations for a given subject. Although GEE yields consistent parameter estimates even when the working correlation structure is misspecified, the exchangeable structure represents a reasonable approximation for models with a small number of multivariate outcomes (Liang & Zeger, 1986; Shults et al., 2009). Figure 1 provides a conceptual diagram of the ERA-GEE model specified in the data-generating procedure. The number of

predictors, components, and outcome variables was held constant across all simulation conditions.

Partitioning Covariates and Subgroup Structure

A primary factor manipulated in the simulation was the type of covariate used to induce subgroup heterogeneity. Both binary and continuous partitioning variables were considered, as they are commonly investigated in prior MOB simulation studies and represent typical covariate types encountered in applied research (Fokkema et al., 2025; Kim & Hwang, 2021; Zeileis et al., 2008). However, when partitioning on a continuous covariate, MOB must perform an exhaustive search over many potential split points to identify the value that maximizes subgroup differences, substantially increasing computational burden (Zeileis et al., 2008). This challenge was particularly amplified in the MOB-ERA-GEE framework due to the additional complexity of multivariate estimation. To mitigate this computational challenge, a binned version of the continuous partitioning covariate was also considered as a more efficient alternative (Vejkanchana & Kuacharoen, 2019; Zakrisson, 2024). In this condition, the original continuous covariate (Z_1) was transformed into a categorical variable by partitioning it into 10 intervals defined by its empirical deciles. Each observation was then assigned to its corresponding decile-based interval, following a similar binning approach to Schlosser et al. (2019). This binning strategy preserves the ordering information contained in the original continuous covariate while substantially reducing the number of candidate split points evaluated by the MOB algorithm. Including this condition allowed the simulation to directly assess the trade-off between computational efficiency and partitioning performance, specifically whether binning could reduce computational burden without substantially compromising split detection or subgroup classification accuracy.

Subgroup membership was determined by a single covariate, Z_I , which varied across simulation conditions. Specifically, Z_I was generated as either a binary variable, $Z_I \sim B(N, p_1)$; a continuous variable, $Z_I \sim U(-1, 1)$; or a binned version of the continuous variable described above. For the continuous and binned continuous conditions, subgroup membership was defined by splitting Z_I at the quantile corresponding to the targeted subgroup proportion p_1 , such that p_1 represented the proportion of observations assigned to the first subgroup. In all conditions, exactly two subgroups were assumed in the data-generating procedure.

To evaluate robustness to irrelevant partitioning variables, three additional noise covariates unrelated to the true subgroup structure were included in all datasets. These consisted of a binary covariate, a continuous uniform covariate, and a normally distributed covariate (Kim & Hwang, 2021; Wickelmaier & Zeileis, 2018; Zeileis et al., 2008). Including these noise variables allowed assessment of whether the MOB algorithm exhibited systemic tendencies to split on covariates with various distributional characteristics. During model estimation, the true partitioning covariate and all noise covariates were jointly considered as candidate partitioning variables.

Subgroup Balance and Parameter Heterogeneity

To examine the impact of subgroup balance on model performance, the proportion of observations assigned to the first subgroup (p_1) was systematically manipulated. Two conditions were considered: a balanced condition with $p_1 = 0.50$, yielding two approximately equal-sized subgroups, and an unbalanced condition with $p_1 = 2/3$, in which the first subgroup was twice the size of the second (i.e., $2/3$ vs. $1/3$; Kim & Hwang, 2021). This manipulation enabled assessment of MOB-ERA-GEE's robustness in detecting parameter instability under unequal subgroup sizes, a scenario frequently encountered in applied research.

Parameter instability was introduced using a deviation parameter (δ), which simultaneously governed subgroup differences in both the ERA component weights and associated regression coefficients. The condition $\delta = 0$ represented the homogeneity case, in which all observations shared identical component structures and regression parameters across subgroups. Values of $\delta > 0$ represented increasing levels of heterogeneity, with larger values producing more pronounced subgroup-specific differences.

Following Kim and Hwang (2021), subgroup differences in the component structure were implemented by systematically shifting the baseline component weight vectors for the two components, w_1 and w_2 , in opposite directions across subgroups. Specifically, the component weight matrices for the two subgroups were defined as

$$W_{Group\ 1} = \begin{bmatrix} w_1 - \delta & 0 \\ 0 & w_2 - \delta \end{bmatrix}, \quad W_{Group\ 2} = \begin{bmatrix} w_1 + \delta & 0 \\ 0 & w_2 + \delta \end{bmatrix}.$$

A parallel adjustment was applied to the regression coefficients linking the component scores to the two outcomes. Let (b_1, b_2) denote the baseline coefficient vectors for the two outcomes.

Subgroup-specific coefficients were then defined as

$$b_{Group\ 1} = (b_1 - \delta, b_2 - \delta), \quad b_{Group\ 2} = (b_1 + \delta, b_2 + \delta).$$

This coordinated manipulation of both component weights and regression coefficients ensured that subgroup differences reflected systematic changes in the underlying measurement and structural components of the ERA model. As a result, increasing values of δ produced progressively stronger violations of parameter homogeneity, allowing for a controlled evaluation of how the magnitude of subgroup heterogeneity affects the performance of MOB-ERA-GEE.

Performance Metrics

To evaluate the ability of MOB-ERA-GEE to detect parameter instability and recover true subgroup structures, four performance metrics were examined: Type I error rate, empirical

power, classification accuracy, and subgroup size recovery. The first three metrics are standard evaluation criteria in MOB and other recursive partitioning simulation studies, ensuring consistency with the existing literature (Kim & Hwang, 2021; Strobl et al., 2015; Wickelmaier & Zeileis, 2018; Zeileis et al., 2008). Subgroup size recovery was included as an additional metric after preliminary analyses indicated that the estimated subgroup proportions did not always align with the data-generating values, particularly under certain design conditions. Because accurate recovery of subgroup proportions was a central feature of the data-generation procedure, an explicit measure of subgroup size recovery was included to explicitly evaluate this aspect of model performance.

Type I error rate was assessed under the homogeneity condition ($\delta = 0$), in which no true subgroup differences were present. It was defined as the proportion of simulated datasets in which MOB-ERA-GEE incorrectly identified at least one partition, thereby reflecting the method's tendency to detect spurious subgroup differences (Kim & Hwang, 2019; Strobl et al., 2015). Empirical power was evaluated under heterogeneity conditions ($\delta > 0$) and was defined as the proportion of datasets in which the method correctly detected and partitioned the data according to the predefined subgroup structure (Kim & Hwang, 2019; Strobl et al., 2015). For both performance metrics, 95% confidence intervals were computed using the binomial distribution. Together, these metrics evaluated the ability of MOB-ERA-GEE to avoid detecting spurious subgroup differences while successfully identifying true underlying heterogeneity.

When MOB-ERA-GEE identified a partition, performance was further evaluated in terms of whether individual observations were assigned to their correct subgroups and whether the estimated subgroup sizes reflected the true generating proportions. Classification accuracy was assessed using Cramér's V , which quantifies agreement between true and estimated subgroup

memberships (Kim & Hwang, 2019; Wickelmaier & Zeileis, 2018). It ranges from 0 to 1, with larger values indicating stronger agreement and greater classification accuracy (Mirkin, 2001). For each simulation condition, the mean and standard deviation of Cramer's V were computed across simulation replications. This measure therefore reflects the extent to which the MOB procedure accurately recovered the underlying subgroup structure beyond merely identifying the presence of a partition.

Subgroup size recovery evaluated the extent to which the fitted partitions reproduced the intended subgroup proportions (50/50 for balanced conditions or 2/3–1/3 for unbalanced conditions). For each simulation condition, the proportion of datasets in which the estimated subgroup sizes matched the true generating proportions was calculated, and 95% confidence intervals were computed using the binomial distribution. Although this exact metric has not been commonly reported in prior MOB simulation studies, similar measures have been used to evaluate correct identification of subgroups in other recursive partitioning methods (e.g., Fokkema et al., 2018; Yang et al., 2020). Taken together, these metrics provided a comprehensive evaluation of MOB-ERA-GEE's ability to detect partitions, correctly classify observations, and recover true subgroup proportions.

Pilot Simulation and Rationale for Final Design

To explore the effects of parameter heterogeneity and sample size on MOB-ERA-GEE, a preliminary pilot simulation was conducted based on ranges utilized in prior recursive partitioning studies (Kim & Hwang, 2021; Schlosser et al., 2019; Wickelmaier & Zeileis, 2018). The full factorial design under consideration included three types of partitioning covariates (binary, continuous, binned continuous), two subgroup balance conditions (balanced, unbalanced), six instability parameters ($\delta = 0, 0.1, 0.2, 0.3, 0.5, 0.8$), and six sample sizes ($N =$

90, 120, 180, 300, 600, 1000). This design resulted in 216 unique conditions, which would have been computationally infeasible for the full simulation.

To address this, the pilot simulation generated ten datasets per condition to provide a preliminary assessment of method behaviour. Across all 216 pilot conditions, the four key performance metrics—Type I error, empirical power, classification accuracy, and subgroup size recovery—were evaluated. Results indicated that multiple combinations of parameter heterogeneity and sample size produced similar performance outcomes. This suggested that including all conditions in the full simulation would provide limited additional insight relative to the substantial computational cost.

Based on these findings, the final simulation design was streamlined to include four instability parameters ($\delta = 0, 0.15, 0.30, 0.60$) and four sample sizes ($N = 90, 180, 300, 600$). This design maintained appropriate coverage across a broad range of conditions while substantially reducing computational burden. For the final simulation, 1,000 datasets were generated for each combination of partitioning covariate type, subgroup balance, parameter heterogeneity, and sample size (Kim & Hwang, 2021; Kok et al., 2021). All simulations for the final design were conducted using the University of Manitoba Grex High Performance Computing System (RRID: SCR_026342), a parallel computing system that enables efficient large-scale computation through task distribution across multiple processors. Once completed, method performance was evaluated using the four key performance metrics described above.

Results

Binary True Covariate Condition

Table 1 presents the Type I error rates, empirical power, classification accuracy, and subgroup size recovery for MOB-ERA-GEE across various sample sizes and levels of

heterogeneity when the true partitioning covariate was binary. Under homogeneity ($\delta = 0$), Type I error rates were consistently below the nominal significance level of 0.05, with the smallest sample size ($N = 90$) exhibiting the greatest conservativeness and larger sample sizes ($N \geq 300$) approaching the nominal level. When heterogeneity was present ($\delta > 0$), empirical power, classification accuracy, and subgroup size recovery improved systematically with both sample size and the magnitude of heterogeneity. Power was particularly sensitive to weak heterogeneity ($\delta = 0.15$), remaining low except in the largest sample ($N = 600$). In contrast, under moderate-to-strong heterogeneity ($\delta = 0.30, 0.60$), power increased sharply across all sample sizes, with larger samples ($N \geq 300$) achieving near-perfect performance. Power was slightly attenuated for unbalanced subgroups in smaller samples and under weaker heterogeneity, though these differences diminished as either increased.

Classification accuracy, quantified by Cramér's V, followed a similar trend to empirical power. In smaller samples ($N \leq 180$) under weak heterogeneity, classification accuracy fell below 0.90, whereas for larger samples ($N \geq 300$), agreement between true and predicted subgroup membership consistently exceeded 0.95 across all levels of heterogeneity. Accuracy was modestly reduced in smaller sample sizes with unbalanced subgroups, but this effect diminished as sample size increased.

Subgroup size recovery was most sensitive to sample size. Recovery rates were below 0.70 for the smallest sample size ($N = 90$), even under strong heterogeneity. They increased sharply at $N = 180$ and reached near-perfect levels for $N \geq 300$ under moderate-to-strong heterogeneity. Unbalanced subgroups demonstrated lower and more variable recovery in smaller samples, with these differences largely resolving at larger samples or stronger levels of heterogeneity.

Computational time was also evaluated across sample size and covariate conditions (Figure 2). Because runtime distributions were moderately skewed in some conditions, particularly for continuous covariates at larger sample sizes, median runtime and the interquartile range (IQR) were used to summarize computational performance. Under the binary partitioning covariate condition, median model fitting times increased gradually with sample size, from 0.12 seconds (IQR = 0.14) at $N = 90$ to 0.39 seconds (IQR = 0.47), 0.75 seconds (IQR = 0.73), and 1.40 seconds (IQR = 0.86) at $N = 180, 300$, and 600 , respectively. Despite this increase, computational demands remained modest, with median runtimes remaining below two seconds across all binary conditions.

Continuous True Covariate Condition

Table 2 summarizes model performance for MOB-ERA-GEE when the true partitioning covariate was continuous. Similar to the binary condition, the Type I error rate under homogeneity ($\delta = 0$) was conservative at the smallest sample size and converged toward the nominal significance level as sample size increased. Under heterogeneous conditions ($\delta > 0$), empirical power displayed trends largely consistent with the binary covariate condition. Power remained low for smaller samples ($N \leq 180$) and weak heterogeneity, but increased substantially as heterogeneity strengthened. Additionally, larger samples ($N \geq 300$) achieved near-perfect power even at moderate levels of heterogeneity. As before, unbalanced subgroups showed slightly reduced power in smaller samples, with these differences diminishing as sample size or heterogeneity increased.

In contrast to the binary condition, classification accuracy differed markedly when the true partitioning covariate was continuous. Cramér's V values were substantially lower overall (ranging from 0.26 to 0.68), indicating difficulty in correctly assigning individual observations to

their true subgroups. Within each sample size, classification accuracy generally improved with increasing heterogeneity. However, at a given level of heterogeneity, larger sample sizes tended to yield lower Cramér's V values, likely because the increased number of candidate split points raises the probability of selecting a threshold that deviates from the true generative value. Additionally, across all conditions, balanced subgroups consistently showed slightly higher classification accuracy compared to unbalanced subgroups.

Subgroup size recovery was also considerably reduced relative to the binary condition, ranging from 0.01 to 0.41. Recovery generally improved with increasing heterogeneity, but the effect of sample size was less consistent. Under weak heterogeneity, recovery showed modest improvement as sample size increased, whereas under moderate-to-strong heterogeneity, larger samples ($N \geq 300$) often exhibited lower recovery, mirroring the trends observed in Cramér's V . Unbalanced subgroups also displayed greater variability and occasionally higher recovery rates than balanced subgroups. Together, these findings suggest that the exhaustive split search procedure required for continuous covariates increases the likelihood of selecting thresholds that deviate from the true generative split-point, reducing subgroup classification and size recovery.

As illustrated in Figure 2, computational efficiency was also examined for the continuous partitioning covariate condition. Relative to the binary condition, median model fitting times increased substantially with sample size, reflecting the additional computational burden associated with evaluating an exhaustive set of potential split points along a continuous scale. Median runtimes were approximately 0.21 seconds (IQR = 0.23) at $N = 90$, 22.23 seconds (IQR = 32.67) at $N = 180$, 76.45 seconds (IQR = 93.04) at $N = 300$, and 314.53 seconds (approximately 5.2 minutes; IQR = 347.41) at $N = 600$. Overall, these results demonstrate that,

unlike the binary condition, computation time increased sharply with sample size under the continuous covariate condition.

Binned Continuous True Covariate Condition

Table 3 presents the performance of MOB-ERA-GEE for the binned continuous condition, in which the true continuous partitioning covariate was discretized into 10 quantiles prior to analysis. Type I error rates again remained below the nominal level under homogeneity, with conservative behaviour in smaller samples and convergence toward the nominal level as sample size increased. Empirical power closely mirrored the patterns observed in both the binary and continuous covariate conditions. Power was lowest when heterogeneity was weak and sample sizes were small, but increased steadily with both larger samples and stronger heterogeneity. For sample sizes of $N \geq 300$, power approached near-perfect levels under moderate-to-strong heterogeneity.

Classification accuracy for the binned continuous covariate was intermediate between the binary and continuous covariate conditions, with Cramér's V ranging from 0.45 to 0.75, and followed patterns similar to those observed in the continuous case. Within each sample size, accuracy generally improved as heterogeneity increased. At moderate-to-strong levels of heterogeneity, smaller samples ($N \leq 180$) consistently exhibited higher classification accuracy than larger samples, reflecting similar challenges associated with selecting an optimal split point as the number of candidate thresholds increases. Across all conditions, balanced subgroups tended to yield slightly higher classification accuracy than unbalanced subgroups.

Subgroup size recovery for the binned continuous condition was also higher than in the continuous condition but remained substantially lower than in the binary condition. Consistent with the continuous covariate results, recovery improved with increasing heterogeneity within

each sample size. When heterogeneity was weak, recovery improved with increasing sample size, particularly at $N = 600$; however, under strong heterogeneity, larger samples often showed lower recovery. Unbalanced subgroups again exhibited greater variability and occasionally produced higher recovery rates than the balanced condition.

Computational time was also evaluated for the binned continuous partitioning covariate condition (Figure 2). Median model fitting times increased modestly with sample size, but remained substantially lower than those observed under the fully continuous covariate condition. Median runtimes were approximately 0.24 seconds (IQR = 0.44) at $N = 90$, 1.69 seconds at $N = 180$ (IQR = 2.40), 3.80 seconds at $N = 300$ (IQR = 4.19), and 5.90 seconds at $N = 600$ (IQR = 4.45). Overall, computation time for the binned continuous covariate condition was substantially lower than for the fully continuous covariate condition, highlighting its practical advantage across all sample sizes.

Empirical Application

To demonstrate the practical utility of MOB-ERA-GEE for identifying heterogeneity in predictor–outcome relationships, the method was applied to baseline data from the CLSA to examine associations between social isolation and both physical and cognitive health. Social isolation is increasingly recognized as a complex, multidimensional construct (Courtin & Knapp, 2017; Henriques et al., 2022), encompassing both structural aspects, such as social network size and interaction frequency, as well as functional aspects, including perceived social support (Stringhini et al., 2012; Valtorta et al., 2016). Social isolation has been linked to adverse health outcomes, including cognitive decline (DiNapoli et al., 2014; Guarnera et al., 2023; Iacono et al., 2024) and reduced grip strength (Yu et al., 2020). Prior research also demonstrates a positive correlation between these outcomes in older adults, with reduced grip strength serving as an

early indicator of cognitive decline (Jeong et al., 2018; McGrath et al., 2019; Tou et al., 2021).

Given this interrelation, modelling cognition and grip strength jointly within a multivariate framework allows for simultaneous estimation while accounting for their dependence, reducing the risk of misleading conclusions from separate analyses.

Although ERA-GEE is well-suited for examining the relationship between social isolation and multiple health outcomes, it assumes uniform effects across the population. In reality, these associations are likely influenced by individual and sociodemographic factors—including age, sex, race and ethnicity, immigrant status, education, and income—which can affect both the prevalence of social isolation and its impact on physical and cognitive health (Beech et al., 2024; Government of Canada, 2024; Islam & Gilmour, 2023; Lu et al., 2023; Yu et al., 2020; Zhang et al., 2025). These covariates may also interact in complex ways, producing distinct subgroups that are difficult to define a priori. Attempting to do so may result in arbitrary classifications that obscure meaningful heterogeneity. Integrating MOB with ERA-GEE thus allows for the empirical identification of such subgroups, providing deeper insights into how social isolation relates to cognitive and physical health across diverse segments of the Canadian population.

Methods

Data and Study Population

The present analysis utilized baseline data from the CLSA Comprehensive cohort, collected between 2011 and 2015 (Raina et al., 2009). The cohort comprises community-dwelling adults aged 45 to 85 years who participated in extensive in-home interviews and standardized physical assessments at one of 11 data collection sites across Canada. These assessments captured detailed information on participants' sociodemographic characteristics,

physical health, cognitive functioning, and psychosocial well-being, with the overarching goal of better understanding the biological, psychological, and social determinants of aging (Raina et al., 2009).

At baseline, the CLSA applied several eligibility exclusions, including individuals who were unable to communicate in English or French, had cognitive impairment of sufficient severity to prevent reliable self-report without assistance, resided in one of Canada's three territories, were full-time members in the Canadian Armed Forces, lived in long-term care institutions, or resided on federal or provincial First Nations reserves or settlements (Raina et al., 2009). These criteria yielded a baseline cohort of $N = 30,097$ participants.

From this baseline cohort, the analytical sample used in the present analysis was derived using complete case analysis. Accordingly, participants with missing data on any of the variables included in the MOB-ERA-GEE model were excluded via listwise deletion, resulting in a final analytical sample of $N = 9,620$; accordingly, 20,477 participants from the baseline cohort were excluded due to incomplete data on one or more analytic variables. Complete case analysis is most defensible when the probability of missingness is unrelated to the variables involved in the analysis, or when any systematic differences can be adequately accounted for by observed information. If missingness is related to participants' sociodemographic characteristics, social isolation indicators, cognitive or physical functioning, or other variables relevant to the analysis, the final analytical sample may differ systematically from the broader eligible CLSA cohort. Consequently, the resulting parameter estimates may be subject to selection bias and should be interpreted with caution, particularly with respect to their generalizability beyond participants with complete data. However, given that the primary objective of the empirical analysis is to demonstrate the applied utility of MOB-ERA-GEE rather than to draw definitive population-

level conclusions about the CLSA cohort, the findings from the present analysis should be interpreted primarily as a methodological application. The CLSA received ethics approval from research ethics boards across participating institutions, and the present analysis was approved by the University of Manitoba Research Ethics Board 1 (protocol #HE2026-0030).

Measures

Social Isolation. Social isolation was operationalized using the 29-item CLSA-SII, developed by Wister et al. (2019). As mentioned previously, the CLSA-SII is a multidimensional measure of social isolation that captures three primary domains: (1) structural/objective, measured by 23 items assessing community and social participation, network size, frequency of contact, living arrangements, and marital status, (2) functional/objective, measured by the four subscales of the Medical Outcomes Study of Social Support Survey, including emotional/informational support, affectionate support, tangible support, and positive social interactions, and (3) functional/subjective dimension, measured by two items reflecting perceived loneliness and desire for greater social engagement. The CLSA-SII has demonstrated good concurrent validity and has been widely used in aging research, supporting its reliability and relevance (Lee et al., 2023; Wister et al., 2019; Wister et al., 2022). A full list of the 29 CLSA-SII items and their assigned domains is provided in Table 4.

Unlike the standard CLSA-SII scoring procedure—which aggregates items within domains and combines the domain scores into a single composite index—the present study retained all 29 items to define the predictor–component structure within the ERA framework. Consistent with the CLSA-SII conceptualization, each item was organized into three lower-dimensional components corresponding to the structural objective, functional objective, and functional subjective domains. This approach preserves the theoretical structure of the original

index while constructing components that empirically estimate each item's contribution to maximizing explained variance in the outcomes, rather than assuming all items contribute equally. All CLSA-SII items were self-reported and may therefore be subject to recall and social desirability bias (Holbrook, 2008; Lee et al., 2023).

Handgrip Strength. Handgrip strength is a widely used marker of overall muscle strength and physical health in older adults, and predicts adverse health outcomes, including disability, cognitive decline, and mortality (McGrath et al., 2019). In the CLSA, handgrip strength (kg) was assessed using a wireless handgrip dynamometer during standardized physical assessments, with participants seated in a straight-backed chair (Canadian Longitudinal Study on Aging [CLSA], 2016). Each participant completed three trials, and the maximum value attained was used in the present analysis.

Executive Functioning. Executive functioning was used as an overall indicator of cognitive health, capturing higher-order cognitive processes such as planning, reasoning, problem solving, and cognitive flexibility. Deficits in executive functioning often appear early in cognitive decline and can predict broader cognitive deterioration independent of memory, making it a reliable marker of overall cognition (Corbo & Casagrande, 2022; Clark et al., 2012). Executive functioning was assessed using a continuous composite score derived from a six-test cognitive battery administered during the CLSA assessment. This battery included the Animal Fluency Test, the Mental Alternation Test, the Stroop Interference Test, and the FAS Phonemic Fluency Task (O'Connell et al., 2022; CLSA, 2022). Test scores were combined and standardized to produce a composite score with a mean of 100, with higher scores indicating better executive functioning and cognitive performance.

Potential Covariates. Six sociodemographic variables were included as candidate partitioning variables in the MOB procedure: age, sex, race and ethnicity, immigrant status, education, and household income. These variables were selected *a priori* based on evidence that both social isolation prevalence and its impact on health vary systematically across demographic and socioeconomic groups (Beech et al., 2024; Government of Canada, 2024; Islam & Gilmour, 2023; Lu et al., 2023; Yu et al., 2020; Zhang et al., 2025).

Age was measured as participants' self-reported chronological age (in years) at baseline. Sex was assessed through a binary response to the question, "Are you male or female?". Race and ethnicity were self-reported using 13 response categories. To ensure adequate subgroup sizes and stable estimation, these categories were collapsed into six broader groups using a strategy similar to Morin et al. (2022): White only, Black only, East Asian only (Chinese, Korean, Japanese), South or Southeast Asian only (including Filipino), other racial or cultural origin only (including West Asian, Arab, and Latin American), and multiple racial or cultural origins.

Immigrant status was coded as a binary variable based on whether participants self-reported being born in Canada or outside Canada. Educational attainment was categorized into four levels based on participants' highest reported credential: (1) no post-secondary degree certificate, or diploma; (2) post-secondary education below the bachelor's level; (3) bachelor's degree; and (4) university degree or certificate above the bachelor's level. Self-reported annual household income before taxes and deductions was grouped into five categories: (1) less than \$20,000; (2) \$20,000 to \$49,999; (3) \$50,000 to \$99,999; (4) \$100,000 to \$149,999; and (5) \$150,000 or more.

Statistical Analysis

Descriptive statistics for continuous variables were reported as means and standard deviations, while categorical variables were reported as percentages. The final ERA-GEE model specified is illustrated in Figure 3. Parameter instability testing and split-point selection were performed using the “mob” function from the “partykit” R package (version 1.2-24; Zeileis & Hothorn, 2022). Recursive partitioning proceeded until no candidate partitioning variables exhibited significant parameter instability at $\alpha = .05$, or until additional splits would have produced nodes with fewer than 200 observations. Following estimation of the final tree, 1,000 bootstrap samples were generated to construct 95% confidence intervals for the subgroup-specific ERA-GEE parameter estimates and evaluate their statistical significance. Ten-fold cross-validation was also applied as a post-pruning diagnostic to assess potential overfitting.

Results

Baseline Characteristics

A summary of the descriptive statistics for the study population is presented in Table 4. Participants were predominantly middle-aged to older adults, with an approximately equal distribution of males and females. Most participants were Canadian-born and identified as White, and the majority were married or living with a partner. Educational attainment was generally high, with most participants having completed some form of post-secondary education. Household income was largely concentrated in the middle to upper strata, with relatively limited representations in the lowest income category.

Parameter Instability and Subgroup Identification

Results from the parameter instability tests are presented in Table 5. At the root node (Node 1), household income demonstrated the strongest evidence of parameter instability,

indicated by the largest Supremum Lagrange Multiplier test statistic and smallest p-value. Consequently, it was selected as the first splitting variable, dividing the sample into two groups: participants with an annual household income below \$20,000 (Node 2) and those with incomes of \$20,000 or higher (Node 3). Further splitting was not pursued in Node 2 due to the minimum node size constraint, and parameter instability tests within Node 3 revealed no additional covariates with significant instability after accounting for income. Thus, the final MOB-ERA-GEE model for this sample consisted of two income-defined subgroups: one comprising participants in the lowest income category and the other including all remaining participants.

Heterogeneity in the Composition of Social Isolation

Subgroup-specific component weight estimates from the final ERA-GEE models are presented in Table 6. In the lower-income subgroup (Node 2), none of the item-level weights defining the social engagement, social support, or subjective well-being components were statistically significant, as all confidence intervals included zero. Because the lower-income subgroup had a small sample size, representing only approximately 2.75% of the analytical sample, these non-significant estimates should be interpreted cautiously as imprecise subgroup-specific estimates rather than as evidence that the social isolation components were not meaningful for lower-income participants. In contrast, many weights in the higher-income subgroup (Node 3) were significant across all three social isolation domains. For social engagement, nine of the 23 predictors contributed significantly to the component, reflecting both the number of social connections—including relatives, household members, and acquaintances from work, school, or other activities—and the frequency of participation in social activities, such as family or friend gatherings, religious events, and neighbour interactions. Some predictors contributed positively to the formation of the component, whereas others contributed negatively.

For social support, all four predictors contributed significantly: positive social interactions and tangible support had positive contributions, whereas affectionate interactions and emotional/informational support contributed negatively. Finally, for subjective well-being, only perceived loneliness was significant, suggesting that this component was largely driven by participants' feelings of social isolation.

Differential Associations Between Social Isolation Components and Health Outcomes

Subgroup-specific regression coefficient estimates are presented in Table 7. In the lower-income subgroup (Node 2), none of the associations between the ERA components and health outcomes reached statistical significance, consistent with the absence of significant component weights. Because the lower-income subgroup represented only approximately 2.75% of the analytical sample, its subgroup-specific estimates are likely to be less precise. The non-significant associations should therefore be interpreted cautiously rather than as definitive evidence that social isolation components were unrelated to executive functioning or handgrip strength. In contrast, several significant associations were observed in the higher-income subgroup (Node 3), with distinct patterns across each component. Specifically, social engagement was positively associated with both health outcomes, indicating that more frequent participation and larger social networks corresponded to higher executive functioning and stronger grip strength. Social support was positively associated with grip strength only, suggesting that practical and positive interactions support physical functioning. Lastly, subjective well-being was positively associated with executive functioning, indicating that a stronger sense of social connectedness corresponded to better cognitive performance. Across both subgroups, executive functioning and handgrip strength showed positive associations based on Spearman's

rank correlations (lower-income: $\rho = .12, p = .048$; higher-income: $\rho = .03, p = .001$), thereby supporting their joint modelling within the ERA-GEE framework.

Model Complexity and Cross-Validation Results

Cross-validation (CV) was conducted to evaluate whether increasing tree complexity improved model fit without introducing overfitting. Because only one meaningful split was identified in the data, two levels of tree complexity were considered: depth 1, representing a single global ERA-GEE model assuming a homogeneous population, and depth 2, which allows the data to split into two heterogeneous subgroups, each with its own set of ERA-GEE parameter estimates. As shown in Figure 4, mean CV loss was very similar between the two tree depths, with depth 1 ($M = 1.8436, SE = 0.0077$) slightly higher than depth 2 ($M = 1.8414, SE = 0.0078$). However, the difference between the depths was minimal (0.0022), and the standard error bars overlapped substantially. These results suggest that the subgroup-specific model did not introduce clear evidence of overfitting, although it also did not provide a meaningful improvement in predictive performance over the simpler tree.

Discussion

The present thesis developed and evaluated the MOB-ERA-GEE framework, a unified methodology that integrates two well-established frameworks, MOB and ERA-GEE. This integration was designed to address three methodological challenges that frequently arise in observational research but are often treated separately: multicollinearity among predictors, correlated multivariate outcomes, and covariate-dependent parameter instability. Simulation results and an empirical application to CLSA data demonstrate that the proposed approach enables supervised dimension reduction and subgroup detection to operate within a single coherent estimating framework. Further, simulations indicate that MOB-ERA-GEE reliably

detects parameter instability and recovers true subgroup structures under favourable conditions. The empirical application further underscores the practical value of this approach, revealing income-based heterogeneity in the association between social isolation and both cognitive and physical health—patterns that would have been obscured under a homogeneous ERA-GEE model. Taken together, these findings establish MOB-ERA-GEE as a flexible and interpretable framework for analyzing complex multivariate observational data in which both dimensionality reduction and subgroup heterogeneity are of interest.

Simulation

The simulation study provides systematic evidence regarding the operating characteristics of MOB-ERA-GEE across varying conditions of sample size, subgroup balance, and effect heterogeneity. Under parameter homogeneity ($\delta = 0$), the method exhibited conservative Type I error rates, particularly at the smallest sample size ($N = 90$), where rejection rates fell well below the nominal significance level of 0.05. This finding is consistent with previous research on Bonferroni-corrected score-based fluctuation tests within the MOB framework (Frick et al., 2014; Kim & Hwang, 2021; Seibold et al., 2019). Because these tests control the family-wise error rate across multiple candidate partitioning covariates at each node, they require stronger evidence before a partition is introduced, which may result in conservative splitting behaviour when sample sizes are small and statistical information is limited. Although this conservatism reduces the likelihood of identifying spurious subgroup partitions, it also reflects a trade-off between Type I error control and sensitivity, as stricter partitioning criteria reduce the power to detect true subgroup differences in smaller samples.

As sample size increased, this conservatism gradually diminished and MOB-ERA-GEE became better calibrated. In particular, Type I error rates moved closer to the nominal

significance level, while statistical power increased steadily across all simulation conditions. This improvement was particularly evident at moderate-to-large sample sizes ($N \geq 300$), where both balanced and unbalanced subgroup structures demonstrated comparable performance and more reliable detection of true heterogeneity. Taken together, these findings suggest that MOB-ERA-GEE adopts a conservative partitioning strategy in smaller samples, reducing the likelihood of spurious partition detection, albeit at the cost of reduced sensitivity. As sample size increases, the procedure becomes better calibrated and more effective at detecting true subgroup heterogeneity, making it particularly well suited for large-scale observational studies, where substantial sample sizes are typically available.

In contrast to the relatively consistent patterns observed for Type I error and power across the three covariate conditions, true partitioning covariate type played an important role in both classification accuracy and subgroup size recovery. When the true partitioning covariate was binary and significant parameter instability was detected, MOB-ERA-GEE was faced with only a single possible split point. Consequently, the algorithm consistently partitioned the sample at the threshold specified in the data-generating procedure, yielding highly stable classification of observations and accurate recovery of subgroup sizes. This pattern closely mirrors findings reported by Kim and Hwang (2021), who observed similarly strong classification accuracy for binary covariates in the univariate MOB-ERA framework, particularly as sample size and effect heterogeneity increased.

In contrast, continuous and binned continuous partitioning variables yielded lower exact subgroup recovery, particularly in smaller samples. This decrease does not imply that MOB-ERA-GEE performs poorly when using continuous or binned continuous partitioning covariates. Rather, it reflects a fundamental property of the MOB framework when applied to covariates

with many candidate split points. Unlike binary covariates, which admit only a single possible partition, continuous and multi-category covariates present a large set of possible thresholds. In these settings, MOB will select the threshold that maximizes evidence of parameter instability in the individual-level score functions, even if that split diverges from the prespecified simulation threshold.

From a practical perspective, this behaviour is not necessarily a limitation. Allowing the algorithm to select among multiple plausible split points enables the discovery of subgroup boundaries that may not be known or anticipated a priori. As a result, reduced recovery of the exact simulated subgroup structure reflects flexibility in threshold selection rather than failure to detect heterogeneity. This finding is consistent with the statistical nature of score-based instability tests, which may yield alternative—but equally plausible—partitions when multiple candidate split points are present (Zeileis et al., 2008). To my knowledge, prior work has not systematically examined how continuous or binned continuous partitioning covariates affect subgroup recovery metrics, making the present study among the first to document and clarify these effects.

Building on these differences in subgroup recovery, computational demand also varied systematically by covariate type. While continuous partitioning covariates allow MOB-ERA-GEE to identify empirically optimal thresholds, they come with substantially increased computational demand, particularly in larger samples. Binned covariates, by contrast, incur only a modest increase in processing time compared to binary covariates, which remain the most computationally efficient. Accordingly, when computational efficiency is a practical consideration, utilizing naturally occurring binary partitioning covariates (e.g., biological sex, left- vs. right-handedness, treatment vs. control) may be computationally preferable when they

are substantively justified. If continuous covariates are of substantive interest and can be grouped into meaningful ordered categories based on prior theory or empirical literature (e.g., age categorized into seven meaningful groups defined by Statistics Canada (2016)), binning may further reduce computational burden. However, for exploratory analyses aimed at discovering novel subgroup boundaries in a data-driven manner, retaining continuous covariates in their original form may be preferable, despite the longer computational time.

Empirical Application

The empirical application of MOB-ERA-GEE to CLSA data highlights the framework's practical utility for detecting heterogeneity in multivariate outcome settings. In the present analysis, only one covariate—household income—introduced meaningful heterogeneity in the social isolation–health relationship, yielding two subgroups with distinct ERA-GEE parameter estimates. This relatively simple tree structure contrasts with many previous MOB applications, which typically yield deeper trees characterized by multiple covariate splits (e.g., see Fokkema et al., 2025; Kim & Hwang, 2021; Tang & Donnell, 2019; Zeileis et al., 2008). At least two methodological considerations may help explain this limited tree growth.

First, many MOB studies embed the algorithm within comparatively simple univariate global models, such as linear or logistic regression, generalized linear mixed models, or univariate ERA (Fokkema et al., 2025; Kim & Hwang, 2021; Zeileis et al., 2008). Relative to these models, ERA-GEE is a comparatively rich base model that captures multivariate predictor–outcome relationships and within-subject correlation. Consequently, ERA-GEE may explain a larger proportion of systematic variability in the data, leaving limited residual heterogeneity available for MOB to detect and partition. In contrast, simpler univariate base models may leave

more unexplained variance, thereby providing greater opportunity for MOB to identify multiple covariate-driven splits.

Second, MOB-ERA-GEE relies on score-based instability tests rather than likelihood-ratio comparisons of subgroup-specific refitted models. Score-based tests fit the global model once and examine whether cumulative contributions to the score function vary systematically when observations are ordered by a candidate partitioning variable (Zeileis et al., 2008). Likelihood-ratio tests, by comparison, evaluate each potential split by fitting separate models for candidate subgroups and directly comparing their likelihoods (Brandmaier et al., 2013; Chow, 1960). Because score-based tests assess instability through deviations from the global model rather than through direct subgroup model refitting, they may be less sensitive when subgroup differences are small or distributed across multiple parameters. As a result, score-based procedures may be less likely to identify additional splits than likelihood-ratio-based approaches.

Evidence for this pattern has been observed in other complex modelling contexts. For example, within the structural equation model (SEM) trees framework, Fokkema and Zeileis (2024) found that likelihood-ratio-based SEM trees tended to yield larger tree structures and implement a greater number of splits than their score-based counterparts. This finding may help explain why applications of SEM trees typically rely on methods that implement likelihood-ratio tests as opposed to score-based methods: refitting subgroup-specific models allows for more direct evaluation of improvements in model fit and greater sensitivity to heterogeneous parameter structures (Brandmaier et al., 2025). Accordingly, these approaches more often identify multiple significant partitioning variables and yield comparatively large tree structures relative to the present analysis (e.g., Alhadabi, 2021; Grassi & Tarantino, 2023). Taken together, the high complexity of ERA-GEE as a base model—coupled with MOB’s reliance on score-

based tests for parameter instability—likely limits the amount of detectable residual heterogeneity and sensitivity of the partitioning procedure. This provides a plausible explanation as to why the present MOB-ERA-GEE empirical analysis identified only a single meaningful partition, despite expectations of more extensive tree growth.

Nonetheless, applying ERA-GEE within each income-defined subgroup enabled a refined decomposition of the CLSA-SII that extends beyond its original index formulation. The standard CLSA-SII aggregates item scores and assigns equal weights to each indicator, without considering how individual items relate to the outcome(s) of interest (Wister et al., 2022). In contrast, the present analysis demonstrated that only a subset of the CLSA-SII items contributed meaningfully to the structural objective, functional objective, and functional subjective domains when these domains were jointly modelled with the health outcomes. This illustrates a key methodological advantage of ERA for multidimensional constructs, as items are empirically weighted according to their contribution to the outcome(s), rather than being assumed to exert equal influence. As a result, the final component scores better reflect the core aspects of social isolation most strongly associated with physical and cognitive health, yielding more informative and outcome-relevant domains.

Subgroup-specific ERA-GEE estimates further revealed distinct patterns in the associations between the three domains of social isolation and the health outcomes. In the lower-income subgroup, which comprised only 2.75% of the sample, none of the social isolation components emerged as statistically significant predictors of either executive functioning or grip strength. However, given the relatively small size of this subgroup, these null findings should be interpreted with caution, as they may reflect limited statistical power and imprecise parameter estimation rather than the absence of meaningful associations.

In contrast, the higher-income subgroup, representing the vast majority of participants, exhibited several positive associations: social engagement was positively associated with both health outcomes, social support was positively associated with grip strength, and subjective well-being was positively associated with executive functioning. Had a homogeneous ERA-GEE model been used, it would have assumed the same social isolation–health associations across all participants, largely reflecting the patterns observed in the majority group. Consequently, the greater uncertainty and lack of statistically significant associations observed in the lower-income subgroup would not have been distinguished from the stronger associations observed in the higher-income subgroup. Instead, these subgroup-specific patterns would have been combined within a single aggregate model, potentially obscuring heterogeneity in both the magnitude and precision of the estimated relationships. Importantly, the comparable CV loss observed between both the global and subgroup-specific models indicates that these differences were captured without substantial overfitting. Together, these findings underscore the value of integrating MOB with ERA-GEE for identifying subtle, subgroup-specific differences in multivariate outcome relationships.

Although these findings demonstrate the value of MOB-ERA-GEE for identifying meaningful subgroup heterogeneity, they also highlight an important methodological consideration regarding the relationship between model complexity, terminal node size, and the stability of subgroup-specific estimation. Within the MOB-ERA-GEE framework, the identification of a statistically meaningful partition does not necessarily imply that each resulting terminal node contains sufficient information to support stable estimation of a complex base model. Rather, the precision and reliability of subgroup-specific parameter estimates depend on

the amount of information available within each node relative to the complexity of the model being estimated.

This issue is clearly demonstrated in the empirical application, where the lower-income subgroup comprised only 2.75% of the sample, yet was required to support estimation of the same ERA-GEE base model as the substantially larger higher-income subgroup. Because this model includes a relatively large number of parameters—including component weights, regression coefficients, and GEE nuisance covariance parameters—smaller terminal nodes may provide less information for stable estimation compared to larger nodes. Consistent with this possibility, the lower-income subgroup exhibited substantially wider confidence intervals for both the component weights and regression coefficients, suggesting greater uncertainty in the subgroup-specific estimates.

Although there are currently no established sample size recommendations for the stable estimation of the ERA-GEE framework, the broader multivariate regression literature has long recognized that the relationship between sample size and the number of estimated parameters can influence the stability and precision of model estimates, particularly when effects are small or predictors are highly correlated (Babiyak, 2004; Green, 1991). While these general principles cannot be directly translated to the ERA setting, they suggest that the balance between available information and model complexity may also be an important consideration for MOB-ERA-GEE.

Accordingly, the present findings highlight a potential trade-off within the MOB-ERA-GEE framework. As recursive partitioning produces progressively smaller terminal nodes, the amount of information available to support estimation of complex subgroup-specific models may decrease, potentially reducing the precision and stability of the resulting parameter estimates. Although the present study was not designed to formally evaluate this relationship, the findings

suggest that the practical utility of MOB-ERA-GEE depends not only on its ability to detect heterogeneity, but also on whether sufficient information is available within each terminal node to support reliable subgroup-specific estimation. Future methodological research examining the interplay among model complexity, terminal node sample size, and estimation stability would therefore help establish practical guidance for the application of MOB-ERA-GEE in complex multivariate settings.

Strengths and Limitations

The present thesis demonstrates several important methodological strengths of the MOB-ERA-GEE framework. First, the approach uses data-driven partitioning to estimate subgroup-specific effects without requiring pre-defined subgroups or arbitrary cut-off points. This allows meaningful patterns of heterogeneity to emerge directly from the data, rather than from researcher-imposed classifications that could obscure important subgroup differences. Second, the method retains the core strengths of ERA-GEE, enabling the analysis of multiple correlated outcomes and high-dimensional, collinear predictors while producing marginally interpretable estimates. As a result, subgroup-specific associations can be detected, interpreted, and compared across subgroups, provided that terminal node sizes are sufficient to support stable estimation. Finally, the resulting tree structure provides a clear and intuitive visualization of how covariates drive subgroup formation. By visually representing the hierarchy of selected partitioning variables, the tree highlights factors that are influential in shaping subgroup-specific predictor–outcome associations, and supports the interpretation and communication of complex patterns.

Despite these strengths, several methodological considerations of the current MOB-ERA-GEE framework warrant discussion. First, the present implementation relies on an alternating least squares (LSE) estimation procedure within the generalized estimating equation (GEE)

framework. In this setting, parameter estimates are obtained by solving estimating equations based on the marginal mean structure, without requiring full specification of the joint outcome distribution (Lee et al., 2019). By contrast, maximum-likelihood estimators (MLE) rely on explicit distributional assumptions and utilize the full likelihood function. When outcomes are continuous and approximately Gaussian, LSE-based GEE estimators and MLE estimators typically yield asymptotically equivalent results (Mair & Wilkie, 2024). However, when outcomes are discrete, skewed, or otherwise non-normal, likelihood-based estimators may achieve greater statistical efficiency under correct distributional specification as they leverage information from the full outcome distribution (Baissa & Rainey, 2020). At the same time, GEE-based LSE offers robustness to distributional misspecification (Kim et al., 2020; Lee et al., 2019). Because MOB relies on score-based instability tests derived from the estimating equations, instability detection in the current framework is driven by the marginal mean structure. As a result, while subgroup detection remains statistically valid, the efficiency of identifying subtle parameter instability may be reduced in certain non-Gaussian contexts.

Second, the current MOB-ERA-GEE framework does not incorporate ridge-type penalty terms in the estimation of component weights or regression coefficients. In the original ERA-GEE formulation, ridge regularization is applied jointly to both component weights and regression coefficients to stabilize supervised dimension reduction and to control variance inflation arising from correlated predictors and correlated components. In such settings, penalization can substantially improve estimation stability and generalization performance. However, in a tree-based framework, introducing joint regularization raises nontrivial questions regarding whether penalty parameters should be held constant across nodes or re-tuned within each subgroup. Because ridge penalties directly modify the estimating equations, node-specific

tuning could confound genuine parameter instability with differences in regularization strength, whereas enforcing a single global penalty may restrict subgroup-specific adaptation. To maintain comparability of model parameters across candidate partitions and ensure that detected splits reflect structural heterogeneity rather than tuning artifacts, the present implementation adopts an unpenalized specification. Although this choice preserves a coherent global model structure within which subgroup differences are evaluated, extensions incorporating carefully controlled regularization remain an important direction for future methodological development.

Third, the exhaustive evaluation of candidate split points for continuous partitioning covariates can be computationally intensive. Although the ERA-GEE model is re-estimated only once per node, evaluating each candidate split in the MOB-ERA-GEE framework requires repeated calculation of the cumulative score-based instability statistic for each potential partition. This repeated evaluation can substantially increase processing time, particularly as sample size, model complexity, or the number of continuous partitioning variables grows. Researchers should therefore consider computational feasibility when applying MOB-ERA-GEE to large-scale datasets or analyses involving numerous continuous covariates. In practice, this burden may be reduced by restricting candidate partitioning variables supplied to the algorithm based on substantive theory or prior empirical evidence, or through preliminary variable-screening procedures before tree construction.

Future Research Directions

The limitations described above suggest several avenues for future research. First, reformulating the ERA-GEE framework within a true maximum-likelihood setting would allow the method to leverage outcome-specific distributional assumptions (e.g., binomial, Poisson, ordinal) and could improve both efficiency and sensitivity to parameter instability, particularly

for non-Gaussian outcomes. Second, incorporating regularization techniques—such as ridge or elastic net penalties—into the ERA-GEE parameter estimation procedure could enhance stability in ultra-high-dimensional settings (Kim et al., 2020; Lee et al., 2019). Third, computational scalability could be improved by implementing parallel evaluation of candidate split points, thereby reducing processing time when analyzing continuous covariates in large or complex datasets.

Conclusion

In conclusion, the present thesis establishes MOB-ERA-GEE as a flexible and interpretable framework for addressing multicollinearity among predictors, modelling multivariate outcomes, and detecting subgroup-specific heterogeneity in a data-driven manner. Simulation studies demonstrated that the method can reliably identify subgroup heterogeneity and highlighted under which conditions optimal performance occurred. An empirical application to CLSA data further illustrated its practical utility, uncovering meaningful subgroup differences in the social isolation–health relationship that may have been obscured under homogeneous models. Together, these findings position MOB-ERA-GEE as a significant methodological advancement for large-scale observational data, offering a principled, data-driven framework for detecting meaningful heterogeneity in complex multivariate relationships.

References

- Alhadabi, A. (2021). Science interest, utility, self-efficacy, identity, and science achievement among high school students: An application of SEM tree. *Frontiers in Psychology, 12*, 634120. <https://doi.org/10.3389/fpsyg.2021.634120>
- Ambike, M., Sawant, S., & Singh, S. (2025). Enhancing behavioral intentions toward body donation through a theory of planned behavior-based, culturally tailored educational intervention: A quasi-experimental study. *BMC Public Health, 25*(1), 4203. <https://doi.org/10.1186/s12889-025-25544-5>
- Baissa, D. K., & Rainey, C. (2020). When BLUE is not best: Non-normal errors and the linear model. *Political Science Research and Methods, 8*(1), 136–148. <https://doi.org/10.1017/psrm.2018.34>
- Beech, B. M., Bruce, M. A., Siddhanta, A., Marshall, G. L., Whitfield, K. E., & Thorpe, R. J. (2024). Racial differences in the association between loneliness and cognitive impairment among older black and white men. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 79*(11), 1. <https://doi.org/10.1093/gerona/glae227>
- Brandmaier, A. M., Prindle, J. J., Arnold, M., & Van Lissa, K. J. (2025). Semtree: Recursive partitioning for structural equation models. R package version 0.9.23. <https://CRAN.R-project.org/package=semtree>
- Brandmaier, A. M., Von Oertzen, T., McArdle, J. J., & Lindenberger, U. (2013). Structural equation model trees. *Psychological Methods, 18*(1), 71–86. <https://doi.org/10.1037/a0030001>

- Braveman, P., & Gottlieb, L. (2014). The social determinants of health: It's time to consider the causes of the causes. *Public Health Reports (1974)*, 129(Suppl 2), 19–31.
<https://doi.org/10.1177/00333549141291s206>
- Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* (1st ed.). Chapman and Hall/CRC. <https://doi-org.uml.idm.oclc.org/10.1201/9781315139470>
- Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, 49(5), 1716–1735. <https://doi.org/10.3758/s13428-016-0814-1>
- Canadian Longitudinal Study on Aging. (2016). *Hand grip strength*. https://www.clsa-elcv.ca/wp-content/uploads/2023/06/sop_dcs_0028_v2.2_2016apr18_baseline_final_watermark.pdf
- Canadian Longitudinal Study on Aging. (2022). *Derived variables – Cognition (COG) normative data (comprehensive assessment)*. https://www.clsa-elcv.ca/wp-content/uploads/2023/06/dv_cog_norms_com_v1.0_2022apr04-1.pdf
- Chen, M., Cao, X., Wang, A., Zhu, Y., Lu, G., Zhang, L., & Shen, L. (2024). A global perspective on risk factors for social isolation in community-dwelling older adults: A systematic review and meta-analysis. *Archives of Gerontology and Geriatrics*, 116, 105211. <https://doi.org/10.1016/j.archger.2023.105211>
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3), 591–605. <https://doi.org/10.2307/1910133>
- Clark, L. R., Schiehser, D. M., Weissberger, G. H., Salmon, D. P., Delis, D. C., & Bondi, M. W. (2012). Specific measures of executive function predict cognitive decline in older

- adults. *Journal of the International Neuropsychological Society*, 18(1), 118–127.
<https://doi.org/10.1017/S1355617711001524>
- Corbo, I., & Casagrande, M. (2022). Higher-level executive functions in healthy elderly and mild cognitive impairment: A systematic review. *Journal of Clinical Medicine*, 11(5), 1204.
<https://doi.org/10.3390/jcm11051204>
- Courtin, E., & Knapp, M. (2017). Social isolation, loneliness and health in old age: A scoping review. *Health & Social Care in the Community*, 25(3), 799–812.
<https://doi.org/10.1111/hsc.12311>
- DiNapoli, E. A., Wu, B., & Scogin, F. (2014). Social isolation and cognitive function in Appalachian older adults. *Research on Aging*, 36(2), 161–179.
<https://doi.org/10.1177/0164027512470704>
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1), 54–75.
<https://doi.org/10.1214/ss/1177013815>
- Fokkema, M., Henninger, M., & Strobl, C. (2025). One model may not fit all: Subgroup detection using model-based recursive partitioning. *Journal of School Psychology*, 109, 101394. <https://doi.org/10.1016/j.jsp.2024.101394>
- Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods*, 50(5), 2016–2034. <https://doi.org/10.3758/s13428-017-0971-x>

- Fokkema, M., & Zeileis, A. (2024). Subgroup detection in linear growth curve models with generalized linear mixed model (GLMM) trees. *Behavior Research Methods*, 56(7), 6759–6780. <https://doi.org/10.3758/s13428-024-02389-1>
- Frick, H., Strobl, C., Zeileis, A., & Gilli, M. (2014). To split or to mix? Tree vs. mixture models for detecting subgroups. In M. Gilli, G. Gonzalez-Rodríguez, & A. Nieto-Reyes (Eds.), *21st international conference on computational statistics* (pp. 379–386). <https://www.zora.uzh.ch/server/api/core/bitstreams/16238ff4-99c0-4be1-aa98-08d79999d4bc/content?trackerId=6df927d77c13c1a5>
- Government of Canada. (2024). *Social isolation of seniors: A focus on new immigrant and refugee seniors in Canada*. <https://www.canada.ca/en/employment-social-development/corporate/seniors-forum-federal-provincial-territorial/social-isolation-immigrant-refugee.html#h2.5>
- Grassi, M., & Tarantino, B. (2023). SEMtree: Tree-based structure learning methods with structural equation models. *Bioinformatics (Oxford, England)*, 39(6). <https://doi.org/10.1093/bioinformatics/btad377>
- Guarnera, J., Yuen, E., & Macpherson, H. (2023). The impact of loneliness and social isolation on cognitive aging: A narrative review. *JAD Reports*, 7(1), 699–714. <https://doi.org/10.3233/ADR-230011>
- Hardin, J. W., & Hilbe, J. M. (2013). *Generalized estimating equations* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b13880>
- Henriques, A., Talih, M., Pastor-Valero, M., Fraga, S., Dias, I., Matijasevich, A., & Barros, H. (2022). A multidimensional perspective of the relation between social isolation and

- depression among Portuguese older adults. *Health & Social Care in the Community*, 30(4), 1412–1421. <https://doi.org/10.1111/hsc.13471>
- Hu, F. B., Goldberg, J., Hedeker, D., Flay, B. R., & Pentz, M. A. (1998). Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology*, 147(7), 694–703. <https://doi.org/10.1093/oxfordjournals.aje.a009511>
- Huang, P.-C., Latner, J. D., Bevan, N., Griffiths, M. D., Chen, J.-S., Huang, C. H., O'Brien, K. S., & Lin, C.-Y. (2024). Internalized weight stigma and psychological distress mediate the association of perceived weight stigma with food addiction among young adults: A cross-sectional study. *Journal of Eating Disorders*, 12(1), Article 150. <https://doi.org/10.1186/s40337-024-01112-x>
- Iacono, A., Oremus, M., Maxwell, C. J., & Tyas, S. L. (2024). Functional social isolation mediates the association between depression and executive function in older women: Findings from the Canadian Longitudinal Study on Aging Comprehensive cohort. *Aging, Neuropsychology, and Cognition*, 31(4), 661–681. <https://doi.org/10.1080/13825585.2023.2226855>
- Islam, M. K., & Gilmour, H. (2023). Immigrant status and loneliness among older Canadians. *Health Reports*, 34(7), 3–18. <https://doi.org/10.25318/82-003-x202300700001-eng>
- Jeong, S., Choi, S., Kim, K., Kim, S. M., Kim, S., & Park, S. M. (2018). Association among handgrip strength, body mass index and decline in cognitive function among the elderly women. *BMC Geriatrics*, 18(1), Article 225. <https://doi.org/10.1186/s12877-018-0918-9>

- Kim, S. (2020). *Interpretable data reduction in prediction modeling: Extended redundancy analysis and its extensions and applications* (Publication No. 28384000) [Doctoral dissertation, McGill University]. ProQuest Dissertations and Theses Global.
- Kim, S., & Hwang, H. (2021). Model-based recursive partitioning of extended redundancy analysis with an application to nicotine dependence among US adults. *British Journal of Mathematical and Statistical Psychology*, 74(3), 567–590.
- Kim, S., Lee, S., Cardwell, R. L., Kim, Y., Park, T., Hwang, H., Kim, J.-S., Molenaar, D., Wiberg, M., González, J., & Böckenholt, U. (2020). An application of regularized extended redundancy analysis via generalized estimating equations to the study of co-occurring substance use among US adults. *Quantitative Psychology*, 322, 365–376. https://doi.org/10.1007/978-3-030-43469-4_27
- Kok, B. C., Choi, J. S., Oh, H., & Choi, J. Y. (2021). Sparse extended redundancy analysis: Variable selection via the exclusive LASSO. *Multivariate Behavioral Research*, 56(3), 426–446. <https://doi.org/10.1080/00273171.2019.1694477>
- Lee, S., Kim, S., Kim, Y., Oh, B., Hwang, H., & Park, T. (2019). Pathway analysis of rare variants for the clustered phenotypes by using hierarchical structured components analysis. *BMC Medical Genomics*, 12, 100. <https://doi.org/10.1186/s12920-019-0517-4>
- Lee, A., McArthur, C., Ioannidis, G., Mayhew, A., Adachi, J. D., Griffith, L. E., Thabane, L., & Papaioannou, A. (2023). Associations between social isolation index and changes in grip strength, gait speed, bone mineral density (BMD), and self-reported incident fractures among older adults: Results from the Canadian Longitudinal Study on Aging (CLSA). *PloS One*, 18(10), e0292788. <https://doi.org/10.1371/journal.pone.0292788>

- Lewis, B. A., Williams, D. M., Frayeh, A., & Marcus, B. H. (2016). Self-efficacy versus perceived enjoyment as predictors of physical activity behaviour. *Psychology & Health, 31*(4), 456–469. <https://doi.org/10.1080/08870446.2015.1111372>
- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13–22. <https://doi.org/10.1093/biomet/73.1.13>
- Lu, M., Bronskill, S. E., Strauss, R., Boblitz, A., Guan, J., Im, J. H. B., Rochon, P. A., Gruneir, A., & Savage, R. D. (2023). Factors associated with loneliness in immigrant and Canadian-born older adults in Ontario, Canada: A population-based study. *BMC Geriatrics, 23*(1), Article 380. <https://doi.org/10.1186/s12877-023-04092-w>
- Mair, C., & Wilkie, C. (2024). *Maximum likelihood or least squares* [Lecture notes]. Bookdown. <https://bookdown.org/colettemair0/bookdown/maximum-likelihood-or-least-squares.html>
- McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, linear, and mixed models* (2nd ed.). John Wiley & Sons.
- McGrath, R., Robinson-Lane, S. G., Cook, S., Clark, B. C., Herrmann, S., O'Connor, M. L., & Hackney, K. J. (2019). Handgrip strength is associated with poorer cognitive functioning in aging Americans. *Journal of Alzheimer's Disease, 70*(4), 1187–1196. <https://doi.org/10.3233/JAD-190042>
- Mirkin, B. (2001). Eleven ways to look at the chi-squared coefficient for contingency tables. *The American Statistician, 55*(2), 111–120. <https://doi.org/10.1198/000313001750358428>
- Morin, S. N., Berger, C., Papaioannou, A., Cheung, A. M., Rahme, E., Leslie, W. D., & Goltzman, D. (2022). Race/ethnic differences in the prevalence of osteoporosis, falls and fractures: A cross-sectional analysis of the Canadian Longitudinal Study on

- Aging. *Osteoporosis International*, 33(12), 2637–2648. <https://doi.org/10.1007/s00198-022-06539-z>
- O’Connell, M. E., Kadlec, H., Griffith, L. E., Maimon, G., Wolfson, C., Taler, V., Simard, M., Tuokko, H., Voll, S., Kirkland, S., & Raina, P. (2022). Methodological considerations when establishing reliable and valid normative data: Canadian Longitudinal Study on Aging (CLSA) neuropsychological battery. *Clinical Neuropsychologist*, 36(8), 2168–2187. <https://doi.org/10.1080/13854046.2021.1954243>
- Pan, W. (2001). Akaike’s information criterion in generalized estimating equations. *Biometrics*, 57(1), 120–125. <https://doi.org/10.1111/j.0006-341X.2001.00120.x>
- Raina, P., Wolfson, C., Kirkland, S., Griffith, L. E., Balion, C., Cossette, B., Dionne, I., Hofer, S., Hogan, D., van den Heuvel, E. R., Liu-Ambrose, T., Menec, V., Mugford, G., Patterson, C., Payette, H., Richards, B., Shannon, H., Sheets, D., Taler, V., . . . Young, L. (2009). The Canadian Longitudinal Study on Aging (CLSA). *Canadian Journal on Aging / La Revue canadienne du vieillissement*, 28(3), 221–229. <https://doi.org/10.1017/S0714980809990055>
- Schlosser, L., Hothorn, T., & Zeileis, A. (2019). *The Power of Unbiased Recursive Partitioning: A Unifying View of CTree, MOB, and GUIDE*. <https://arxiv.org/pdf/1906.10179>
- Seibold, H., Hothorn, T., & Zeileis, A. (2019). Generalised linear model trees with global additive effects. *Advances in Data Analysis and Classification*, 13(3), 703–725. <https://doi.org/10.1007/s11634-018-0342-1>
- Shults, J., Sun, W., Tu, X., Kim, H., Amsterdam, J., Hilbe, J. M., & Ten-Have, T. (2009). A comparison of several approaches for choosing between working correlation structures in

- generalized estimating equation analysis of longitudinal binary data. *Statistics in Medicine*, 28(18), 2338–2355. <https://doi.org/10.1002/sim.3622>
- Statistics Canada. (2016). *Classification of age group*.
<https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=252430>
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, 36(2), 111–147. <https://doi.org/10.1111/j.2517-6161.1974.tb00994.x>
- Stringhini, S., Berkman, L., Dugravot, A., Ferrie, J. E., Marmot, M., Kivimaki, M., & Singh-Manoux, A. (2012). Socioeconomic status, structural and functional, measures of social support, and mortality: The British Whitehall II cohort study, 1985-2009. *American Journal of Epidemiology*, 175(12), 1275–1283. <https://doi.org/10.1093/aje/kwr461>
- Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch trees: A new method for detecting differential item functioning in the Rasch model. *Psychometrika*, 80(2), 289–316.
<https://doi.org/10.1007/S11336-013-9388-3>
- Takane, Y., & Hwang, H. (2005). An extended redundancy analysis and its applications to two practical examples. *Computational Statistics & Data Analysis*, 49(3), 785–808.
<https://doi.org/10.1016/j.csda.2004.06.004>
- Tang, H., & Donnell, E. T. (2019). Application of a model-based recursive partitioning algorithm to predict crash frequency. *Accident Analysis & Prevention*, 132, Article 105274.
<https://doi.org/10.1016/j.aap.2019.105274>
- Tou, N. X., Wee, S.-L., Pang, B. W. J., Lau, L. K., Jabbar, K. A., Seah, W. T., Chen, K. K., & Ng, T. P. (2021). Associations of fat mass and muscle function but not lean mass with

cognitive impairment: The Yishun Study. *PloS One*, 16(8), e0256702.

<https://doi.org/10.1371/journal.pone.0256702>

Valtorta, N. K., Kanaan, M., Gilbody, S., & Hanratty, B. (2016). Loneliness, social isolation and social relationships: What are we measuring? A novel framework for classifying and comparing tools. *BMJ Open*, 6(4), e010799. <https://doi.org/10.1136/bmjopen-2015-010799>

Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.)*, 6(2). <https://doi.org/10.4172/2161-1165.1000227>

Vejkanchana, N., & Kuacharoen, P. (2019). Continuous variable binning algorithm to maximize information value using genetic algorithm. *Applied Informatics* 1051, 158–172. https://doi.org/10.1007/978-3-030-32475-9_12

Walsh, T. P., Gill, T. K., Evans, A. M., Yaxley, A., Chisholm, J. A., Kow, L., Arnold, J. B., & Shanahan, E. M. (2018). Changes in foot pain, structure and function following bariatric surgery. *Journal of Foot and Ankle Research*, 11(1). <https://doi.org/10.1186/s13047-018-0277-y>

Wickelmaier, F., & Zeileis, A. (2018). Using recursive partitioning to account for parameter heterogeneity in multinomial processing tree models. *Behavior Research Methods*, 50(3), 1217–1233. <https://doi.org/10.3758/s13428-017-0937-z>

Wister, A., Cosco, T., Mitchell, B., Menec, V., & Fyffe, I. (2019). Development and concurrent validity of a composite social isolation index for older adults using the CLSA. *Canadian Journal on Aging*, 38(2), 180–192. <https://doi.org/10.1017/S0714980818000612>

- Wister, A. V., Li, L., & Mitchell, B. A. (2022). A study of social isolation, multimorbidity and multiple role demands among middle-age adults based on the Canadian Longitudinal Study on Aging. *International Journal of Aging & Human Development*, 94(3), 312–343. <https://doi.org/10.1177/00914150211040451>
- Yang, J., Dahabreh, I. J., & Steingrimsdottir, J. A. (2020). Causal interaction trees: Tree-based subgroup identification for observational data. *arXiv.Org*. <https://doi.org/10.48550/arxiv.2003.03042>
- Yu, B., Steptoe, A., Niu, K., & Jia, X. (2020). Social isolation and loneliness as risk factors for grip strength decline among older women and men in China. *Journal of the American Medical Directors Association*, 21(12), 1926–1930. <https://doi.org/10.1016/j.jamda.2020.06.029>
- Zakrisson, H. (2024). Tree-based machine learning methods with non-life insurance applications (PhD dissertation, Department of Mathematics, Stockholm University). <https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-226748>
- Zeileis, A., & Hothorn, T. (2022). Parties, models, mobsters: A new implementation of model-based recursive partitioning in R. *Packag. Vignette*. <https://cran.r-project.org/web/packages/partykit/vignettes/mob.pdf>
- Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, 17(2), 492–514. <https://doi.org/10.1198/106186008X319331>
- Zhang, W., Zhang, J., & Gao, N. (2025). Social isolation and cognitive decline in older adults: A longitudinal study across 24 countries. *BMC Geriatrics*, 25(1), 775. <https://doi.org/10.1186/s12877-025-06430-6>

Table 1

Type I Error, Power, Cramér's V, and Subgroup Size Recovery by Sample Size and Parameter Instability (Binary Covariate Condition)

Performance Metric	Balance	δ	Sample Size			
			90	180	300	600
Type I Error	(N/A)	0	.01 [.008, .018]	.03 [.024, .040]	.03 [.021, .035]	.04 [.035, .052]
Power	Balanced	.15	.05 [.036, .062]	.17 [.163, .211]	.58 [.552, .613]	.99 [.979, .993]
		.30	.22 [.191, .242]	.90 [.884, .921]	.99 [.994, 1.0]	1.00 [.996, 1.0]
		.60	.69 [.661, .718]	1.00 [.996, 1.0]	1.00 [.996, 1.0]	1.00 [.996, 1.0]
	Unbalanced	.15	.03 [.021, .043]	.18 [.160, .208]	.50 [.472, .534]	.95 [.932, .960]
		.30	.18 [.156, .204]	.79 [.765, .815]	.99 [.984, .996]	1.00 [.996, 1.0]
		.60	.58 [.548, .609]	1.00 [.998, .998]	1.00 [.996, 1.0]	1.00 [.996, 1.0]
Cramér's V	Balanced	.15	.79 (.370)	.88 (.295)	.97 (.155)	1.00 (.029)
		.30	.94 (.177)	.98 (.096)	.99 (.000)	1.00 (.000)
		.60	.98 (.030)	.99 (.029)	.99 (.000)	1.00 (.000)
	Unbalanced	.15	.83 (.336)	.87 (.312)	.95 (.195)	.99 (.062)
		.30	.93 (.206)	.98 (.107)	.99 (.030)	1.00 (.000)
		.60	.97 (.068)	.99 (.042)	.99 (.000)	1.00 (.000)
Size Recovery	Balanced	.15	.04, [.030, .055]	.17 [.151, .198]	.58 [.544, .605]	.99 [.979, .993]
		.30	.20 [.174, .224]	.90 [.874, .913]	1.00 [.993, 1.0]	1.00 [.996, 1.0]
		.60	.66 [.627, .686]	1.00 [.994, 1.0]	1.00 [.994, 1.0]	1.00 [.996, 1.0]
	Unbalanced	.15	.03 [.019, .040]	.17 [.143, .189]	.49 [.453, .515]	.94 [.924, .956]
		.30	.17 [.150, .197]	.78 [.756, .807]	.99 [.983, .995]	1.00 [.996, 1.0]
		.60	.56 [.530, .591]	.99 [.984, .996]	1.00 [.996, 1.0]	1.00 [.996, 1.0]

Note. δ = degree of parameter instability. Results are based on 1,000 simulation replications per condition. Type I error, power, and subgroup size recovery are reported as proportions. Values in square brackets represent 95% confidence intervals across simulation replications. Cramér's V is reported as an average classification agreement measure across replications, with values in parentheses

indicating standard deviations. Balanced and unbalanced conditions correspond to equal (50/50) and unequal (2/3–1/3) true subgroup proportions, respectively.

Table 2

Type I Error, Power, Cramér's V, and Subgroup Size Recovery by Sample Size and Parameter Instability (Continuous Covariate Condition)

Performance Metric	Balance	δ	Sample Size			
			90	180	300	600
Type I Error	(N/A)	0	.01 [.008, .017]	.03 [.021, .035]	.04 [.032, .050]	.04 [.034, .052]
Power	Balanced	.15	.04 [.030, .055]	.18 [.160, .208]	.56 [.529, .590]	.99 [.980, .994]
		.30	.20 [.180, .230]	.92 [.899, .933]	1.00 [.996, 1.0]	1.00 [.996, 1.0]
		.60	.67 [.640, .689]	1.00 [.996, 1.0]	1.00 [.996, 1.0]	1.00 [.996, 1.0]
	Unbalanced	.15	.03 [.023, .045]	.17 [.148, .195]	.49 [.459, .521]	.96 [.946, .970]
		.30	.17 [.146, .192]	.80 [.778, .827]	.99 [.975, .991]	1.00 [.996, 1.0]
		.60	.58 [.545, .606]	.99 [.986, .997]	1.00 [.996, 1.0]	1.00 [.996, 1.0]
Cramér's V	Balanced	.15	.49 (.295)	.46 (.265)	.40 (.215)	.30 (.192)
		.30	.67 (.182)	.50 (.192)	.40 (.201)	.29 (.294)
		.60	.68 (.681)	.53 (.202)	.44 (.219)	.32 (.211)
	Unbalanced	.15	.56 (.328)	.41 (.242)	.37 (.187)	.29 (.176)
		.30	.64 (.211)	.47 (.202)	.38 (.180)	.26 (.153)
		.60	.63 (.200)	.50 (.204)	.41 (.211)	.31 (.198)
Size Recovery	Balanced	.15	.01 [.007, .021]	.05 [.034, .060]	.06 [.045, .074]	.05 [.041, .069]
		.30	.06 [.046, .075]	.12 [.098, .137]	.09 [.076, .112]	.04 [.031, .056]
		.60	.20 [.175, .225]	.16 [.139, .184]	.12 [.099, .139]	.07 [.053, .084]
	Unbalanced	.15	.02 [.013, .031]	.03 [.019, .039]	.05 [.037, .064]	.05 [.036, .063]
		.30	.11 [.093, .132]	.14 [.122, .165]	.08 [.065, .098]	.03 [.023, .045]
		.60	.41 [.376, .437]	.22 [.192, .243]	.14 [.117, .160]	.09 [.075, .110]

Note. δ = degree of parameter instability. Results are based on 1,000 simulation replications per condition. Type I error, power, and subgroup size recovery are reported as proportions. Values in square brackets represent 95% confidence intervals across simulation replications. Cramér's V is reported as an average classification agreement measure across replications, with values in parentheses

indicating standard deviations. Balanced and unbalanced conditions correspond to equal (50/50) and unequal (2/3–1/3) true subgroup proportions, respectively.

Table 3

Type I Error, Power, Cramér's V, and Subgroup Size Recovery by Sample Size and Parameter Instability (Binned Continuous Covariate Condition)

Performance Metric	Balance	δ	Sample Size			
			90	180	300	600
Type I Error	(N/A)	0	.02 [.012, .024]	.02 [.018, .032]	.04 [.033, .051]	.04 [.030, .047]
Power	Balanced	.15	.04 [.026, .049]	.21 [.186, .236]	.59 [.554, .636]	.98 [.965, .984]
		.30	.21 [.188, .238]	.91 [.895, .930]	1.00 [.994, 1.0]	1.00 [.996, 1.0]
		.60	.69 [.662, .719]	1.00 [.996, 1.0]	1.00 [.996, 1.0]	1.00 [.996, 1.0]
	Unbalanced	.15	.03 [.022, .044]	.18 [.154, .201]	.42 [.390, .451]	.93 [.917, .948]
		.30	.15 [.126, .170]	.75 [.717, .771]	.98 [.972, .989]	1.00 [.996, 1.0]
		.60	.51 [.477, .539]	.97 [.962, .982]	1.00 [.996, 1.0]	1.00 [.996, 1.0]
Cramér's V	Balanced	.15	.51 (.344)	.63 (.254)	.50 (.219)	.59 (.228)
		.30	.75 (.182)	.66 (.186)	.50 (.216)	.56 (.229)
		.60	.75 (.141)	.67 (.177)	.53 (.227)	.56 (.226)
	Unbalanced	.15	.45 (.340)	.59 (.259)	.48 (.236)	.53 (.226)
		.30	.65 (.227)	.61 (.206)	.49 (.221)	.52 (.221)
		.60	.67 (.190)	.62 (.199)	.51 (.218)	.53 (.216)
Size Recovery	Balanced	.15	.02 [.011, .027]	.09 [.077, .113]	.10 [.085, .122]	.25 [.224, .278]
		.30	.12 [.04, .145]	.33 [.299, .357]	.19 [.167, .215]	.25 [.228, .282]
		.60	.36 [.329, .388]	.37 [.337, .396]	.24 [.209, .261]	.25 [.226, .280]
	Unbalanced	.15	.02 [.015, .033]	.09 [.072, .107]	.12 [.099, .138]	.28 [.251, .307]
		.30	.12 [.099, .138]	.33 [.302, .360]	.24 [.216, .270]	.29 [.262, .318]
		.60	.39 [.357, .418]	.46 [.425, .487]	.28 [.251, .307]	.31 [.278, .335]

Note. δ = degree of parameter instability. Results are based on 1,000 simulation replications per condition. Type I error, power, and subgroup size recovery are reported as proportions. Values in square brackets represent 95% confidence intervals across simulation replications. Cramér's V is reported as an average classification agreement measure across replications, with values in parentheses

indicating standard deviations. Balanced and unbalanced conditions correspond to equal (50/50) and unequal (2/3–1/3) true subgroup proportions, respectively.

Table 4*Variables and Summary Statistics for the CLSA Comprehensive Cohort (Baseline)*

Variable name	Variable description	<i>M</i>	<i>SD</i>
Outcomes			
COG_CONSTR_EF4	Executive Functioning Latent Construct for 6-COG-test battery	99.94	15.02
GS_EXAM_MAX	Max grip strength (kg) attained from all three trials	35.65	11.68
Predictors			
F1: Structural Objective SPA_OUTS	Frequency of Participation in Family / Friends Activities out of the Household (Past 12 Months; 1 = Never 2 = At least once a year 3 = At least once a month 4 = At least once a week 5 = At least once a day)	3.59	0.66
SPA_CHRCH	Frequency of Participation in Religious Activities (Past 12 Months; 1 = Never 2 = At least once a year 3 = At least once a month 4 = At least once a week 5 = At least once a day)	2.19	1.23
SPA_CLUB	Frequency of Participation in Clubs or Fraternal Organization Activities (Past 12 Months; 1 = Never 2 = At least once a year 3 = At least once a month 4 = At least once a week 5 = At least once a day)	1.54	0.94
SPA_EDUC	Frequency of Participation in Educational or Cultural Activities (Past 12 Months; 1 = Never 2 = At least once a year 3 = At least once a month 4 = At least once a week	2.71	0.80

Variable name	Variable description	<i>M</i>	<i>SD</i>
	5 = At least once a day)		
SPA_NEIBR	Frequency of Participation in Association Activities (Past 12 Months; 1 = Never 2 = At least once a year 3 = At least once a month 4 = At least once a week 5 = At least once a day)	2.10	0.99
SPA_OTACT	Frequency of Participation in Other Recreational Activities (Past 12 Months; 1 = Never 2 = At least once a year 3 = At least once a month 4 = At least once a week 5 = At least once a day)	2.99	1.12
SPA_SPORT	Frequency of Participation in Sports or Physical Activities with Others (Past 12 Months; 1 = Never 2 = At least once a year 3 = At least once a month 4 = At least once a week 5 = At least once a day)	3.37	1.17
SPA_VOLUN	Frequency of Participation in Volunteer or Charity Work (Past 12 Months; 1 = Never 2 = At least once a year 3 = At least once a month 4 = At least once a week 5 = At least once a day)	2.43	1.15
SN_CHILD_NB	Number of living children	2.61	1.21
SN_FRND_NB	Number of close friends	5.98	6.14
SN_NEIBR_NB	Number of neighbours	11.70	11.84
SN_SIBLIV_NB	Number of living siblings	3.04	2.10

Variable name	Variable description	<i>M</i>	<i>SD</i>
SN_RELLIV_NB	Number of living relatives	34.35	26.63
SN_PERWSCH_NB	Number of people known through work or school	52.92	40.37
SN_PERCOM_NB	Number of people known through community involvement	38.29	37.40
SN_PERACT_NB	Number of people known through other activities	26.78	31.96
SN_SEECHILD	Last get together with children outside of household (1 = More than 1 year ago 2 = Within the past year 3 = Within the past 6 months 4 = Within the past month 5 = Within the last week or two 6 = Within the last day or two)	4.91	1.19
SN_SEESIB	Last get together with siblings outside of household (1 = More than 1 year ago 2 = Within the past year 3 = Within the past 6 months 4 = Within the past month 5 = Within the last week or two 6 = Within the last day or two)	3.62	1.55
SN_SEEREL	Last get together with relatives outside of household (1 = More than 1 year ago 2 = Within the past year 3 = Within the past 6 months 4 = Within the past month 5 = Within the last week or two 6 = Within the last day or two)	4.07	1.54
SN_SEEFRND	Last get together with close friends outside of household (1 = More than 1 year ago 2 = Within the past year 3 = Within the past 6 months 4 = Within the past month)	5.02	1.03

Variable name	Variable description	<i>M</i>	<i>SD</i>
	5 = Within the last week or two 6 = Within the last day or two)		
SN_SEENEIBR	Last get together with neighbours (1 = More than 1 year ago 2 = Within the past year 3 = Within the past 6 months 4 = Within the past month 5 = Within the last week or two 6 = Within the last day or two)	4.29	1.71
SN_LIVH_NB	Number of people living in household (excluding the participant)	1.20	0.93
SDC_MRTL	Marital/partner status (0 = Single, never married / never lived with partner, widowed, divorced, or separated 1 = Married / living with a partner in a common-law relationship)	% Married: 76.51	
F2: Functional Objective SSA_DPAFF	Affection Subscale from the MOS Social Support Survey	89.24	16.32
SSA_DPEMO	Emotional and Informational Support Subscale from the MOS Social Support Survey	82.68	16.76
SSA_DPSOC	Positive Social Interaction Subscale from the MOS Social Support Survey	84.82	16.73
SSA_DPTNG	Tangible Social Support Subscale from the MOS Social Support Survey	83.56	18.65
F3: Functional Subjective DEP_LONLY	Within the past week, how often did the participant feel lonely? (1 = All of the time (5-7 days) 2 = Occasionally (3-4 days) 3 = Some of the time (1-2 days) 4 = Rarely or never (less than 1 day))	3.68	0.67
SPA_MORAC	Desire to participate in more activities (past 12 months;	% Yes: 45.34	

Variable name	Variable description	<i>M</i>	<i>SD</i>
	1 = “Yes” 2 = “No”)		
Partitioning Covariates			
AGE_NMBR	Age (Years)	63.56	9.10
SEX_ASK	Sex (Male / Female)	% Male: 50.54	
SDC_DCGT	Cultural / Racial Background (White only, Black only, East Asian only, South/Southeast Asian only, Other (Only), Multiple)	%: 95.12/0.60/0. 62/1.23/1.03/ 1.39	
SDC_COB	Country of birth (Canadian-Born / Foreign-Born)	% Canadian- Born: 82.49	
ED_HIGH	Education highest degree (1 = No post-secondary education 2 = Post-secondary education below bachelor’s level 3 = Bachelor's degree 4 = University degree or certificate above bachelor’s degree)	%: 8.89/38.85/26 .38/25.88	
INC_TOT	Total household income (1 = Less than \$20,000 2 = \$20,000 or more, but less than \$50,000 3 = \$50,000 or more, but less than \$100,000 4 = \$100,000 or more, but less than \$150,000 5 = \$150,000 or more)	%: 2.75/18.73/37 .08/22.33/19. 11	

Note. $N = 9,620$. *CLSA* = Canadian Longitudinal Study on Aging; *MOS* = Medical Outcomes

Study. All variable names originally ended in the suffix “_COM” (Comprehensive Cohort). This suffix has been removed for readability.

Table 5

Parameter Instability Test Results from MOB-ERA-GEE for the CLSA Comprehensive Cohort (Baseline)

Node	Age		Sex		Cultural / Racial Background		Country of Birth		Education		Income	
	supLM statistic	<i>p</i> -value	supLM statistic	<i>p</i> -value	supLM statistic	<i>p</i> -value	supLM statistic	<i>p</i> -value	supLM statistic	<i>p</i> -value	supLM statistic	<i>p</i> -value
1	49.37	< .001	45.77	< .001	72.10	< .001	12.85	.24	34.17	.07	111.79	< .001
3	13.18	.97	3.04	1.00	27.92	.99	5.74	.97	20.45	.89	17.52	.98

Note. supLM = Supremum Lagrange Multiplier test statistic. Node 2 did not attempt a split as the number of observations was below the pre-specified minimum of $N = 200$, as determined by the model's pre-pruning criterion.

Table 6*Subgroup-Specific Component Weight Estimates from MOB-ERA-GEE for the CLSA**Comprehensive Cohort (Baseline)*

Subgroup	Component	Predictors	Estimate	95% CI	
				LL	UL
Node 2	F1: Social Engagement	SPA_OUTS_COM	-.41	-.57	.07
		SPA_CHRCH_COM	.32	-.10	.56
		SPA_CLUB_COM	-.03	-.36	.34
		SPA_EDUC_COM	-.11	-.43	.29
		SPA_NEIBR_COM	.20	-.15	.44
		SPA_OTACT_COM	.02	-.32	.32
		SPA_SPORT_COM	-.05	-.34	.27
		SPA_VOLUN_COM	-.01	-.48	.44
		SN_CHILD_NB_COM	.27	-.08	.47
		SN_FRND_NB_COM	.11	-.19	.28
		SN_NEIBR_NB_COM	.20	-.11	.43
		SN_SIBLIV_NB_COM	-.26	-.49	.08
		SN_RELLIV_NB_COM	-.02	-.30	.30
		SN_PERWSCH_NB_COM	-.38	-.57	.12
		SN_PERCOM_NB_COM	-.33	-.57	.12
		SN_PERACT_NB_COM	.12	-.24	.45
		SN_SEECHILD_COM	.13	-.23	.39
		SN_SEESIB_COM	-.04	-.36	.30
		SN_SEEREL_COM	.24	-.18	.50
		SN_SEEFRND_COM	.13	-.22	.41
		SN_SEENEIBR_COM	.09	-.22	.32
		SN_LIVH_NB_COM	.08	-.32	.33
		SDC_MRTL_COM	-.52	-.74	.01
	F2: Social Support	SSA_DPAFF_COM	1.07	-.25	1.33
		SSA_DPEMO_COM	.63	-.50	1.12
		SSA_DPSOC_COM	-.71	-1.46	.94
		SSA_DPTNG_COM	-.33	-1.23	1.12
	F3: Subjective Well-Being	DEP_LONLY_COM	.05	-.98	1.00
		SPA_MORAC_COM	.99	-.10	1.03
Node 3	F1: Social Engagement	SPA_OUTS_COM	-.12	-.18	-.05
		SPA_CHRCH_COM	-.37	-.43	-.29
		SPA_CLUB_COM	.03	-.05	.11
		SPA_EDUC_COM	-.01	-.09	.09
		SPA_NEIBR_COM	.01	-.06	.08
		SPA_OTACT_COM	-.02	-.09	.05
		SPA_SPORT_COM	.06	-.01	.13
		SPA_VOLUN_COM	.01	-.08	.09
		SN_CHILD_NB_COM	-.05	-.12	.02

	SN_FRND_NB_COM	-.08	-.15	-.02
	SN_NEIBR_NB_COM	-.04	-.11	.02
	SN_SIBLIV_NB_COM	.01	-.08	.08
	SN_RELLIV_NB_COM	-.14	-.20	-.07
	SN_PERWSCH_NB_COM	.59	.53	.64
	SN_PERCOM_NB_COM	-.04	-.12	.05
	SN_PERACT_NB_COM	.15	.07	.21
	SN_SEECHILD_COM	-.06	-.12	.01
	SN_SEESIB_COM	-.02	-.10	.04
	SN_SEEREL_COM	-.06	-.13	.01
	SN_SEEFRND_COM	-.05	-.12	.03
	SN_SEENEIBR_COM	.11	.04	.17
	SN_LIVH_NB_COM	.30	.22	.38
	SDC_MRTL_COM	.38	.30	.45
F₂: Social Support	SSA_DPAFF_COM	-.49	-.78	-.18
	SSA_DPEMO_COM	-1.12	-1.36	-.71
	SSA_DPSOC_COM	.62	.29	.88
	SSA_DPTNG_COM	1.08	.78	1.23
F₃: Subjective Well-Being	DEP_LONLY_COM	1.01	0.66	1.02
	SPA_MORAC_COM	-.20	-.81	.42

Note. *CI* = confidence interval; *LL* = lower limit; *UL* = upper limit. Total *N* = 9,620. The

reported 95% CIs were derived from 1,000 bootstrap samples. Estimates with 95% CIs excluding zero are shown in boldface font.

Table 7*Subgroup-Specific Regression Coefficient Estimates from MOB-ERA-GEE for the CLSA**Comprehensive Cohort (Baseline)*

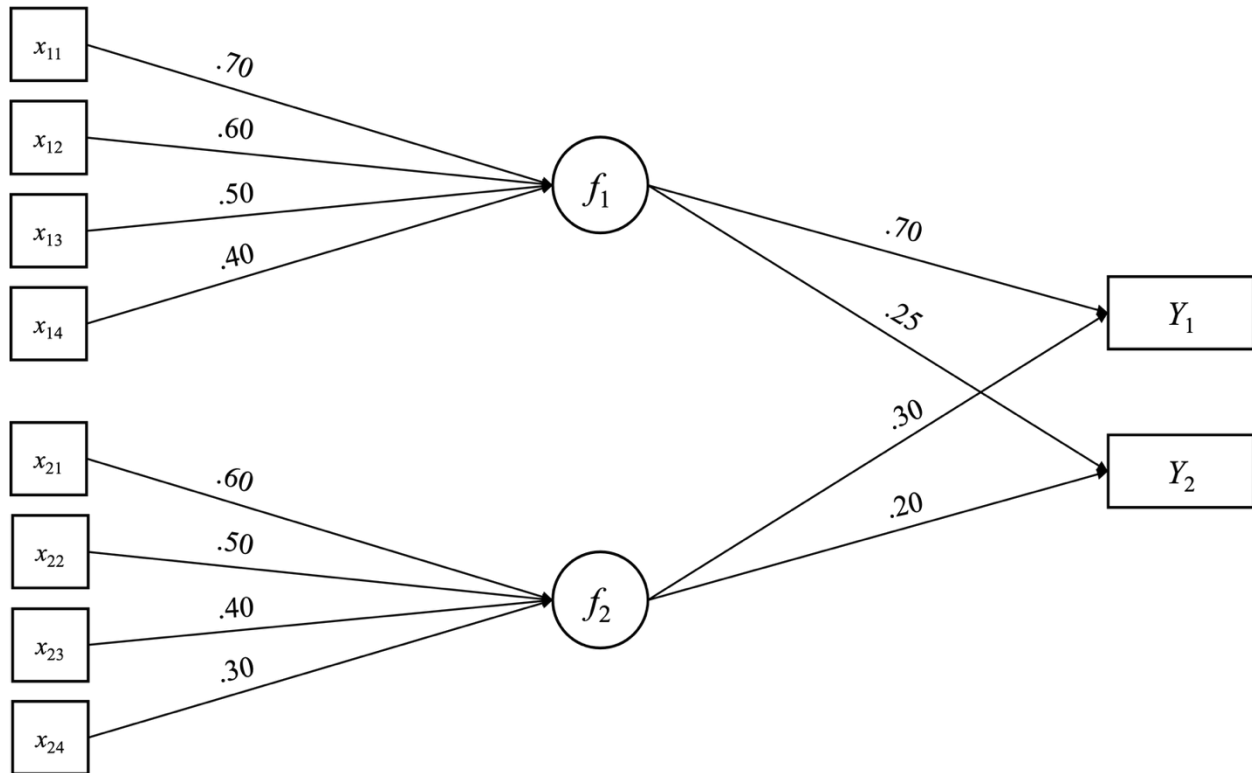
Subgroup	Components	Outcomes	Estimate	95% CI	
				<i>LL</i>	<i>UL</i>
Income < \$20,000 (<i>N</i> = 265)	F ₁ : Social Engagement	Y ₁ : Executive Functioning	-.21	-.53	.19
		Y ₂ : Grip Strength	-.35	-.57	0.00
		Y ₂ : Grip Strength	-.18	-.32	.05
	F ₂ : Social Support	Y ₁ : Executive Functioning	.02	-.27	.28
		Y ₂ : Grip Strength	-.18	-.32	.05
		Y ₂ : Grip Strength	-.11	-.25	.09
Income ≥ \$20,000 (<i>N</i> = 9,355)	F ₁ : Social Engagement	Y ₁ : Executive Functioning	.10	.05	.14
		Y ₂ : Grip Strength	.29	.27	.32
		Y ₂ : Grip Strength	.10	.08	.12
	F ₂ : Social Support	Y ₁ : Executive Functioning	-.02	-.05	.02
		Y ₂ : Grip Strength	.10	.08	.12
		Y ₂ : Grip Strength	.03	.01	.07
	F ₃ : Subjective Well-Being	Y ₁ : Executive Functioning	.03	.01	.07
		Y ₂ : Grip Strength	.03	.01	.07
		Y ₂ : Grip Strength	.02	-.01	.04

Note. *CI* = confidence interval; *LL* = lower limit; *UL* = upper limit. Total *N* = 9,620. The

reported 95% CIs were derived from 1,000 bootstrap samples. Estimates with 95% CIs excluding zero are shown in boldface font.

Figure 1

Conceptual Diagram of the ERA-GEE Model Specified in the Simulation Study

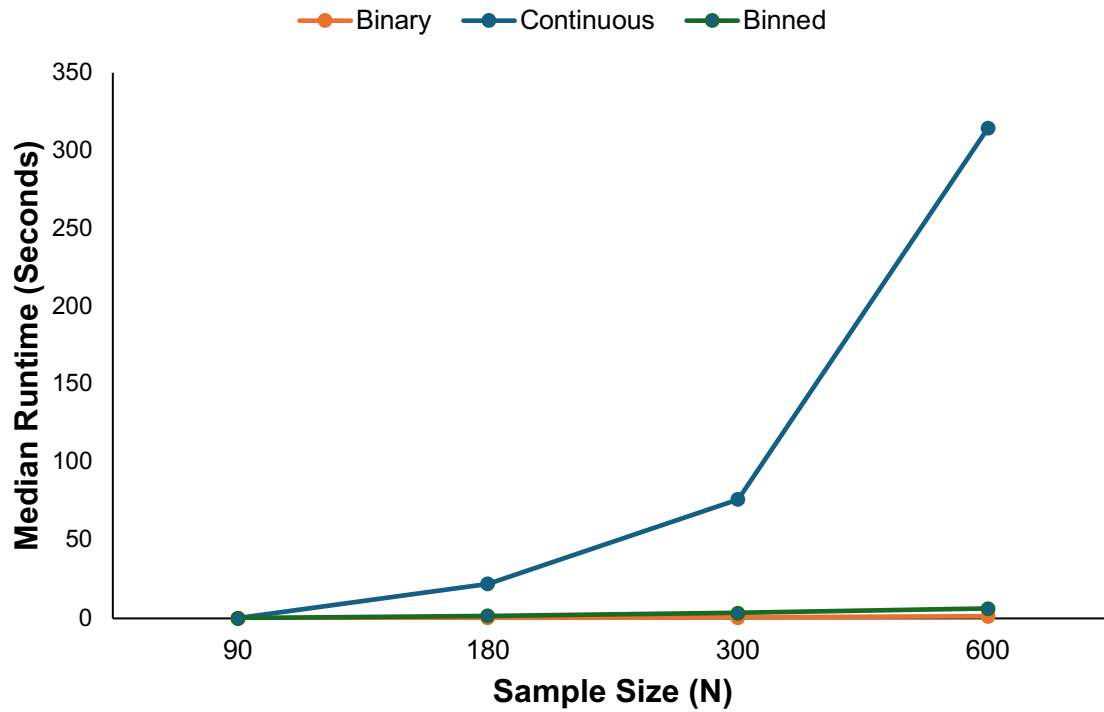


Note. Square boxes represent the eight standardized predictor variables, while circles denote the two component scores constructed as weighted linear combinations of these predictors.

Rectangles represent continuous outcome variables (Y_1 and Y_2). Arrows from predictors to components indicate fixed component weights, while arrows from components to outcomes indicate fixed regression coefficient estimates. A residual correlation of 0.30 is specified between Y_1 and Y_2 .

Figure 2

Computational Performance of MOB-ERA-GEE Across Sample Size and Covariate Conditions



Note. Median run times were calculated across 1,000 Monte Carlo replications. The binned covariate condition was created by partitioning the original continuous covariate into 10 decile-based intervals, thereby reducing the number of candidate split points evaluated by the MOB algorithm.

Figure 3

Final MOB-ERA-GEE Model Specified for the CLSA Comprehensive Cohort (Baseline) Data

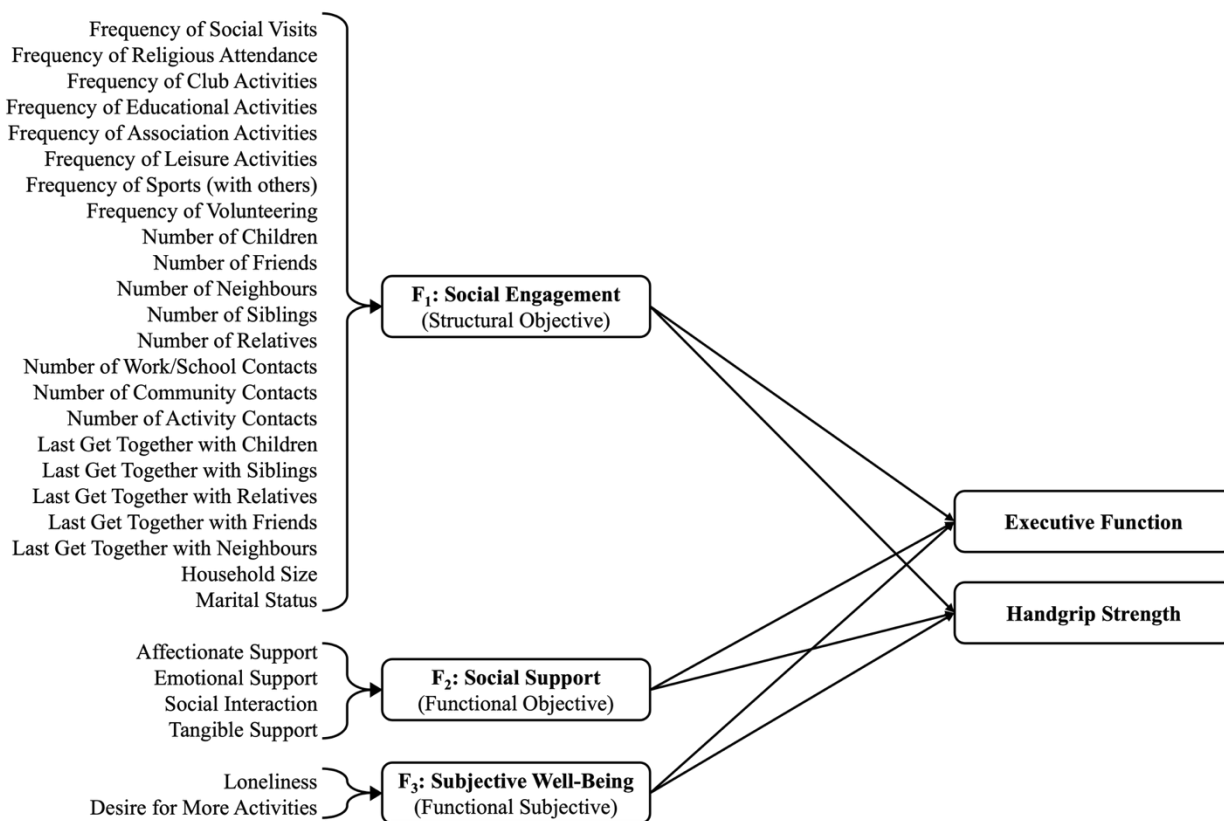
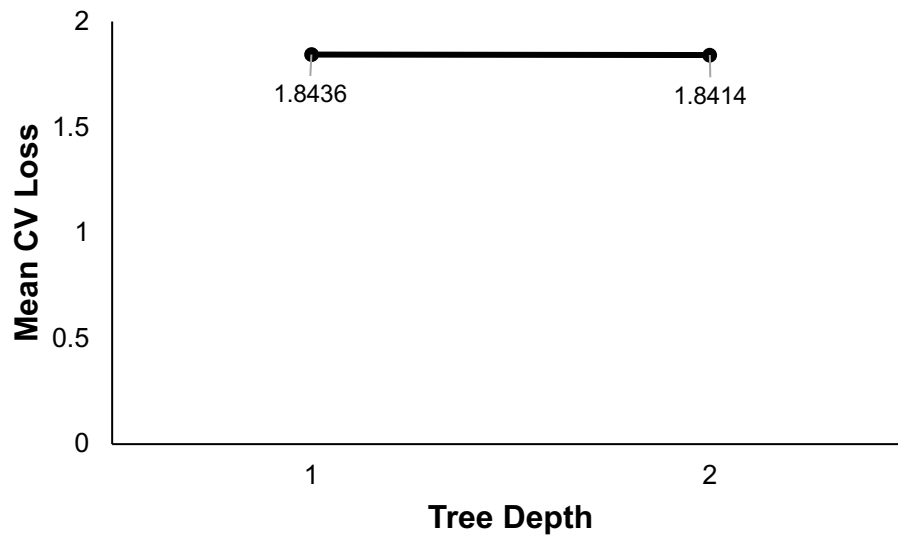


Figure 4

Comparison of 10-Fold Cross-Validation Loss Across MOB-ERA-GEE Tree Depths



Note. CV = cross-validation. Tree depth 1 represents a single homogeneous group, while depth 2 allows one additional split, creating two subgroups based on detected parameter instability.