

Clustering of health behaviours in Canadians: A multiple behaviour analysis of data from the Canadian Longitudinal Study on Aging

Abstract

Background: Health behaviours such as physical inactivity, unhealthy eating, smoking tobacco, and alcohol use are each leading risk factors for non-communicable chronic disease. Better understanding which behaviours tend to co-occur (i.e., cluster together) and co-vary (i.e., are correlated) may provide novel opportunities to develop more comprehensive interventions to promote multiple health behaviour change. However, whether co-occurrence or co-variation based approaches are better suited for this task remains relatively unknown.

Purpose: To compare the utility of co-occurrence vs co-variation based approaches for understanding the interconnectedness between multiple health impacting behaviours.

Methods: Using baseline and follow-up data (N=40,268) from the Canadian Longitudinal Study of Aging, we examined the co-occurrence and co-variation of health behaviours. We used cluster analysis to group individuals based on their behavioural tendencies across multiple behaviours and to examine how these clusters are associated with demographic characteristics and health indicators. We compared outputs from cluster analysis to behavioural correlations and compared regression analyses of clusters and individual behaviours predicting future health outcomes.

Results: Seven clusters were identified, with clusters differentiated by six of the seven health behaviours included in the analysis. Sociodemographic characteristics varied across several clusters. Correlations between behaviours were generally small. In regression analyses individual behaviours accounted for more variance in health outcomes than clusters.

Conclusions: Co-occurrence based approaches may be more suitable for identifying sub-groups for intervention targeting while co-variation approaches are more suitable for building an understanding of the relationships between health behaviours.

Keywords: health behaviours, multiple behaviours, cluster analysis, CLSA

Non-communicable chronic diseases such as chronic respiratory disease, diabetes, cardiovascular disease, and cancer cause two thirds of annual deaths in Canada and worldwide [1,2,3]. Furthermore, nearly 12% of people aged 65 or older have lived with two or more chronic conditions during their lifetime [4]. Health behaviours such as smoking, excessive alcohol consumption, physical inactivity, and unhealthy eating are strongly associated with quality of life and are leading risk factors for chronic diseases [5]. With approximately four in five adult Canadians engaging in at least one of the health impacting behaviours associated with non-communicable chronic diseases, the prevalence of risky health behaviours is high [4].

The consequences and risk factors of multimorbidity (living with 2 or more chronic conditions) has been studied extensively [6,7,8]; however, research seeking to understand the relationships between life satisfaction, general health, and different combinations of health behaviours has received comparably little attention. Our daily lives are characterized by multiple interconnected social, personal, family, health, and work-related behaviours, each contesting for the limited energy, motivation, and time available [9]. Despite this, health risk behaviours are generally promoted and studied in isolation resulting in interventions and guidelines for healthy living siloed by individual behaviours. For example, historically Canada has had separate guidelines for alcohol consumption [10, 11], and physical activity and sleep [12], although recent guidelines are beginning to incorporate multiple health behaviours (e.g., guidelines for movement behaviours including sleep, sedentary activity, and physical activity [13]). The move towards guidelines that cover multiple health behaviours provide an opportunity to develop an evidence base to reflect an understanding of which health behaviours are interconnected, and how these patterns of interconnectedness are associated with health care utilization, life satisfaction, physical health, and mental health. This in turn may provide

new opportunities to promote multiple health behaviour change in guidelines and beyond. Indeed, interventions could be tailored to reflect the real-world complexities of health behaviours through an understanding of which behaviours are interconnected and for whom.

When investigating the interconnectedness of multiple health behaviours there are two general approaches: *person-centered approaches* which assess co-occurrence of behaviours and group people into categories, and *variable centered approaches* which assess co-variation of behaviours through the strength and direction of relationships between behaviours. Person-centered approaches include but are not limited to agglomerative cluster analysis, k-means, latent class analysis, behavioural profiles, and Gaussian mixture models. Applied to multiple health behaviours, person-centered approaches aim to segment people into categories based on similarity of behavioural features to identify focused intervention targets (behavioural combinations) and the sociodemographic patterns associated with each group [14-17].

However, to date research in this area often assesses different combinations of behaviours with heterogeneous measurements which result a wide array of behavioural clusters [14-17]. For example, Conry et al [14] investigated the clustering of alcohol use, physical activity, smoking, and unhealthy eating in a sample of Irish adults obtained from the 2007 National Survey of Lifestyle, Attitudes, and Nutrition. Six clusters were identified in this cross-sectional analysis which were labelled as: 1) '*multiple risk factor*' (moderate physical activity, moderate to high alcohol use, variable healthy eating); 2) '*mixed lifestyle*' (those who had never smoked, reported moderate physical activity, and variable alcohol consumption); 3) '*physically inactive*' (people with low levels of physical activity, poor eating, who reporting some smoking and high alcohol use); 4) '*temperate*' (moderately active and moderate drinkers who had never smoked); 5) '*former smokers*' (former smokers who reported high physical activity, moderate alcohol

use, and healthy eating); and 6) '*healthy lifestyle*' (characterized by people who had never smoked, high physical activity, highest healthy eating, moderate alcohol use). In another example, Buck and Frosini [15] examined the clustering of unhealthy eating, alcohol use, smoking, and physical inactivity among adults aged 16-74 using 2003-2008 data from the Health Survey of England. Findings indicated that in 2008, 63% of the sample engaged in one or two unhealthy behaviours, 25% engaged in three or more risky health behaviours, and 5% reported engaging in all four measured health behaviours. Only 7% of the sample did not engage in any measured risky health behaviours. Finally, in contrast to these data-driven approaches for clustering, Shaw and Agahi [18] used a descriptive approach called 'health behaviour profiles' to assess all possible combinations of co-occurring health risk behaviours in American adults 50 years or older using baseline data from the Health and Retirement study [19]. Overall, twelve health behaviour profiles were created using all combinations of physically active vs inactive, smokers vs non-smokers, and those who reported no vs moderate vs heavy alcohol consumption. The percentage of people represented in each profile varied widely with the six most prevalent profiles including: 1) 'physically active, non-drinkers, who smoke' (4.2%); 2) 'physically inactive, non-drinkers, who smoke' (6.5%); 3) 'physically inactive, moderate drinkers, who do not smoke' (8.6%); 4) 'physically active, moderate drinkers, who do not smoke' (10.1%); 5) 'physically active, non-drinkers, who do not smoke' (23.7%); and 6) 'physically inactive, non-drinkers, who do not smoke' (34.1%).

The second approach for modelling the relationships between health behaviours are *variable centered approaches*. The purpose of these types of analyses is to identify the associations between health behaviours to determine the strength and direction of the (usually linear) associations. Examples of variable-centered approaches include (but are not limited to)

correlations, multiple regression, network psychometrics, structural equation modelling, and lag-1 temporal time series analysis. The variable-centered approach can help to identify important associations between health behaviours such as the strong positive relationship between healthy eating and exercise [20]. Additionally, revealing the absence of linear relationships, as is the case with the relative independence of sedentary behaviours and physical activity [21], can inform research, policy and interventions to consider these behaviours as independent from one another.

Although the person- and variable-centred approaches are used in the multiple health behaviour literature, to our knowledge there have been no direct comparisons between them. It remains unknown whether person- and variable-centered approaches produce complimentary or divergent insights in the context of multiple health behaviours. It is also unknown whether person or variable analysis is more suitable for understanding the relationships between behaviours and health outcomes (e.g., life satisfaction and general health, onset of chronic conditions, BMI). To this end, we analyzed baseline and follow-up data from the Canadian Longitudinal Study on Aging (CLSA; [22]), to 1) identify patterns of co-occurring and co-varying behaviours and assess how sociodemographic and health indicators are associated with these patterns; 2) compare outputs from these two methods; and 3) compare the ability of clusters vs individual behaviours to predict future health indicators.

Methods

The Canadian Longitudinal Study of Aging (CLSA) is a longitudinal, nationally representative study designed to measure societal, biological, physical, and psychosocial factors related to healthy aging [22]. Baseline data collection for the CLSA was collected between 2010-2015 comprising two approaches. First, the 'tracking' cohort (n = 21,241) completed data collected via

an hour-long computer assisted phone interviews. Second, the 'comprehensive' cohort (n=30,097) completed an in-person interview lasting 90-minutes as well as a data collection site visit. Additionally, a 'maintaining contact questionnaire' was administered over the phone for the comprehensive and tracking cohorts. The maintaining contact questionnaire, tracking cohort, and comprehensive cohort form the baseline data collected used in this analysis. Detailed methodological information is shown in the published protocol [23]. With follow-up CLSA data subsequently made available, additional analysis was also performed using follow-up data. In the follow-up wave of data collection participants again completed the 'tracking' cohort (n = 17,050) and the 'comprehensive' cohort (n = 27,765) packages. A total of n = 6,523 participants who completed baseline data collection did not provide follow-up data. Data are available from the Canadian Longitudinal Study on Aging (www.clsa-elcv.ca) for researchers who meet the criteria for access to de-identified CLSA data

Participants

Participants were recruited through random-digit dialing, provincial health registries, and the Canadian Community Health Survey on Healthy Aging [22, 24]. Exclusion criteria for the CLSA included: residents living in three territories and First Nations reserves, full time members of the Canadian Armed Forces, people living with cognitive impairments, and individuals living in institutions (including 24-hour nursing homes; [22]). Participants included in the study were n=51,338 French and English-speaking Canadians (51% female) between the ages of 45-85 at time of enrollment. The average participant age is 62.98 years (SD = 10.4) with 26% between 45-54 years, 32% between 55-64 years, 23% between 65-74 years, and 18% between 75-85 years of age. A full description of demographic characteristics of the sample, as well as summary data across all measured variables is available in the CLSA baseline data report [25]. A total of

n = 6,523 participants who completed baseline data collection did not provide follow-up data. Thus, we have a sample size of n = 44,815 participants who completed follow-up data collection.

Variables

Health Behaviours. *Physical activity* and *sedentary behaviour* were measured as independent behaviours with the Physical Activity Scale for the Elderly (PASE; [26]) which assesses the frequency of sedentary behaviour, walking, light physical activity, moderate physical activity, strenuous physical activity, and exercise. Items asked participants to report on their activity levels over the previous 7 days on a 1 (never) to 4 (often, 5-7 days) scale. A Statistics Canada report focusing on the relationship between physical activity and lung functioning [27] merged light and moderate physical activity together and also merged strenuous physical activity and exercise together based on issues with question prompts and conceptual overlap between question items. To facilitate dimension reduction, we opted for a similar approach in which the PASE subscale items were merged to represent: sitting, walking, light/moderate physical activity (renamed 'light sports' to avoid confusion with 'light-to-moderate physical activity'; e.g.,[28]), and strenuous physical activity/ exercise. *Fruit and vegetable consumption* was assessed with one item from the Seniors in the Community Risk Evaluation for Eating and Nutrition questionnaire [29]. The item asks respondents how many servings of fruits and vegetables they eat in a day. The original scale was scored 1 (seven or more) to 7 (less than two); however, items were reverse coded such that higher scores indicate more fruit and vegetable consumption. *Smoking behaviour* was measured using a skip-question framework in the CLSA. We assigned a value of 0 to each respondent who responded 'no' to the question 'have you ever smoked a whole cigarette'. A similar approach has been applied to skip structure data when missing data represent the absence of a behaviour or psychological feature [30]. Participants who answered 'yes' to the

question 'have you ever smoked a whole cigarette' were subsequently asked whether they smoke not at all, occasionally, or daily, in the past 30 days. Ultimately, this created four levels distinguishing between people who have never smoked (coded 0), those who have not smoked within 30 days (1), those who smoke occasionally (2), and those who smoke daily (3). *Alcohol use* was assessed with a single item asking participants how often they drank alcohol in the past 12 months on a scale from 1 (almost every day) to 7 (less than once a week). Responses were reverse coded so that higher values indicate greater alcohol consumption. Finally, *sleep* was measured with a single item. Participants were asked how many hours of sleep they get, on average, during the past month and could respond with any value between 0-24. This variable was originally included in the analysis plan [23] but was subsequently removed due to high (41.5%) prevalence of missingness.

Sociodemographic Indicators. We included age, as grouped in the CLSA dataset (45-54; 55-64; 65-74; 75-85), sex (male/female), marital status (single, married or common-law, widowed, divorced, separated), household income (<\$20k, \$20-\$49k, \$50-\$99k, \$100-\$149k, \$150k+), retirement status (completely retired, partly retired, not retired), and working status (yes/no to 'are you currently working at a job or business').

Social Support. Participants responded to 19 questions from the Medical Outcomes Study (MOS) Social Support Survey [31]. The MOS is scored on 5 subscales: tangible social support, affection, positive social interaction, and emotional and informational support. A MOS 'overall support index' is also scored in the CLSA baseline dataset. To reduce the number of constructs in our analyses, we used the overall support index, scored from 0 (low support available) to 100 (high support available).

General Health and Life Satisfaction. Three single item measures were selected from the CLSA's general health module: an indicator of general health ('in general, would you say your health is excellent, very good, good, fair, or poor?'), mental health ('in general, would you say your mental health is excellent, very good, good, fair, or poor?'), and perceptions of healthy aging (in terms of your own healthy aging, would you say it is excellent, very good, good, fair, or poor?). Items were originally scored on a 1 (excellent) to 5 (poor) but were reverse coded. Additionally, a composite score from the Satisfaction with Life Questionnaire (SWLS; [32]) was used. The SWLS is scored according to Diener [33] from 1 (extremely dissatisfied) to 7 (extremely satisfied). A measure of body mass index (BMI) was used as an indicator of physical health.

Chronic Conditions. Participants reported any diagnosed chronic conditions during baseline and follow-up data collection (e.g., chronic respiratory disease, diabetes, cardiovascular disease, cancer). A summary variable which classifies people into those living with at least one chronic condition at follow-up (1) and those not living with any chronic conditions (0) was used as an indicator of health.

Analysis

Cluster Analysis

Prior to performing cluster analysis, all health behaviour variables (walking, sitting, light sports, exercise, smoking, alcohol) were standardized (i.e., mean centered) using the scale function in base R [34]. Listwise deletion was applied to missing data in health behaviour variables resulting in a remaining total sample of $n = 40,268$ for baseline behaviours. We then performed hierarchical agglomerative cluster analysis with five linkage methods (e.g., complete-linkage, single-linkage, average-linkage, centroid-linkage, and Ward's method) using 'hclust' function

supported by the package 'fastcluster' [35] to optimize performance. Gower distance was computed using the 'daisy' function in the 'cluster' package [36].

We examined cluster analysis outputs by looking at summary statistics for health behaviour variables for each linkage method. Two of the five linkage methods produced interpretable and useful clustering solutions (i.e., Ward and complete-linkage) while the other methods resulted in clusters with nearly all participants forming a single cluster with a small number of participants (often one per cluster) forming the remaining groups. Next, we employed a data-driven approach to determine the optimal number of clusters using complete linkage and Ward's method. We used the NbClust package [37] to provide the top three clustering solutions for both linkage methods, resulting in six options for combinations of linkage measures and k . Four of six options produced clustering solutions with 2-3 clusters with minimal variability across behaviours. Of the remaining two options (Ward $k=4$ and $k=7$) the research team opted for the clustering solution with 7 clusters as this option produced more behavioural variability (i.e., more clusters defined by higher/lower scores on a given behaviour).

Multinomial Logistic Regressions (Baseline)

We conducted four multinomial logistic regressions predicting cluster membership with baseline data to determine whether clusters are associated with 1) sociodemographic factors, 2) indicators of physical and mental health, 3) non-health behaviours, and 4) health care utilization. Analysis was performed using the 'multinom' function from the 'nnet' package [38]. Results are presented in appendix II.

Comparing Person and Variable Approaches

Comparisons between person- and variable-based approaches were conducted in two ways. First, baseline behaviours associated with one another via clusters are descriptively compared

with associations assessed with partial polychoric correlations. Partial polychoric correlations (ρ) were computed using the same baseline ($n= 40,268$) sample used for cluster analysis and visualized as a network (Figure 2). Polychoric correlations are appropriate for ordered categorical data [40]. Secondly, individual health behaviours and clusters were used as predictors in separate regression analyses to predict health outcomes (general health, healthy aging, and the presence of chronic conditions). Ordinary Least Squares (OLS) regression was used to predict general health and mental health while logistic regression was used to predict the presence of chronic conditions. The reference group for the regression analysis was Cluster 4 ('frequent alcohol use and infrequent walkers') due to most of that cluster's health behaviours being close to the final sample average. Variance explained (R^2) values are used to compare OLS models while AIC is used to compare model fit for logistic regression. No covariates were included in either model.

Results

Descriptive Analysis of Clusters

Standardized means and standard deviations for health behaviours in each cluster are presented in Table 1. Demographic information for each cluster is presented in Table 2. Ridge plots illustrating the density distributions of responses for each health behaviour across clusters are presented in Figure 1. Descriptive summaries highlighting the characteristics of the final sample and each cluster, interpreted using unstandardized scales, are provided below and are accompanied by radar charts using standardized scales in Figure 2.

Final sample. Following listwise deletion for missing values in health behaviour variables there were 40,268 people included in the final sample. On average, people engaged in walking activities 3-4 days a week ($M = 3.1$; $SD = 1.1$), sitting activities close to 4-5 days a week ($M =$

3.9; SD = 0.4), light sports were mostly performed between 'never' and 'seldom (1-2 days per week; M = 1.3; SD = 0.5), people engaged in strenuous exercise approximately 1 day per week (M = 1.6; SD = .8), ate 4 servings of fruits and vegetables per day (M = 4.0; SD = 1.8), were non-smokers (M = 0.9; SD = 0.8), and consumed alcohol near the middle of a 7 point scale ranging from monthly to daily (M = 4.2; SD = 2.0). The overall sample was balanced by sex (50.2% Female). The majority were in married or common law relationships (70.8%) with household incomes between \$50,000-\$99,000 per year (35.6%). The distribution of age groups was 27.1% (age 45-54), 32.8% (age 55-64), 23.5% (age 65-74), and 16.6% (age 75-85).

Cluster 1: Physically Active Healthy Eaters. People assigned to this cluster comprised 19% of all participants and engaged in more walking activities and exercise than the final sample and ate slightly more daily servings of fruits and vegetables. Specifically, people in Cluster 1 engaged in walking activities, on average, closer to 6-7 days a week than the 3-4 days overall average (M = 4.0; SD = 0.2) and strenuous exercise between 1-2 days a week and 3-4 days a week (M = 2.4; SD = 0.7). Average daily fruit and vegetable consumption was closer to 5 servings per day (M = 4.7; SD = 1.7) compared to the overall average of 4 servings per day. When compared to the proportion of people earning \$150,000 or more annually in the final sample (15.4%), more people in this cluster earned \$150,000 or more (20.5%).

Cluster 2: Frequent Walkers with Infrequent Strenuous Exercise and Infrequent Alcohol Use. People in this cluster represented 18% of participants and engaged in more frequent walking activities but less frequent strenuous exercise and alcohol consumption when compared to the overall sample. Walking activities were closer to 6-7 days a week than 3-4 days (M = 3.8; SD = 0.4) while the weekly average for strenuous exercise was closer to 'never' than 'seldom' (M = 1.2; SD = .4) and alcohol consumption was closer to monthly than daily (M

= 2.7; SD = 1.6). Demographically, there were 5.6% fewer males in this group than the final sample and 5% more people earning \$20,000-\$49,000 annually.

Cluster 3: Infrequent Alcohol Users, Walkers, Fruit/Vegetable Consumption, Light Sports, and Exercise. In this group (10% of participants), all health behaviours were performed less frequently than the group average except for slightly more sitting activities. Notably, the frequencies of walking, light physical activity, and strenuous exercise were each closer to 'never' than 'seldom (1-2 days)' (M = 1.3, 1.2, 1.3; SD = 0.5, 0.4, 0.6). Alcohol consumption was lower than average (M = 2.3; SD = 1.5) indicating that people in this group consumed alcohol closer to monthly than daily. Daily fruit and vegetable servings were closer to 3 servings a week (M = 3.4; SD = 1.7) than the overall average of 4 servings (M = 4.0; SD = 1.8). Additionally, the group was comprised of non-smokers. Demographically, there are more people aged 78-85 in this group (21.1%) compared to overall (16.6%), less Males (41.2%) than overall (49.8%), and the distribution of annual income was skewed towards lower income brackets compared to the final sample with 6.4% more people in Cluster 3 than the final sample earning \$20,000-\$49,000 and 6.3% less people earning \$150,000 per year or more.

Cluster 4: Frequent Alcohol Users and Infrequent Walkers. The largest of the seven clusters (27%) was defined by near average frequencies of health behaviours with two exceptions. First, the average frequency of walking activities was lower in this cluster with people engaging in walking activities 1-2 days per week (M = 2.1; SD = 0.8) compared to 3-4 days per week in the final sample (M = 3.1; SD = 1.1). Second, alcohol consumption was higher (M = 5.3; SD = 1.4) than the final sample average (M = 4.2; SD = 2.0) meaning that people in this cluster were closer to daily alcohol consumption than monthly consumption on the 1 (<

once a month) to 7 (almost every day) scale. There were slightly more people in married or common law relationships in this cluster (75.3%) compared to the overall sample (70.8%).

Cluster 5: Frequent Walkers with Infrequent Strenuous Exercisers with Higher Alcohol Use. Comprised of 13% of participants, Cluster 5 is similar to Cluster 2 with higher than average walking frequencies ($M = 4.0$; $SD = 0.1$) and lower than average strenuous exercise ($M = 1.2$; $SD = 0.3$). However, these two clusters are differentiated by alcohol consumption with the average drinking frequency for this group being 1 point away from 'almost every day' on a 1-7 scale ($M = 6.0$; $SD = 1.0$). Differences in demographics also distinguish these two clusters: there were fewer people aged 45-54 in this cluster compared to overall (20.0% vs 27.1%), more Males (55.7% vs 49.8%), and more people in married or common law relationships (76.2% vs 70.8%).

Cluster 6: Occasional and Daily Smokers who Infrequently Eat Fruits and Vegetables and Exercise. Nearly all participants who smoked occasionally or daily were included in this cluster (8% of total). Participants in this cluster also ate, on average, 1 less serving of fruits and vegetables per week ($M = 3.0$; $SD = 1.7$) compared to the overall sample ($M = 4.0$; $SD = 1.8$). Additionally, the average level of strenuous exercise in this group was closer to 'never' ($M = 1.3$; $SD = 0.6$) than the overall sample whose average was closer to 'seldom (1-2 days per week'; $M = 1.6$; $SD = 0.8$). Demographically, this group was skewed towards younger age groups (e.g., 37.3% aged 45-54 vs 27.1% final sample) and lower income brackets (e.g., 10.6% with income <\$20,000 vs 4.1% final sample). Lastly, this group was comprised of 15.7% less married or common law individuals, compared to overall, and 7.1% more single people and 5.8% more divorced participants.

Cluster 7: Infrequent Sedentary Activities. The smallest cluster by group membership (6%), people assigned to this cluster engaged in sitting activities, on average, between ‘seldom (1-2 days)’ and ‘sometimes (2-4 days)’ ($M = 2.6$; $SD = 0.6$) compared to the overall sample who, on average, participated in sitting activities closer to ‘often (5-7 days)’ ($M = 3.9$; $SD = 0.4$). Demographically, this group contains 10.3% more people aged 45-54 than the overall sample.

Partial Correlations (Baseline)

Partial polychoric correlations are visualized as a network in Figure 3. Correlations ranges from $\rho = -.13$ for smoking and fruit/vegetable consumption and $\rho = .14$ for exercise and fruit/vegetable consumption. The average correlation was $\rho = +/- .06$.

Predicting Follow-Up Indicators from Baseline Behaviours

Two sets of exploratory regression analyses were performed with baseline health behaviour clusters and individual health behaviours, predicting three health outcome indicators at follow-up. Summary statistics for health indicators are presented in Appendix III while regression outputs are summarized in Table 3. While controlling for age and sex, clusters predicted general health $R^2 = .05$, $F(8, 37466) = 228.4$, $p < .001$, mental health $R^2 = .02$, $F(8, 37485) = 78.82$, $p < .001$, and chronic conditions ($AIC = 12,629$). Individual behaviours also predicted general health $R^2 = .08$, $F(9, 37465) = 368.9$, $p < .001$, mental health $R^2 = .03$, $F(9, 37484) = 128.3$, $p < .001$, and chronic conditions ($AIC = 12,607$).

Discussion

Multiple health behaviours are robustly associated with the development of preventable non-communicable diseases and people engage in different combinations of these behaviours to varying degrees. To identify which behaviours are associated with one another to support multiple health behaviour change interventions, it may help to first identify which behaviours

co-occur and/or co-vary; co-occurrence and co-variation are assessed through person and variable centered approaches. In this study, we compared outputs from person centered (cluster analysis) and variable centered (partial correlation) approaches. Using representative data from the Canadian Longitudinal Study of Aging (CLSA), our cluster analysis produced seven groups of individuals based on similarities of frequencies they engage in key health behaviours (e.g., walking, sitting, light sports, exercise, fruit and vegetable consumption, smoking, and alcohol use). Overall, clusters were differentiated by six of the seven health behaviours included in the analysis with the most variability observed in weekly walking frequency, strenuous exercise, and alcohol consumption. Specifically, three clusters were partly characterized by walking frequency and two were characterized by strenuous exercise and alcohol consumption, respectively. Of the remaining health behaviours, there was little variability in weekly 'light sports' frequencies within the seven clusters, while one cluster was generally defined by a relative extreme of a single behaviour (infrequent sedentary activities). Sociodemographic characteristics varied across several clusters while associations between self-reported physical/mental health and cluster memberships were generally small.

In contrast, a partial correlation approach revealed small associations between health behaviours ranging from ranges from $\rho = -.13$ for smoking and fruit/vegetable consumption and $\rho = .14$ for exercise and fruit/vegetable consumption. Minimal effect sizes of interest are not well established in the multiple health behaviour change literature and it is unknown whether the small effect sizes observed in this study represent more than the 'crud factor', the idea that in the behavioral research everything correlates with everything else [40]. For example, in some fields within psychology a correlation less than $\rho = .10$ is not considered hypothesis

supporting as the observed relationships between theoretically relevant and irrelevant constructs can reach this level of effect size [41].

A comparison between co-occurrence and co-variation approaches reveals strengths and limitations to each approach. Regarding limitations, neither approach modelled some known phenomena. For example, the combination of high physical activity and frequent sedentary behaviour is common in individuals who participate in sports and strenuous exercise [49]; this distinction was not captured in the cluster analysis which illustrates the trade-offs between parsimony and nuance using hierarchical cluster analysis to describe co-occurring health behaviours. Additionally, the clustering algorithm revealed associations that were overlooked with variable centered analyses. Specifically, three clusters were defined by varying combinations of walking frequency and alcohol consumption while correlations between the two variables were negligible. Taken together, these findings highlight the need for alignment between methods and research objectives with person centered approaches more suitable for identifying sub-groups for intervention targeting purposes and variable centered approaches more appropriate for understanding the strength and direction of relationships between interconnected behaviours.

In addition to comparing insights into health behaviour associations from person and variable centered approaches we also investigated the ability of these approaches to predict future health outcomes. Between 2%-8% of variability in general and mental health at follow-up were accounted for by baseline clusters or individual health behaviours. Although individual behaviours accounted for more explained variance and had better model fit, these differences were small and firm conclusions regarding the comparative utility between approaches are not yet warranted on the basis of one study but point to opportunities for future research. When

classifying whether people reported any chronic conditions at follow-up, the baseline cluster of 'infrequent alcohol use, walking, light sports, and exercise' was the strongest predictor of having at least one condition while the individual behaviour of exercise was the strongest predictor of not having a chronic condition. To the best of our knowledge, no health behaviour clustering studies in adults have produced a grouping similar to the cluster we named 'infrequent alcohol use, walking, light sports, and exercise'. Behaviourally, this cluster was defined by little to no physical activity of any kind, non-smoking, and less frequent alcohol and fruits/vegetable consumption than average. The only behaviour that was above the final sample average were sedentary behaviours. People in this group tended to be older, have lower annual incomes, have higher BMI's, use healthcare services more frequently, not be employed, and be women. Taken together, the 'infrequent alcohol use, walking, light sports, and exercise' cluster may present a relatively homogenous behavioural sub-group to target for researchers and practitioners interested in conducting health behaviour interventions

This research is subject to limitations worth noting when interpreting the findings. First, many of the items selected for planned analysis are self-report which have known and inherent strengths and weaknesses [43]. Second, direct comparisons between multiple health behaviour studies is difficult due to variations in sample, measurement characteristics, and inconsistent naming conventions. Although heterogeneous samples and measurement variability may be useful for establishing the presence of robust phenomena in the form of co-occurring behaviours, we encourage future analysis to clearly label clusters to include each prominent health behaviour. For example, a cluster defined as 'occasional and daily smokers who infrequently eat fruits and vegetables and exercise' is more clearly defined than 'smokers with other risk behaviours'. Third, the health behaviours included in the cluster analysis were

not exhaustive (e.g., sleep hygiene, substance use, sexual risk behaviours were not included) and some behaviours were overrepresented such as of physical activity. While grouping people based on the types of physical activities scored with the PACE scale [26] enabled us to explore variability in walking activities, the way in which physical activity frequency was measured and the nature of the categories made it difficult to evaluate the health behaviours of this sample relative to behavioural guidelines.

Although a single comparisons between methods is not definitive, person centered approaches appear better suited than variable centered approaches or the purposes of identifying and prioritizing targets for multiple health behaviour change interventions. However, given the limitations of cluster analysis, we suggest that future person-centered research employ the 'behaviour profile approach' [18] with measures linked to behavioural guidelines in order to identify all possible combinations of 'meets guidelines/does not meet guidelines' for behaviours that contribute to negative health outcomes. Such approaches should ideally focus on datasets that include measures of health behaviour that provide an ability to directly link behaviour performance to thresholds recommended in guidelines. For variable centered approaches, the issue of measurement heterogeneity can be addressed through the use of meta-analysis. Although meta-analytic work on the associations between health behaviours has not yet conducted, some studies are planned for the future [44].

In conclusion, the scope, size, and rigour of the CLSA dataset provided an unprecedented opportunity to investigate how health behaviours are interconnected and to compare methods for modelling this interconnectivity. Our findings show how the population of older adults in Canada can be segmented by the multiple health behaviours that characterise people's lives and that these segmented clusters are socially patterned and associated with different health

outcomes. Comparing a person- and variable-centered approach can lead to insights about behaviours that may be overlooked with a single approach. Additionally, our analyses highlights opportunity for behavioural measures to be tied to national guidelines, which could lead to even more actionable analyses. The 'health behaviour profile' approach may be especially useful for future person-centered analysis, and a systematic review with meta-analysis could help establish associations between behaviours using a variable centered approach in future research. Understanding which behaviours co-occur and co-vary, and for whom, is an important first step towards developing tailored health behaviour change interventions. Future research will further develop our understanding of how interconnected health behaviours influence health outcomes over time using longitudinal data with multiple follow-up assessments.

Disclaimer

The opinions expressed in this manuscript are the author's own and do not reflect the views of the Canadian Longitudinal Study on Aging

References

1. Public Health Agency of Canada. (2016). How healthy are Canadians?: A trend analysis of the health of Canadians from a healthy living and chronic disease perspective. http://publications.gc.ca/collections/collection_2017/aspc-phac/HP40-167-2016-eng.pdf
2. World Health Organization. (2021). Non-communicable diseases: Fact sheet. Retrieved from: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>.
3. GBD 2015 Risk Factors Collaborators. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet (London, England)*, 388(10053), 1659. doi: [10.1016/S0140-6736\(16\)31679-8](https://doi.org/10.1016/S0140-6736(16)31679-8)
4. Public Health Agency of Canada. (2017). At-a-glance—The 2017 Canadian chronic disease indicators. *Health Promot Chronic Dis Prev Can*, 37, 248–251. <https://doi.org/10.24095/hpcdp.37.8.03>
5. Fisher EB, Fitzgibbon ML, Glasgow RE, Haire-Joshu D, Hayman LL, Kaplan RM, Nanney MS, Ockene JK. Behavior matters. *Am J Prev Med*. 2011 May 1;40(5):e15-30.. <https://doi.org/10.1016/j.amepre.2010.12.031>
6. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, Meinow B, Fratiglioni L. Aging with multimorbidity: a systematic review of the literature. *Ageing Res Rev*. 2011 Sep 1;10(4):430-9. <https://doi.org/10.1016/j.arr.2011.03.003>
7. Nunes BP, Flores TR, Mielke GI, Thumé E, Facchini LA. Multimorbidity and mortality in older adults: a systematic review and meta-analysis. *Arch Gerontol Geriatr*. 2016 Nov 1;67:130-8 <https://doi.org/10.1016/j.archger.2016.07.008>
8. Prados-Torres A, Calderón-Larrañaga A, Hanco-Saavedra J, Poblador-Plou B, van den Akker M. Multimorbidity patterns: a systematic review. *J Clin Epidemiol*. 2014 Mar 1;67(3):254-66. <https://doi.org/10.1016/j.jclinepi.2013.09.021>

9. Presseau J, Tait RI, Johnston DW, Francis JJ, Sniehotta FF. Goal conflict and goal facilitation as predictors of daily accelerometer-assessed physical activity. *Health Psychol.* 2013 Dec;32(12):1179. <https://doi.org/10.1037/a0029430>
10. Butt P, Beirness D, Gliksman L, Paradis C, Stockwell T. Alcohol and health in Canada: a summary of evidence and guidelines for low-risk drinking. Ottawa, ON: Canadian Centre on Substance Abuse. 2011.
11. Canadian Center on Substance Abuse and Addition. (2018). Canada's low risk alcohol consumption guidelines. Retrieved from: <https://www.camh.ca/-/media/files/canadas-low-risk-guidelines-pdf.pdf>
12. Tremblay MS, Warburton DE, Janssen I, Paterson DH, Latimer AE, Rhodes RE, Kho ME, Hicks A, LeBlanc AG, Zehr L, Murumets K. New Canadian physical activity guidelines. *Appl Physiol Nutr Metab.* 2011 Feb;36(1):36-46. <https://doi.org/10.1139/H11-009>
13. Ross R, Chaput JP, Giangregorio LM, Janssen I, Saunders TJ, Kho ME, Poitras VJ, Tomasone JR, El-Kotob R, McLaughlin EC, Duggan M. Canadian 24-Hour Movement Guidelines for Adults aged 18–64 years and Adults aged 65 years or older: an integration of physical activity, sedentary behaviour, and sleep. *Appl Physiol Nutr Metab.* 2020;45(10):S57-102. <https://doi.org/10.1139/apnm-2020-0467>
14. Conry MC, Morgan K, Curry P, McGee H, Harrington J, Ward M, Shelley E. The clustering of health behaviours in Ireland and their relationship with mental health, self-rated health and quality of life. *BMC Public Health.* 2011 Dec;11(1):1-0.. <https://doi.org/10.1186/1471-2458-11-692>
15. Buck D, Frosini F. Clustering of unhealthy behaviours over time. London: The Kings Fund. 2012 Aug:1-24.
16. Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. *Prev Med.* 2015 Dec 1;81:16-41.. <https://doi.org/10.1016/j.ypmed.2015.07.003>
17. Schneider S, Huy C, Schuessler M, Diehl K, Schwarz S. Optimising lifestyle interventions: identification of health behaviour patterns by cluster analysis in a German 50+ survey. *Eur J Public Health.* 2009 Jun 1;19(3):271-7. <https://doi.org/10.1093/eurpub/ckn144>

18. Shaw BA, Agahi N. A prospective cohort study of health behavior profiles after age 50 and mortality risk. *BMC Public Health*. 2012 Dec;12(1):1-0.. <https://doi.org/10.1186/1471-2458-12-803>
19. Heeringa SG, Connor JH. Technical description of the Health and Retirement Survey sample design. Ann Arbor: University of Michigan. 1995 May 16.
20. Sallis JF, Prochaska JJ, Taylor WC. A review of correlates of physical activity of children and adolescents. *Med Sci Sports Exerc*. 2000 May 1;32(5):963-75.
21. Santos DA, Silva AM, Baptista F, Santos R, Vale S, Mota J, Sardinha LB. Sedentary behavior and physical activity are independently related to functional fitness in older adults. *Exp Gerontol*. 2012 Dec 1;47(12):908-12. <https://doi.org/10.1016/j.exger.2012.07.011>
22. Raina PS, Wolfson C, Kirkland SA, Griffith LE, Oremus M, Patterson C, Tuokko H, Penning M, Balion CM, Hogan D, Wister A. The Canadian longitudinal study on aging (CLSA). *Can J Aging*. 2009 Sep;28(3):221-9.
23. van Allen Z, Bacon SL, Bernard P, Brown H, Desroches S, Kastner M, Lavoie K, Marques M, McCleary N, Straus S, Taljaard M. Clustering of Unhealthy Behaviors: Protocol for a Multiple Behavior Analysis of Data From the Canadian Longitudinal Study on Aging. *JMIR Res Protoc*. 2021 Jun 11;10(6):e24887. *PsyArXiv*. <https://doi.org/10.31234/osf.io/y2a7m>
24. Wolfson C, Raina PS, Kirkland SA, Pelletier A, Uniat J, Furlini L, Angus CL, Strople G, Keshavarz H, Szala-Meneok K. The Canadian community health survey as a potential recruitment vehicle for the Canadian longitudinal study on aging. *Can J Aging*. 2009 Sep;28(3):243-9. <https://doi.org/10.1017/S0714980809990031>
25. Raina P, Wolfson C, Kirkland S, Griffith L. The Canadian Longitudinal Study on Aging (CLSA) report on health and aging in Canada.
26. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. *J Clin Epidemiol*. 1993 Feb 1;46(2):153-62. [https://doi.org/10.1016/0895-4356\(93\)90053-4](https://doi.org/10.1016/0895-4356(93)90053-4)

27. Dogra S, Good J, Gardiner PA, Copeland JL, Stickland MK, Rudoler D, Buman MP. Effects of replacing sitting time with physical activity on lung function: An analysis of the Canadian Longitudinal Study on Aging. *Health Rep.* 2019 Mar 20;30(3):12-23. <https://www.doi.org/10.25318/82-003-x201900300002-eng>
28. Saint-Maurice PF, Troiano RP, Berrigan D, Kraus WE, Matthews CE. Volume of light versus moderate-to-vigorous physical activity: similar benefits for all-cause mortality?. *J Am Heart Assoc.* 2018 Apr 2;7(7):e008815. <https://doi.org/10.1161/JAHA.118.008815>
29. Keller HH, Goy R, Kane SL. Validity and reliability of SCREEN II (Seniors in the community: risk evaluation for eating and nutrition, Version II). *Eur J Clin Nutr.* 2005 Oct;59(10):1149-57. <https://doi.org/10.1038/sj.ejcn.1602225>
30. Borsboom D, Cramer AO. Network analysis: an integrative approach to the structure of psychopathology. *Annu Rev Clin Psychol.* 2013 Mar 28;9:91-121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
31. Sherbourne CD, Stewart AL. The MOS social support survey. *Soc Sci Med.* 1991 Jan 1;32(6):705-14. [https://doi.org/10.1016/0277-9536\(91\)90150-B](https://doi.org/10.1016/0277-9536(91)90150-B)
32. Diener ED, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. *J Pers Assess.* 1985 Feb 1;49(1):71-5. https://doi.org/10.1207/s15327752jpa4901_13
33. Diener E. Understanding scores on the satisfaction with life scale. Illinois: Psychology. University of Illinois at Urban Champaign
34. Team RC. R: A language and environment for statistical computing.
35. Müllner D. fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *J Stat Softw.* 2013 May 29;53:1-8. <https://doi.org/10.18637/jss.v053.i09>
36. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. Cluster: cluster analysis basics and extensions. R package version. 2012;1(2):56.
37. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw.* 2014 Nov 3;61:1-36. <https://doi.org/10.18637/jss.v061.i06>

38. Ripley B, Venables W, Ripley MB. Package 'nnet'R package version, 7, 3–12.
39. Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*. 1979 Dec;44(4):443-60. <https://doi.org/10.1007/BF02296207>
40. Orben A, Lakens D. Crud (re) defined. *Adv Methods Pract Psychol Sci*. 2020 Jun;3(2):238-47. <https://doi.org/10.1177/2515245920917961>
41. Ferguson CJ, Heene M. Providing a lower-bound estimate for psychology's "crud factor": The case of aggression. *Prof Psychol Res Pr*. 2021 Dec 1;52(6):620-6. <https://doi.org/10.1037/pro0000386>
42. Weiler R, Aggio D, Hamer M, Taylor T, Kumar B. Sedentary behaviour among elite professional footballers: health and performance implications. *BMJ Open Sport Exerc Med*. 2015 Jul 1;1(1):e000023. <https://doi.org/10.1136/bmjsem-2015-000023>
43. Del Boca FK, Noll JA. Truth or consequences: the validity of self-report data in health services research on addictions. *Addiction*. 2000 Nov;95(11s3):347-60. <https://doi.org/10.1046/j.1360-0443.95.11s3.5.x>
44. Silva C, Presseau J, Dinsmore J, van Allen Z, Marques M. Protocol for two interrelated systematic reviews of multiple health behaviour change interventions in healthcare. [10.31234/osf.io/7dwrv](https://doi.org/10.31234/osf.io/7dwrv)

Table Titles

Table 1. Standardized means and standard deviations for baseline health behaviours and clusters

Table 2. Counts and percentages of sociodemographic variables for final sample and each cluster.

Table 3. Regressions predicting follow-up health indicators from baseline behaviours and clusters

Table 3. Standardized means and standard deviations for baseline health behaviours and clusters

Cluster	n	Walking				Sitting				Light Sports				Exercise				Fruit/Veg				Smoking				Alcohol			
		M ¹	SD ¹	M ²	SD ²	M ¹	SD ¹	M ²	SD ²	M ¹	SD ¹	M ²	SD ²	M ¹	SD ¹	M ²	SD ²	M ¹	SD ¹	M ²	SD ²	M ¹	SD ¹	M ²	SD ²	M ¹	SD ¹	M ²	SD ²
1	7703	0.81	0.18	3.96	0.20	0.17	0.53	3.96	0.20	0.38	1.28	1.49	0.65	1.09	0.94	2.44	0.73	0.37	0.96	4.67	1.72	-0.17	0.62	0.74	0.46	0.03	0.95	4.29	1.92
2	7223	0.62	0.38	3.75	0.43	0.24	0.25	3.99	0.10	-0.25	0.71	1.16	0.36	-0.52	0.49	1.18	0.38	-0.12	0.99	3.79	1.78	-0.42	0.68	0.54	0.51	-0.77	0.77	2.67	1.56
3	4094	-1.53	0.41	1.32	0.47	0.22	0.34	3.98	0.13	-0.19	0.83	1.20	0.43	-0.33	0.82	1.33	0.64	-0.35	0.93	3.38	1.68	-0.38	0.68	0.58	0.51	-0.97	0.75	2.29	1.51
4	10723	-0.80	0.69	2.14	0.78	0.23	0.32	3.99	0.12	0.02	1.00	1.30	0.51	0.07	1.01	1.64	0.80	0.05	0.99	4.08	1.78	-0.15	0.60	0.75	0.45	0.51	0.69	5.26	1.39
5	5212	0.83	0.11	3.99	0.12	0.20	0.47	3.98	0.18	0.04	0.96	1.31	0.49	-0.55	0.40	1.16	0.31	0.15	0.95	4.27	1.70	-0.03	0.65	0.84	0.49	0.86	0.48	5.97	0.97
6	3079	-0.15	1.05	2.88	1.19	-0.04	1.07	3.88	0.41	-0.16	0.87	1.21	0.44	-0.33	0.80	1.33	0.63	-0.56	0.93	3.00	1.68	2.68	0.42	2.89	0.32	-0.08	1.07	4.08	2.15
7	2234	0.08	0.91	3.14	1.03	-3.32	1.52	2.62	0.58	-0.11	0.84	1.23	0.43	-0.05	0.90	1.55	0.71	-0.05	0.95	3.91	1.71	-0.26	0.65	0.67	0.49	-0.14	0.96	3.94	1.93
Full Sample	40268	0.00	1.00	3.05	1.13	0.00	1.00	3.90	0.38	0.00	1.00	1.29	0.51	0.00	1.00	1.59	0.79	0.00	1.00	4.00	1.80	0.00	1.00	0.86	0.75	0.00	1.00	4.24	2.02

¹Standardized Values²Unstandardized Values

Notes: Standardized means are colorized to highlight the size (color intensity) and direction (green = positive, red = negative) of mean differences. Cluster labels are as follows: Cluster 1 = Physically Active Healthy Eaters; Cluster 2 = Frequent Walkers with Infrequent Strenuous Exercise and Infrequent Alcohol Use; Cluster 3 = Infrequent Alcohol Use, Walking, Light Sports, and Exercise; Cluster 4 = Frequent Alcohol Use and Infrequent Walkers; Cluster 5 = Frequent Walkers with Infrequent Strenuous Exercise but Higher Alcohol Use; Cluster 6 = Occasional and Daily Smokers who Infrequently Eat Fruits and Vegetables and Exercise; Cluster 7 = Infrequent Sedentary Activities

Table 4. Counts and percentages of sociodemographic variables for final sample and each cluster.

	Full Sample		Cluster 1			Cluster 2			Cluster 3			Cluster 4			Cluster 5			Cluster 6			Cluster 7		
	N	%	N	%	%-%	N	%	%-%	N	%	%-%	N	%	%-%	N	%	%-%	N	%	%-%	N	%	%-%
Age Group																							
45-54	10923	27.1	2274	29.5	2.4	1852	25.6	-1.5	1013	24.7	-2.4	2760	25.7	-1.4	1040	20.0	-7.2	1147	37.3	10.1	837	37.5	10.3
55-64	13186	32.7	2613	33.9	1.2	2300	31.8	-0.9	1237	30.2	-2.5	3451	32.2	-0.6	1731	33.2	0.5	1182	38.4	5.6	672	30.1	-2.7
65-74	9461	23.5	1772	23.0	-0.5	1733	24.0	0.5	979	23.9	0.4	2591	24.2	0.7	1434	27.5	4.0	527	17.1	-6.4	425	19.0	-4.5
75-85	6698	16.6	1044	13.6	-3.1	1338	18.5	1.9	865	21.1	4.5	1921	17.9	1.3	1007	19.3	2.7	223	7.2	-9.4	300	13.4	-3.2
NA	0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
Employment Status																							
Yes	16401	40.7	3375	43.8	3.1	2843	39.4	-1.4	1524	37.2	-3.5	4310	40.2	-0.5	1794	34.4	-6.3	1411	45.8	5.1	1144	51.2	10.5
No	1609	4.0	236	3.1	-0.9	325	4.5	0.5	204	5.0	1.0	327	3.0	-0.9	158	3.0	-1.0	267	8.7	4.7	92	4.1	0.1
NA	22258	55.3	4092	53.1	-2.2	4055	56.1	0.9	2366	57.8	2.5	6086	56.8	1.5	3260	62.5	7.3	1401	45.5	-9.8	998	44.7	-10.6
Sex																							
F	20220	50.2	3729	48.4	-1.8	4029	55.8	5.6	2409	58.8	8.6	5025	46.9	-3.4	2310	44.3	-5.9	1563	50.8	0.6	1155	51.7	1.5
M	20048	49.8	3974	51.6	1.8	3194	44.2	-5.6	1685	41.2	-8.6	5698	53.1	3.4	2902	55.7	5.9	1516	49.2	-0.5	1079	48.3	-1.5
NA	0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
Annual Income																							
<\$20k	1669	4.1	208	2.7	-1.4	409	5.7	1.5	280	6.8	2.7	229	2.1	-2.0	122	2.3	-1.8	326	10.6	6.4	95	4.3	0.1
\$20-\$49k	8663	21.5	1360	17.7	-3.9	1912	26.5	5.0	1144	27.9	6.4	1973	18.4	-3.1	931	17.9	-3.7	865	28.1	6.6	478	21.4	-0.1
\$50-\$99k	13928	34.6	2526	32.8	-1.8	2464	34.1	-0.5	1407	34.4	-0.2	3823	35.7	1.1	1941	37.2	2.7	999	32.4	-2.1	768	34.4	-0.2
\$100-\$149k	7439	18.5	1636	21.2	2.8	1117	15.5	-3.0	600	14.7	-3.8	2195	20.5	2.0	1038	19.9	1.4	440	14.3	-4.2	413	18.5	0.0
\$150k+	6214	15.4	1578	20.5	5.1	841	11.6	-3.8	375	9.2	-6.3	1886	17.6	2.2	912	17.5	2.1	260	8.4	-7.0	362	16.2	0.8
NA	2355	5.8	395	5.1	-0.7	480	6.6	0.8	288	7.0	1.2	617	5.8	-0.1	268	5.1	-0.7	189	6.1	0.3	118	5.3	-0.6
Marital Status																							
Single	3230	8.0	580	7.5	-0.5	685	9.5	1.5	355	8.7	0.7	681	6.4	-1.7	292	5.6	-2.4	465	15.1	7.1	172	7.7	-0.3
Married/CL	28520	70.8	5692	73.9	3.1	4843	67.0	-3.8	2673	65.3	-5.5	8076	75.3	4.5	3972	76.2	5.4	1697	55.1	-15.7	1567	70.1	-0.7
Widowed	3600	8.9	568	7.4	-1.6	765	10.6	1.7	501	12.2	3.3	868	8.1	-0.8	428	8.2	-0.7	274	8.9	-0.0	196	8.8	-0.2
Divorced	3879	9.6	684	8.9	-0.8	753	10.4	0.8	458	11.2	1.6	871	8.1	-1.5	413	7.9	-1.7	476	15.5	5.8	224	10.0	0.4
Separated	1027	2.6	175	2.3	-0.3	177	2.5	-0.1	105	2.6	0.0	225	2.1	-0.5	107	2.1	-0.5	165	5.4	2.8	73	3.3	0.7
NA	12	0.0	4	0.1	0.0	0	0.0	-0.0	2	0.0	0.0	2	0.0	-0.0	0	0.0	-0.0	2	0.1	0.0	2	0.1	0.1

Notes: %-% denotes the % difference between final sample and a given cluster and are colorized to highlight the size (color intensity) and direction (green = positive, red = negative) of the differences.

Cluster labels are as follows: Cluster 1 = Physically Active Healthy Eaters; Cluster 2 = Frequent Walkers with Infrequent Strenuous Exercise and Infrequent Alcohol Use; Cluster 3 = Infrequent Alcohol Use, Walking, Light Sports, and Exercise; Cluster 4 = Frequent Alcohol Use and Infrequent Walkers; Cluster 5 = Frequent Walkers with Infrequent Strenuous Exercise but Higher Alcohol Use; Cluster 6 = Occasional and Daily Smokers who Infrequently Eat Fruits and Vegetables and Exercise; Cluster 7 = Infrequent Sedentary Activities. NA = Missing data.

Table 3. Regressions predicting follow-up health indicators from baseline behaviours and clusters

	General Health		Mental Health		Chronic Conditions	
	β	p	β	p	OR	p
Cluster 1	.22	<.001	.10	<.001	.83	<.01
Cluster 2	-.12	<.001	-.09	<.001	1.08	.35
Cluster 3	-.32	<.001	-.18	<.001	1.67	<.001
Cluster 5	.10	<.001	.04	.02	1.00	.97
Cluster 6	-.44	<.001	-.28	<.001	1.16	.15
Cluster 7	.01	.63	-.04	.07	.89	.30
Age	-.01	<.001	.00	.01	1.09	<.001
Sex	-.07	<.001	.05	<.001	.56	<.001
	$R^2 = .05$		$R^2 = .02$		AIC = 12,629	
Walking	.07	<.001	.02	<.001	.92	<.001
Sitting	-.03	<.01	-.01	.46	1.16	<.01
Exercise	.16	<.001	.07	<.001	.85	<.001
Light PA	.08	<.001	.04	<.001	.95	.34
Fruit/Vegetable	.05	<.001	.04	<.001	1.01	.36
Smoking	-.13	<.001	-.08	<.001	1.06	.07
Alcohol	.05	<.001	.03	<.001	.96	<.01
Age	-.01	<.001	.00	<.001	1.09	<.001
Sex	-.05	<.001	.08	<.001	.58	<.001
	$R^2 = .08$		$R^2 = .03$		AIC= 12,607	

Cluster labels are as follows: Cluster 1 = Physically Active Healthy Eaters; Cluster 2 = Frequent Walkers with Infrequent Strenuous Exercise and Infrequent Alcohol Use; Cluster 3 = Infrequent Alcohol Use, Walking, Light Sports, and Exercise; Cluster 4 = Frequent Alcohol Use and Infrequent Walkers; Cluster 5 = Frequent Walkers with Infrequent Strenuous Exercise but Higher Alcohol Use; Cluster 6 = Occasional and Daily Smokers who Infrequently Eat Fruits and Vegetables and Exercise; Cluster 7 = Infrequent Sedentary Activities. OR = Odds Ratio. All predictors listed in table 3 are those included in the model.

Figure 1. Health behaviour ridge (density) plots across clusters.

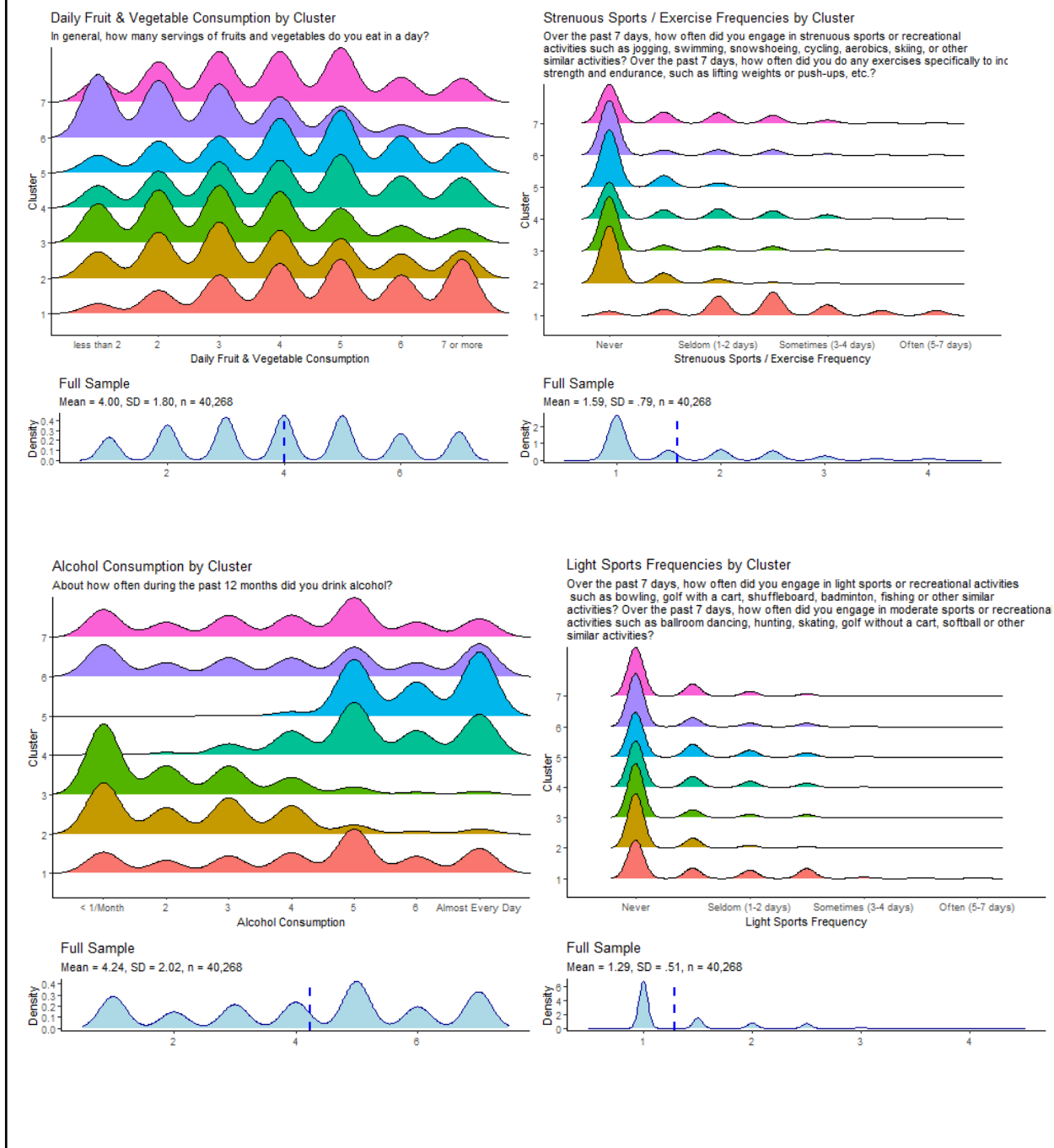


Figure 1 Cont. Ridge (density) plots for each health behaviour across clusters.

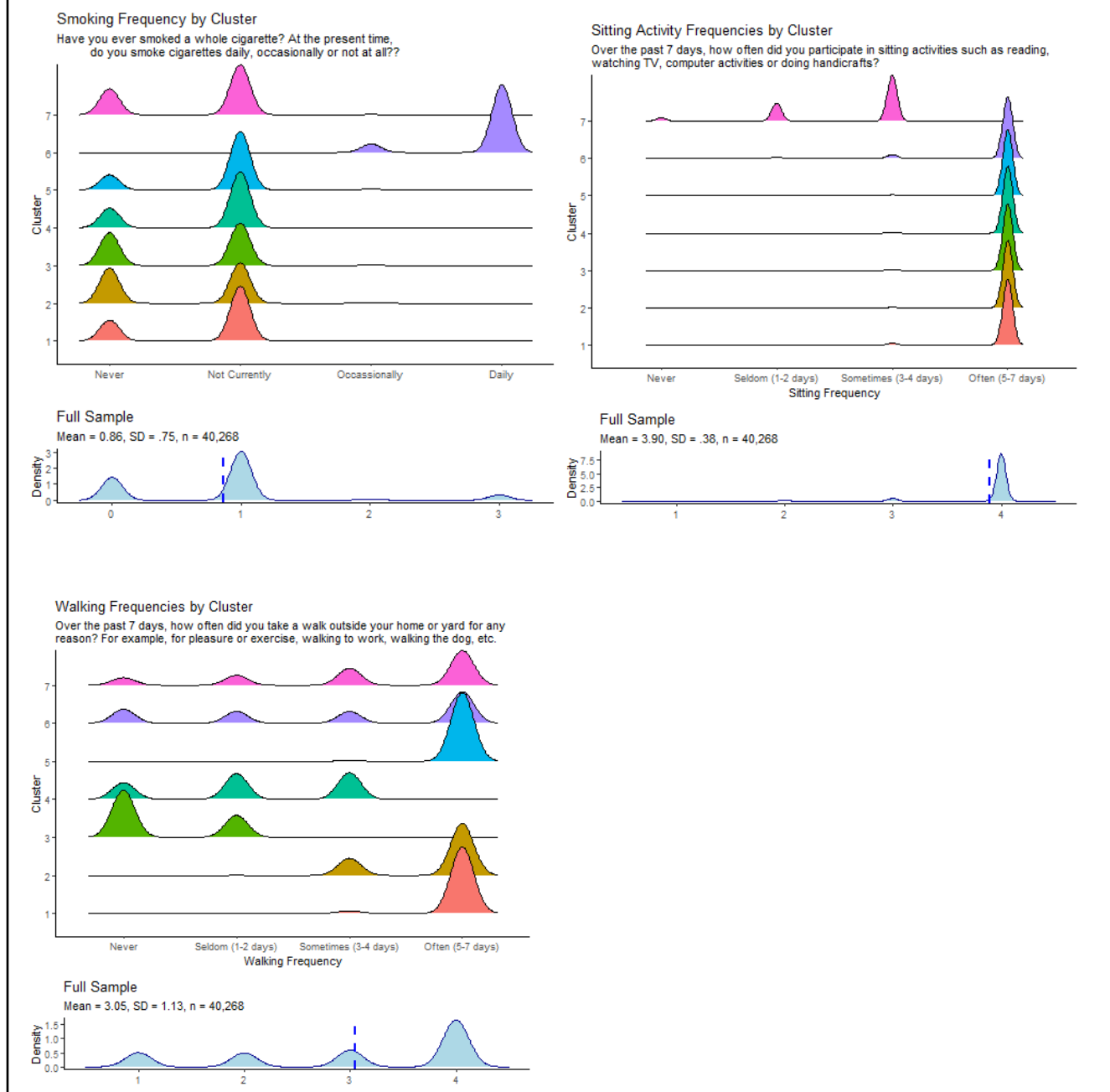


Figure note: Cluster labels are as follows: Cluster 1 = Physically Active Healthy Eaters; Cluster 2 = Frequent Walkers with Infrequent Strenuous Exercise and Infrequent Alcohol Use; Cluster 3 = Infrequent Alcohol Use, Walking, Light Sports, and Exercise; Cluster 4 = Frequent Alcohol Use and Infrequent Walkers; Cluster 5 = Frequent Walkers with Infrequent

Strenuous Exercise but Higher Alcohol Use; Cluster 6 = Occasional and Daily Smokers who Infrequently Eat Fruits and Vegetables and Exercise; Cluster 7 = Infrequent Sedentary Activities.

Figure 2. Radar plots for each cluster (standardized means).

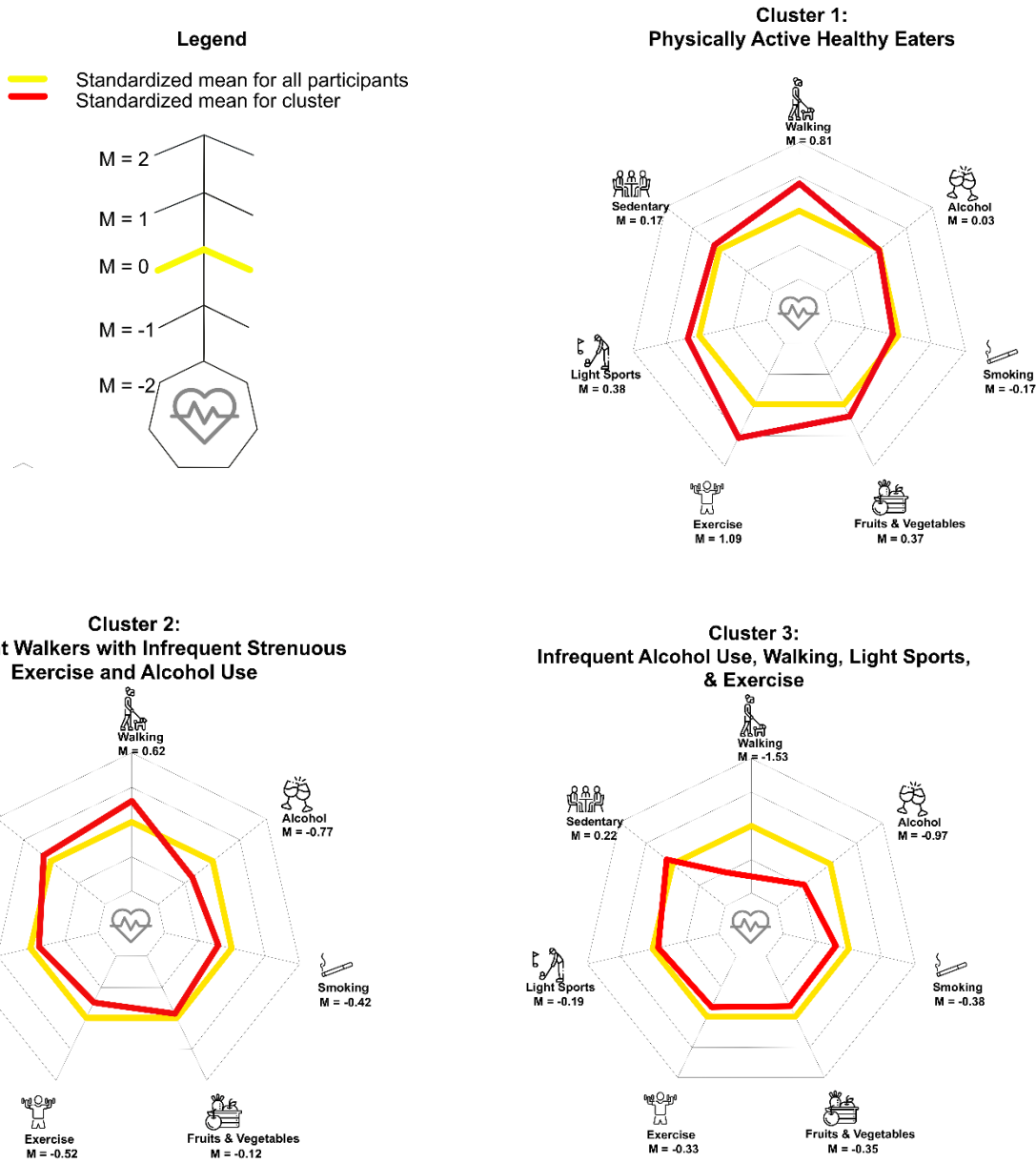
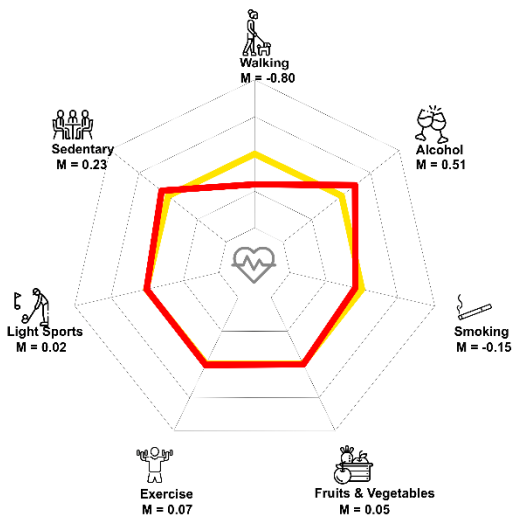
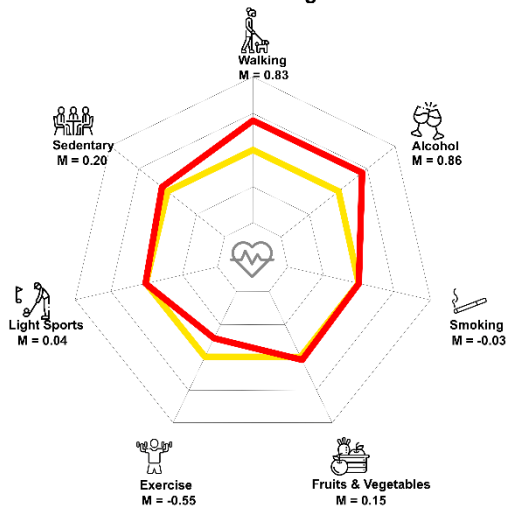


Figure 2 Continued. Radar plots for each cluster (standardized means).

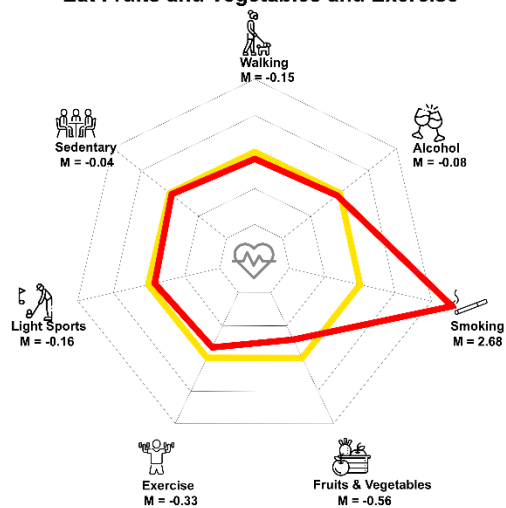
Cluster 4:
Frequent Alcohol Use & Infrequent Walkers



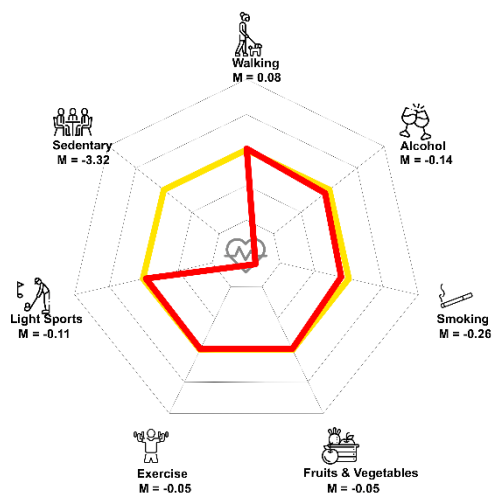
Cluster 5:
Frequent Walkers with Infrequent Strenuous Exercise but Higher Alcohol Use



Cluster 6:
Occasional and Daily Smokers who Infrequently Eat Fruits and Vegetables and Exercise



Cluster 7:
Infrequent Sedentary Activities



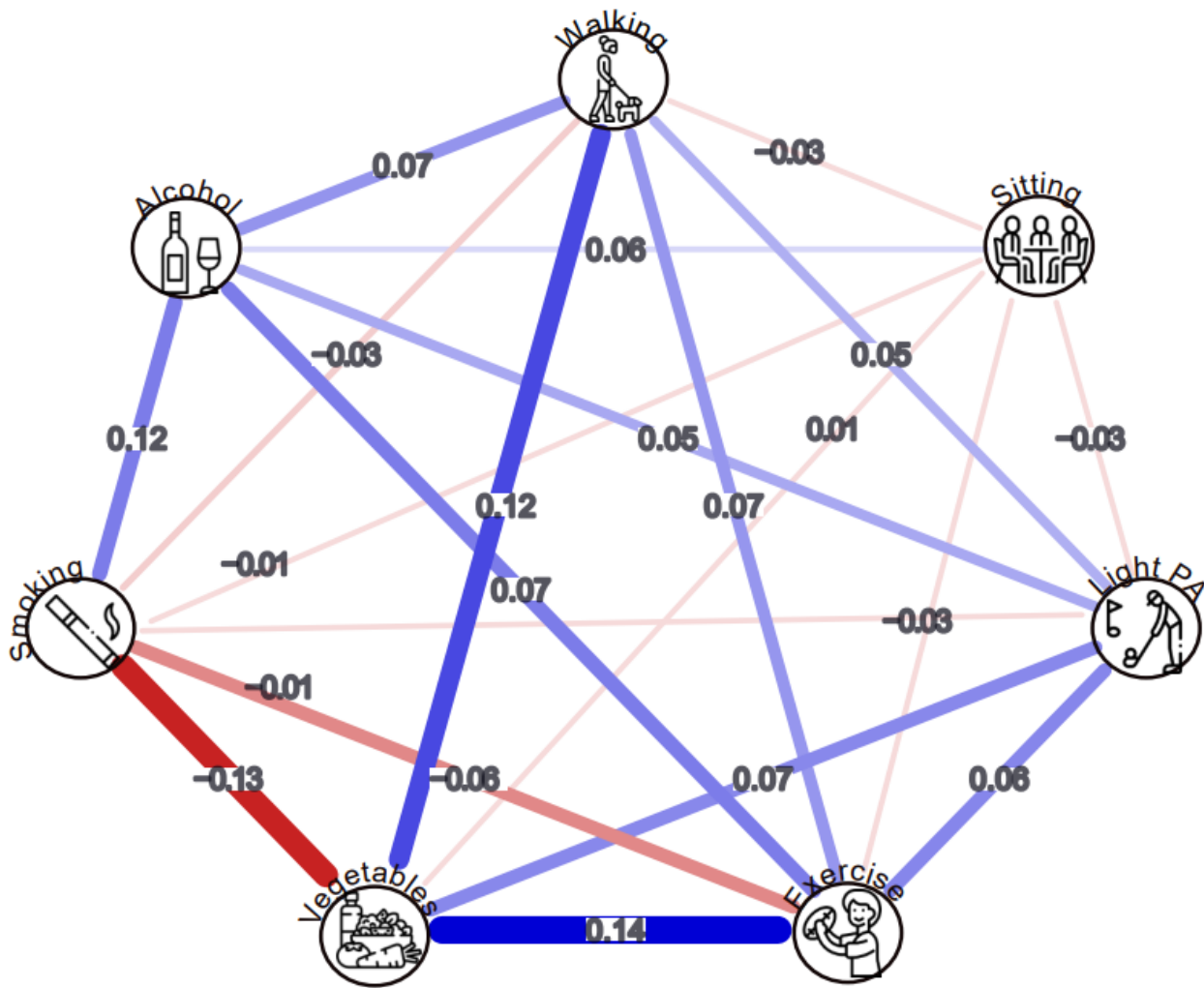


Figure 3. Partial polychoric correlation visualization of baseline health behaviours. Values represent partial polychoric correlations between variables. Red lines represent negative correlations and blue lines represent positive correlations. Line width corresponds to correlation strength.