

Supplement B

Participation in voluntary work and cognitive decline: A Longitudinal Analysis in the Aging Population

Analysis plan

Guusje van Helden

Supervisor: Dr. Nynke Smidt and Prof. Dr. Aat Liefbroer

Study aims

1. Examine the relationship between participation in voluntary work and cognitive decline among individuals between 60-80 years of age and whether this relationship is moderated by age, education and gender.
2. Examine whether there is a dose-response relationship between participation in voluntary work in terms of hours per week and cognitive decline and whether this relationship is moderated by age, education and gender.

Research questions

Is there a relationship between participation in voluntary work and cognitive decline among older individuals?

And is this relationship moderated by age, education and sex?

Study design

Longitudinal study

Population

Baseline adult Lifelines participants who filled in the The Daily Activities and Leisure Activities Addon Study (DALAS) questionnaire (60-80 yr).

Therefore they either have filled in the 2A questionnaire 1 or 2A questionnaire 2 before someone can participate in the DALAS questionnaire.

Individuals who have valid test results on the Cogstate brief battery at both 2A and 3A

Exposure

Voluntary work:

Voluntary work: Defined as “Unpaid non-compulsory work; that is, time individuals give without pay to activities performed either through an organization or directly for others outside their own household” (International Labour Office Geneva, 2011).

Operationalization:

Participation in voluntary work can be defined as “unpaid non-compulsory work; that is that is, time individuals give without pay to activities performed either through an organization or directly for others outside their own household” (International Labour Office Geneva, 2011).

Volunteering can be categorized into two types: formal and informal. Formal volunteering involves activities within an organizational context, while informal volunteering includes activities outside of an organizational context. In this research, both types of volunteering are considered as ‘voluntary work’. -For determining individuals’ volunteer status, questions obtained from the Lifelines DALAS questionnaire are used. The two specific questions to determine volunteer status are: “Have you ever done any volunteer work?” (yes / no) and “Do you currently still do volunteer work?” (yes / no). Individuals who answered ‘no’ on the first question are considered to be never volunteers. Individuals who answered ‘yes’ on the first question but ‘no’ on the second question, are considered to be past volunteers.

Individuals who answered ‘yes’ on both questions are considered to be current volunteers. For the analysis two dummies have been made, using never volunteers as the reference category. In **figure 1** the operationalization can be visualized.

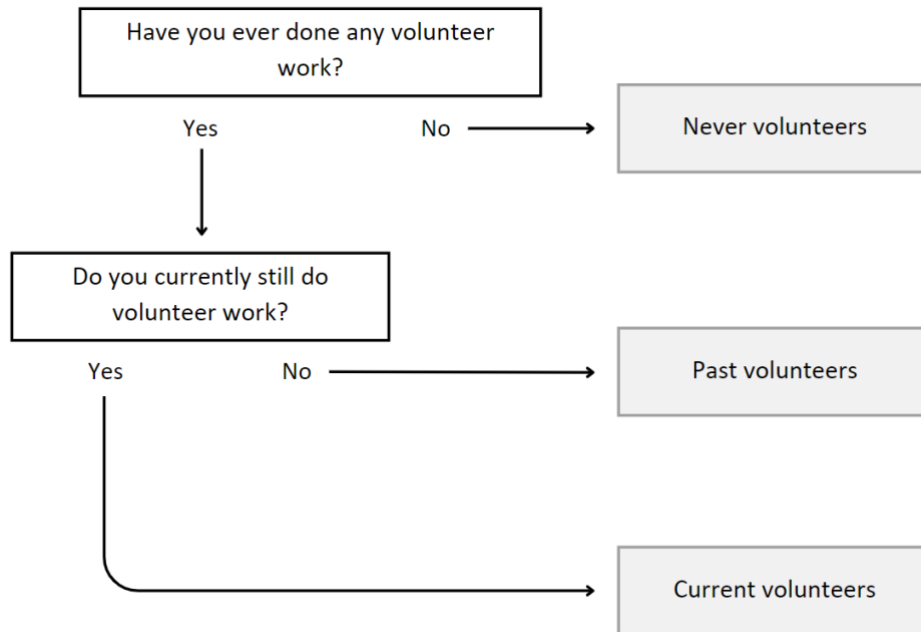


Figure 1 – Operationalization of volunteer status

Measures

Independent variables; participation in voluntary work

A new categorical variable will be made to identify the different types of volunteers.

- 0 = never volunteers (reference)
- 1 = past volunteers
- 2 = current volunteers

Dependent variables; cognitive decline

Cogstate is a battery of computerized cognitive tests used in ~68,000 participants of 13+ years of age during visits to 2A and 3A. Cogstate consists of four subtests, each measuring different cognitive domains. The tests were completed in the following order:

1. Detection task
2. Identification task
3. One back
4. One card learning task

All tests are based on playing card formats. Participants could answer using the computer keyboard represented on the screen. The ‘k’ key was used for yes and the ‘d’ key for no. The outcome measures for the tests are number of errors/correct responses and reaction time (Lifelines Wiki, n.d.).

The detection test (DET)

A reaction time test that measures psychomotor functioning and speed of processing. Participants focus on the center of the screen and respond when the card turns face up, by pressing the “yes” key. The task ends after recording 35 correct trials. The main outcome, reaction time (in milliseconds), which was normalized using log10 transformation (Kuiper et al., 2017). Therefore, lower scores indicate better performance.

The identification task (IDN)

A choice reaction task that measures visual attention. Participants focus on the center of the screen, responding to the question “Is the card red?” with a ‘Yes’ or “No’ key press. The task stops when 30 correct responses have been recorded. The primary outcome is reaction time (in milliseconds), which has also been normalized using log10 transformation (Kuiper et al., 2017). Therefore, lower scores indicate better performance.

One Back task (OBK or ONB)

This test assesses attention and working memory. Participants focus on the card at the center of the screen and they are asked: ‘Is this card the same as that on the immediately previous trial?’. A ‘yes’ press is required for a match and a ‘no’ if it is not a match. The task ends after 40 correct trials. The primary outcome is the proportion of correct answers, which is normalized using arcsine transformation (Kuiper et al., 2017). Therefore, higher scores indicate better performance.

One Card Learning task (OCL)

This is a visual learning and memory task. Participants focus on the card on the screen and answer the question: “Have you seen this card before in this task?”. If the answer is yes, they press the ‘yes’ key, if it is no they press the ‘no’ key. The task ends after 42 trials. The primary outcome measure is the proportion of correct answers, which is normalized using arcsine transformation (Kuiper et al., 2017). Therefore, higher scores indicate better performance.

Cognitive decline

The dependent variable for my research is cognitive decline. For the DET and IDN the outcome is reaction time, which is normalized using log10 transformation. This is probably transformed to normalize the distribution of the data. However, my dependent variable will be cognitive decline, for which I will create a new variable by subtracting the longitudinal Cogstate results for each subtest. It is therefore easier if I use the results for the DET and IDN that are not normalized.

As mentioned before, the DET assesses psychomotor functioning and speed of processing, while the IDN focuses on visual attention. In the context of normal cognitive ageing, it is suggested that cognitive functions as processing information to make a decision, such as measures of speed processing, will decline as we get older (Van Hooren et al., 2007). As the primary outcomes of both of these tests will be the speed of reaction, it makes more sense the values of 3A will be higher than 2A. As I define my outcome variable as cognitive decline, it is easier to interpret that a higher positive outcome means more cognitive decline. Interpreting it this way also means that negative values are associated with cognitive improvement. Therefore, I will subtract the Cogstate results of 3A from 2A:

$$3A - 2A = (\text{decline}) \text{ in reaction time (in milliseconds) for DET and IDN}$$

The OBK/ONB and OCL have different outcome measures, namely the proportion of correct answers, which is normalized using arcsine transformation. Usually, this is also used to make the distribution of the data more symmetric and makes it easier to use for statistical analysis. However, I will also create a new variable for these subtests by subtracting the longitudinal Cogstate results. It is therefore easier if I use the results for the OBK/ONB and OCL that are not normalized.

The ONB assesses working memory, while the OCL assesses visual learning and memory. These cognitive functions are also thought to decline in normal cognitive ageing (Van Hooren et al., 2007). As the primary outcome for both of these tests are based on the proportion of correct answers, I suggest that individuals have lower scores at 3A, compared to their scores at 2A. As mentioned earlier, my primary focus will be on cognitive decline, interpreting a larger positive value as an indication of a more significant cognitive decline. Therefore, I will subtract the Cogstate results of 3A from 2A

$$2A - 3A = (\text{decline}) \text{ in proportion of correct responses (hit rate) for ONB and OCL.}$$

As the scores of all subtests of 2A might influence the degree of cognitive decline, I will take the scores of 2A into account in my linear regression model as a potential confounder.

Due to personal contact with cogstate, I found out that they recommend using the speed measure of the ONB test to measure change over time. Therefore, decline for the ONB test is measured calculating $3A - 2A$.

Moderators

Moderator analysis investigates whether age influences the strength or direction of the relationship between participation in voluntary work and cognitive decline.

- Age groups (0 = 60 - 69 years, 1 = 70 – 80 years)
- Sex (0 = male, 1 = female)
- Educational attainment (0 = low , 1 = middle, 2 = high)

Model build-up

Is there a relationship between participation in voluntary work and cognitive decline among older individuals?

Model 0: univariable regression analysis

- Independent variable (predictor): participation in voluntary work dummies
- Dependent variable (outcome): cognitive decline
- no potential confounders / covariates added

First add all the covariates/confounders separately to see and understand the effect of them. So a univariable regression analysis will be done first for all potential confounders / covariates.

Model 1: multivariable regression analysis

- Add all potential confounders/covariates jointly (age, sex, educational attainment, income, employment status, marital status, self-reported health, presence of cardiovascular risk factors, hearing problems, physical inactivity, obesity, depressive symptoms, social network size, loneliness, social participation, informal caregiving, taking care of grandchildren, history of voluntary work, church attendance and the Cogstate results of 2A)

Is the relationship moderated by age, sex and education?

Model 2: multivariable regression analysis + the effect of moderators

- + Interaction terms – analyze each interaction term separately
 - a. Participation in voluntary work dummies * age groups
 - b. Participation in voluntary work dummies * sex
 - c. Participation in voluntary work dummies * educational attainment

Analysis plan

Statistical analysis

Uni- and multivariable regression model

- The regression model will include cognitive tests scores from the Cogstate as the dependent variable. Independent variables will include the three groups of participation of in voluntary work (categorical)
- Multivariable linear regression analysis will be conducted using SPSS, allowing investigation of the contribution of each predictor variable while controlling for others

Effect in terms of P-values and effect sizes

- Statistical significance will be determined based on a threshold of $p < 0.05$. Effect sizes, such as standardized regression coefficients, will be assessed to quantify the strength and direction of the relationships
- Results will be presented in a regression table, highlighting p-values and effect sizes, and associated confidence intervals.

A lot of variables will not be normally distributed. I first have to investigate and visualize the descriptive and discuss this with my supervisor to make decisions about what to do with the data.

The dataset

I am mainly going to use the Lifelines DALAS questionnaire data. Initially, the missing values were defined by tokens. However, this has already been transformed to dots '.', so I do not need to transform those. I am also going to use some variables from the Lifelines 1A dataset, the tokens have already been transformed. I am going to use some variables from Lifelines 2A dataset, these missing values have also been transformed to dots. Additionally, I am going to use the Cogstate from the Lifelines 3A dataset, this data has not been finished yet. I will receive the additional data at the beginning of February. I will have to transform the missing data that have been defined by tokens into dots myself. For this procedure Lifelines has a manual.

Since I want to use different variables from different datasets, I have to merge the variables I need from Lifelines 1A, 2A and 3A into the DALAS dataset. The DALAS dataset contains all the participants I want to include. Also, for this procedure Lifelines has a manual called "merge instruction" that I am going to follow.

I will copy this merged dataset without tokens and make it my "cleaning dataset". I will work for the next steps in this dataset.

Descriptive statistics

The descriptive will provide insight in how the data is distributed and whether there is missing data and outliers. For every variable from the raw data, I am going to use, I will provide descriptive statistics.

For continuous variables

- Sample size
- Mean
- Median
- Standard deviation
- IQR
- Minimum
- Maximum
- Kurtosis
- Skewness
- Missing

For categorical variables

- Frequency
- Percentage
- Missing

Skewness and Kurtosis

Skewness quantifies the degree of asymmetry in a distribution, indicating whether the left and right sides are symmetrical or not. Skewness in a distribution can lead to right (positive), left (negative), or no skewness. A distribution is right-skewed when its tail extends more to the right of its peak, while it's left-skewed when the tail is longer on the left side of the peak. For a normal distribution, skewness should be around 0 (Turney, 2023). However, the range is around $-1 / -2$ to $1 / 2$.

Kurtosis tells us how much a distributions' tails stand out, which reflects the frequency of outliers. Excess kurtosis specifically measures how much the tails of a distribution deviate from those of a normal distribution. For a normal distribution kurtosis should be around 3 (mesokurtic) (Turney, 2023a).

Visual inspection

Before I do any statistical analysis, I have to visualize the distribution of the data I want to use and check for normality. There are several methods to assess normality.

Visual Inspection for continuous variables

- Histograms: Plot a histogram of your data. A bell-shaped curve is an indication of normal distribution. However, small sample sizes may not always show a clear pattern.
- Q-Q Plots (Quantile-Quantile Plots): These plots compare the quantiles of your data against the quantiles of a normal distribution. A straight line suggests normality.

Visual Inspection for categorical variables

- Use bar charts or pie charts to visualize the distribution of the categories

Identify outliers

As mentioned before, the kurtosis value of each variable reflects the frequency of outliers. However, my making boxplots the outliers will be clearer, and the specific cases can be identified.

Create boxplots to identify skewness and outliers

- Ends of the whiskers show the minimum and maximum scores
- In SPSS the outliers are identified by a small open circle, the extreme outliers are identified by a star
- SPSS - graphs - chart builder - choose boxplot under - choose your variables
- Look at the values with an asterix - those are the outliers, the case numbers are next to them
- Look for the case numbers - go to edit - case - type in case number - identify the variable value for this case number
- In a normally distributed dataset, values that are more than 2 or 3 SD from the mean are considered as outliers
- Outliers can be identified as values below $Q1 - 1.5 * IQR$ or $Q3 + 1.5 * IQR$

Options for outliers

- 1) Leave it in: if it is a legitimate outlier
 - a. If the outlier skews the dataset, use a non-parametric test on skewed data
- 2) Correct: data entry errors
 - a. If it can be determined what the actual value should be
- 3) Winsorize: trim the outliers to match the highest or lowest representative values
- 4) Remove outlier: when an outlier is a multivariate outlier
 - a. This case has extreme scores on several variables, not just one

If it is obvious that an outlier is the result of a measurement error or error of the participant, the datapoints will be removed from the analysis. As there is only one or two cases that are extreme outliers in an variable, the distribution of the data will not be influenced by it. It is not needed to remove these for the analysis.

Always report what I will do with the outliers

Identify missing values

With the descriptives you get an overview of the missing data for each variable. Inspect and report the overall numbers and percentage of missing data. Additionally, inspect and report the percentage of variables with missing data and the percentage of participants with missing data. Additionally, explore patterns of missing values. There might be a pattern within participants or variables.

There are different types of missing data:

- 1) Missing completely at random (MCAR)
 - a. Missing data is not randomly distributed across variables, unrelated to other factors.
- 2) Missing at random (MAR)
 - a. Missing data is not randomly distributed but can be explained by other observed variables
- 3) Missing not at random (MNAR)
 - a. Missing data systematically differs from observed values, related to the values themselves

Issues with missing data

- It can lead to sampling bias, affecting the generalizability of results.
- MCAR and MAR are often considered less problematic and can be ignored. • MNAR can introduce bias to the dataset and cannot be ignored

How to handle missing data

- 1) Acceptance
 - a. Leave the missing data as it is, assuming they are MCAR or MAR
 - b. Make a new category in the variable called “Missing”
 - c. Recode missing values with labels like “N/A” for consistency
- 2) Deletion
 - a. Listwise deletion:
 - i. Remove all cases with missing data across any variable. ii. This might lead to a smaller, biased sample
 - b. Pairwise deletion:
 - i. Keep available data, only remove missing points from specific analysis.
 - ii. Leads to uneven sample sizes for variables but retains more data.
 - c. Imputation
 - i. Replace missing values with estimates based on:
 1. Mean or Median imputation: substitute missing values with the mean or median.
 2. Hot-deck imputation: replace missing values with similar cases within the dataset
 3. Cold-deck imputation: replace missing values with values from similar cases in another dataset (Bhandari, 2023).

If the missing values are inspected, I consider if a case has more than 30% missing data, I will remove it from my analysis. If a variable has more than 20% of missing data, I will investigate what the potential reason could be. If the measure is invalid, this measure should be removed from the analysis.

I will provide a clear overview of the descriptives, outliers and missing values and discuss and decide with my supervisor what choices I will make with regard to these values. Additionally, I will discuss with my supervisor what to do with data that is non-normal distributed. Based on this decision I will

prepare the “cleaning dataset” for further analysis. I will make a copy of this final dataset and call it “working dataset”. This will become the dataset I will continue my analysis with.

Once a decision is made with regard to outliers, missing values and the use of variables I will make a flowchart that includes the missing cases and the amount of excluded individuals etc.

Create and transform variables

- If the data is non-normally distributed, transformation into a new variable might be considered to make them more normal. I will report if I made any transformations.
- I must ensure every variable is correctly identified (continuous or categorical).
- If I need to create new variables, I save the syntaxes in a separate folder and report what I have done to make new variables.

Make descriptives of relevant variables

- Once the data is clean and all the variables are transformed/created and correctly identified, descriptives can be made again and these ones might be included in the research paper itself.
- Make a table with all the relevant variables between volunteers and non-volunteers and it might be interesting to include p-values between the two groups, to indicate any significant differences or not.
 - a. Continuous variables and normality**
 - i. If continuous variables are normally distributed One-Way ANOVA will be conducted to compare the three groups of volunteers.
 - ii. If continuous variables are non-normally distributed non-parametric tests like the the Kruskal-Wallis test will be used.
 - b. Categorical Variables and Chi-squared Test**
 - i. The Chi-squared test is suitable for analyzing categorical variables to determine whether there is a significant association between the variable and the groups (volunteers vs. non-volunteers).

Conduct bivariate analysis between all variables (optional)

- This helps to investigate the nature and strength of the association between the independent variable (participation in voluntary work) and dependent variable (cognitive decline).
- Bivariate analysis helps checking some assumptions required for linear regression, such as linearity.
 - a. Scatterplots and correlation coefficients can reveal whether there is a linear relationship (this is necessary for the independent variable (participation in voluntary work) and the dependent variable (cognitive decline)).
- Also confounders can be added as independent variables for the bivariate analysis to investigate whether specific confounders highly correlate with each other.
 - a. First thing you need to know is to see how the variables correlate with each other.
 - b. Analyse - correlate - bivariate - add variables to the box
 - c. most of the variables continuous - pearson
 - d. Outcome is a table where you can see which variables correlate
- You do not want to include variables that highly correlate, you want to exclude one of them
 - a. This can be an indication of multicollinearity, which can affect the interpretation of regression coefficients
 - b. Visualization can be done using scatterplots

Assumption of linear regression

- 1) Ensure there are no missing values in the variables you want to include
- 2) **Linearity:** There should be a linear relationship between the independent variable(s) and dependent variable. This can be done using scatterplots.

- 3) **Independence of residuals/observations:** The residuals (the differences between observed and predicted values) should be independent of each other. This assumption is often checked using residual plots or Durbin-Watson statistics.
- 4) **Homoscedasticity:** The variance of residuals should be constant across all levels of independent variables
- 5) **Normality of residuals:** The residuals should approximately be normally distributed. This can be checked using histograms, normal probability plots, or statistical tests
- 6) **No perfect multicollinearity:** If you have multiple independent variables, there should be no perfect linear relationship among them. High multicollinearity can destabilize the estimation of coefficients.
- 7) **No autocorrelation:** Residuals should not show a pattern indicating systematic errors over time or across observations. The Durbin-Watson statistic can be used to test for autocorrelation.

Ensure linearity between independent and dependent variable

- Examine e.g. scatterplots to assess the linearity between independent variable (participation in voluntary work) and the dependent variable (cognitive decline)
 - a. Graphs - chart builder - scatter/dot - drag simple scatter into the canvas - move cognitive decline variable to x axis and participation in voluntary work variable to y axis - double click to open chart editor - add regression line by clicking on “add fit line at total” - scatter plot complete.
 - b. When there is only one predictor, you do not need to check collinearity. So in my first analysis investigating the relationship between participation in voluntary work and cognitive decline, this is not necessary
- Establish homoscedasticity (variance of the residuals is constant across all levels of independent variable(s)) and linearity

Perform univariable linear regression analysis

- Analyse - regression - linear - move participation of voluntary work variable into independent window and cognitive decline variable to dependent window - method: enter - statistics: confidence intervals, descriptives, Durbin Watson, case wise diagnostics - continue - plots - move zpred to x and move zresidual to y - click on histogram and normal probability plot
 - a. In residual statistics box: min and max values for stdresidual should not exceed -3.29 or +3.29, if they do you have outliers. *(I will inspect the context of the outlier and based on what kind of outlier it might be I will make a decision whether to remove it or not).*
 - b. Independence of observations: Durbin Watson test should not be less than 1 or greater than 3 - assumptions of independence of observations have been met if it's between 1-3
 - c. Normality: pp plot, if dots must line up along a 45 degree line
 - d. Distribution: there has to be a normal distribution of the dependent variable (cognitive decline)
 - e. Scatter plot of standardised residuals vs predicted values must be elliptical = no pattern
- Repeat this for the other potential covariates/confounders and the decline scores

All of the assumption must met - However, as mentioned before the data might not be normally distributed. In consultation with my supervisor I just run the analysis first and see what the outcome would be. After that I will discuss whether I need to make adjustments with the data or analysis.

Interpret the results

- Correlation = the standardised covariance of two variables.
- Correlation matrix
 - a. Must show that the variables correlate positively or negatively.
- Variables entered/removed box
 - a. Must tell that the only predictor is participation in voluntary work and the only dependent variable is cognitive decline.
- Model summary
 - a. R = correlation
 - b. R square = proportion of variance in cognitive decline accounted for by participation in voluntary work
 - i. If it is for example $0.423 = 42.3\%$ of the variance in cognitive decline was predicted from participation in voluntary work
 - ii. A higher R -squared value suggests a larger effect size
- ANOVA summary table
 - a. The significance value (if significance is met), means that the model that uses participation of voluntary work as a predictor, was significantly better than prediction without participation in voluntary work in the model. There is a statistically significant relationship between my predictor and the outcome variable
 - b. If the ANOVA is not significant, the predictors I choose do not contribute to predicting the outcome any better than using the mean.
 - o *The ANOVA is proof that the model works or not*
- Coefficients
 - a. Standardised and unstandardised coefficients
 - i. Unstandardized coefficients are in the original scale of the variables. They represent the change in the dependent variable (cognitive decline) for one unit change in the independent variable (participation in voluntary work), holding all other variables constant. The units of measurement for the coefficients are in the units of the dependent variable per unit change in the independent variable.
 - ii. Standardized coefficients are in standard deviation units. The coefficients represent the change in the dependent variable in terms of standard deviations for a one-standard-deviation change in the independent variable. o *As I find it easier to interpret it using the coefficients in the units of the dependent and independent variables I will use the unstandardized coefficients.*
 - b. Find A and B for the regression equation
 - i. There is a column for B but not for A
 - ii. Both A and B are located in the column B
 - iii. A is a constant - starting point
 - iv. The B value has a t value associated with it - this is a T-test that measures when adding that variable as a predictor, improves the predictive ability of the model. o *If a T-test for the B-coefficient is not statistically significant, then this predictor does not add to the model*
 - o *If the T-test is significant, then look at the coefficient*
 - c. If the coefficient is negative / positive, it tells whether the dependent variable will increase or decrease due to an increase of the predictor.
 - i. This is the slope value for each predictor
 - d. Make the equation

Asses multicollinearity

Aside from bivariate analysis, the multicollinearity can be investigated when running a multivariable linear regression. The VIF values give an indication whether variables are correlated with each other. This is what you need to include in the linear regression analysis to investigate multicollinearity:

- Dependent variable: cognitive decline
- Independent = the independent variable + all the potential confounders and covariates I want to include
- Statistics: box pops up - click on collinearity diagnostics
- Look at the box “Coefficients”, there is the collinearity statistics “Tolerance and VIF”
 - a. Pay attention to all the VIF values that are ABOVE 10
- Run linear regression again without some variables that have VIF above 10
 - a. If the values are above 2, you might also have a problem
- Run linear regression one more time to exclude the variables above 2
 - a. All VIF values must all be below 2 - this means they are probably not related to each other and could be useful predictors in the model.

Multivariable linear regression

- Combine the dependent variable and the selected independent variables (potential confounders and covariates) into the analysis.

Before adding all potential confounders and covariates at once, I will first add them one at a time to understand the effect of each confounder/covariate. Eventually I will add them all together as independent variables.

- Move all the selected variables into the model.
- Check the assumptions.
- Interpret the results: evaluate model fit
 - a. Examine the correlation matrix for positive / negative correlations.
 - b. Asses the model summary (R and R square)
 - c. Examine the ANOVA summary table for significance
- Interpret the results: evaluate model fit
 - a. Analyse the unstandardised coefficients
 - b. Identify the A (constant) and B(slope) values for the regression equation.
- Diagnostic checks: review residual statistics
 - a. Check min and max values for standardized residuals
 - b. Confirm independence of observations (Durbin-Watson test)
 - c. Examine normality using PP plots and histograms
 - d. Ensure the distribution of the dependent variable is normal
 - e. Check scatter plot of standardized residuals vs. predicted values for elliptical patterns
- Refine the model if needed, by removing or transforming variables
- Make an overview in a table of the associated confidence intervals 95% and the R^3 – below is an example.

Outcome = cognitive decline

Variables	Model 0 univariable	<i>p-value</i>	Model 2 multivariable	<i>p-value</i>
Participation in voluntary work dummies	X (95% CI)		X	
<i>Potential confounders/variables</i>				
Age			x	
Sex			x	
Education			x	
Income			x	
Employment status			x	
Marital status			x	
Self-reported health			x	
CVD risk factors			x	
Hearing problems			x	
Physical inactivity			x	
Obesity			x	
Depressive symptoms			x	
Social network size			x	
Loneliness			x	
Social participation			x	
Informal caregiving			x	
Taking care of grandchildren			x	
History of voluntary work			x	
Church attendance			x	
Cogstate results 2A			x	
R2				

Table 1- Example of the linear regression model

Moderation analysis

I want to investigate whether age, sex and educational attainment moderate the role between participation in voluntary work and cognitive decline.

- You first need to create a new moderator variable – an interaction variable (example with sex)
 - a. The interaction will be between the voluntary work dummies * moderating variable, which in this case is participation in voluntary work * sex
- Only if an interaction term is significant ($p < 0.05$) it will be added to the final model.

To understand the effect of the moderators it is easier to add them first separately into the model and also take into account their effects as confounders (see table below).

Variables	Model 0	Model 1	Model 2
Participation in voluntary work dummies	x	x	x
Age	x		
Age * Participation in voluntary work dummies	x		
Agegroups * participation in voluntary work dummies			
Sex		x	
Sex * Participation in voluntary work dummies		x	
Education			x
Education * participation in voluntary work dummies			x

Table 2 - Example of the moderation analysis

Eventually I can add the interaction terms in my final model, which will look like the table below.

Variables	Model 0 univariable	p-value	Model 1 multivariable	p-value	Model 2 multivariable + interaction terms	p-value
Participation in voluntary work dummies	X (95% CI)		x		x	
<i>Potential confounders/variables</i>						
Age	x		x		x	
Sex	x		x		x	
Education	x		x		x	
Income	x		x		x	
Employment status	x		x		x	
Marital status	x		x		x	
Self-reported health	x		x		x	
CVD risk factors	x		x		x	
Hearing problems	x		x		x	
Physical inactivity	x		x		x	
Obesity	x		x		x	
Depressive symptoms	x		x		x	
Social network size	x		x		x	
Loneliness	x		x		x	
Social participation	x		x		x	
Informal caregiving	x		x		x	
Taking care of grandchildren	x		x		x	
History of voluntary work	x		x		x	
Church attendance	x		x		x	
Cogstate results 2A	x		x		x	

Age * participation in voluntary work dummies					x	
Sex * participation in voluntary work dummies					x	
Education * participation in voluntary work dummies					x	
R2						

Table 3 - Example of the final linear regression model

- Add the interaction term to the final model
 - a. Analyse – regression – linear regression – go to the third window – dependent variable is cognitive decline – add moderator variable is added as independent variable to the other independent variables – run the analysis
 - b. Interpret the coefficients in the third model.
- Make the equation for moderation analysis

$$Y = B0 + B1*X1 + B2*X2 + B3*(X1*X2)$$

Cognitive decline = B0 (constant) + value of 1st independent variable * the slope + value of 2nd independent variable * slope + value of the 1st independent variable * value of 2nd independent variable * the slope of the interaction variable

- a. The constant is the value of the dependent variable when all independent variables are 0.
- This can be done for each moderator separately

Approval form

Prior to working with the data, this approval should be signed by the supervisor as well as the student. Hereby, both student and supervisor commit to the procedures described in the analysis plan.

Place and date: Groningen, 19 March 2024

Name supervisor: Nynke Smidt

Signature supervisor:



A handwritten signature in black ink, appearing to read 'N Smidt', written over a horizontal line.

Name student: Guusje van Helden

Signature student:



A handwritten signature in black ink, appearing to be a stylized 'GH', written over a horizontal line.

Appendix A: specification of certain variables

Income - Lifelines 1A

Participants were asked: “What is the net income per month?” (if you share a household, include the net income of your partner(s)). They could choose from the categories:

- I do not know this
- I prefer not to answer that
- less than € 750
- € 750 - € 1000
- € 1000 - € 1500 • € 1500 - € 2000 • € 2000 - € 2500 • € 2500 - € 3000
- € 3000 - € 3500
- more than €3500

Additionally they were asked how many people live from this income and they could choose from:

- 1 • 2 • 3 • 4 • 5
- 6
- more than 6

From these answers the net household equivalized income can be determined. This can be done choosing the midpoint of the answer, for example for the answer € 750 - € 1000 it will become € 875. This can be divided by the square root of the answered household size (OECD, 2018). So when a household consists of two people, the equivalence factor would be $\sqrt{2} = 1.4$. The net equivalized income would be $\text{€ } 875 / 1.4 = \text{€ } 625$.

Depending on the distribution of the data I will first use is as a continuous variable. I may decide to make categories based on the distribution the data or based on information given by the Central Bureau of Statistics (CBS). Below is an overview of the variables I need and how I will create a new variable.

As it is usual that income variables have a lot of missing values, so it might be that from a lot of people their equivalized income cannot be calculated. Therefore, the missing values will be imputed by the mean equivalized income. Additionally, a new variable will be created that will indicate whose equivalized income can be calculated and take that into account in my analysis.

Variable	Labelled in SPSS
----------	------------------

<p>Income (WORK4A)</p> <p>What is the net income per month?</p> <ul style="list-style-type: none"> • less than € 750 • € 750 - € 1000 • € 1000 - € 1500 • € 1500 - € 2000 • € 2000 - € 2500 • € 2500 - € 3000 • € 3000 - € 3500 • more than € 3500 • i do not know this • i prefer not to answer that 	<p><i>Income_net_adu_q_1_v2</i></p> <p>7 = i do not know this 8 = i prefer not to answer that 9 = less than € 750 10 = € 750 - € 1000 11 = € 1000 - € 1500 12 = € 1500 - € 2000 13 = € 2000 - € 2500 14 = € 2500 - € 3000 15 = € 3000 - € 3500 16 = € 3000 - € 3500 17 = more than 3500</p>
<p>Income (WORK4B)</p> <p>How many people live from this income?</p> <ul style="list-style-type: none"> • 1 • 2 • 3 • 4 • 5 • 6 • more than 6 	<p><i>income_net_adu_q_1_a</i></p> <p>1 = 1 2 = 2 3 = 3 4 = 4 5 = 5 6 = 6 7 = more than 6</p>
<p>Net household equivalized income</p>	<p>choosing the midpoint of the answer of _____</p> <p>$\frac{Income_net_adu_q_1_v2}{\sqrt{householdsize}}$</p>

Table 6 - Overview variables related to income

Cardiovascular risk factors – Lifelines 1A

The “Lifestyle for Brain Health’ (LIBRA) score was developed that assesses an individual’s potential for dementia prevention, based on their lifestyle choices. This also includes some cardiovascular related factors, such as coronary heart disease (CHD), high cholesterol and hypertension (Schiepers et al., 2018). During Lifelines 1A questions with regard to cardiovascular disease were included. CHD refers to several conditions that impact the structure and functioning of the heart (National Heart, Lung, and Blood Institute, National Institutes of Health, 2023). Many complications may be the result such as angina, heart attacks and heart failure. It can be treated using medications, but also with procedures and surgeries like coronary angioplasty, bypass surgery and a heart transplant (National Health Service, 2023).

There are several questions included in Lifelines 1A related to CHD:

Lifelines questions from 1A that are related to CHD are:

- 1) “Have you ever had a heart attack?”
- 2) “Do you have heart failure (decreased pumping of the heart)?”
- 3) “Have you ever had a stroke?”

With regard to high cholesterol they asked questions

- 1) "Have you ever been diagnosed with high cholesterol?"
- 2) "If so how old were you when high cholesterol was first diagnosed? / have you ever been diagnosed with high cholesterol?"
- 3) "How are you being treated for high cholesterol? / have you ever been diagnosed with high cholesterol?"
 - a. I am not being treated
 - b. With a diet
 - c. With medication
 - d. With both

The same applies for questions with regard to hypertension.

- 1) "Have you ever had hypertension?"
- 2) "if so, how old were you when hypertension was first diagnosed? / have you ever had hypertension?"
- 3) "How are you being treated for hypertension? / have you ever had hypertension?"
 - a. I am not being treated
 - b. With a diet
 - c. With medication
 - d. With both

With regard to the questions related to CHD, I can create a new variable that defines participants that have the presence and absence of CHD risk factors. Those who answered 'yes' on one of the related questions will be defined as 'presence of CHD risk factors' = 1. All others, including those who did not answer, will be defined as 'no presence of CHD risk factors' = 0.

As CHD, hypertension and high cholesterol levels are used as separate modifiable risk factors for the LIBRA score, therefore I will use the presence of hypertension and high cholesterol also separately for my analysis. I will only focus on the questions 'Have you ever been diagnosed with high cholesterol' and 'Have you ever had hypertension?'. As it might not be the best indicator whether participants have hypertension or high cholesterol levels at the moment of taking the questionnaire, it is the most easiest way for my analysis to make a distinction between participants.

Consequently, I will recode the variable associated with the diagnosis of high cholesterol. Participants who responded 'yes' to the question regarding the presence of high cholesterol will be categorized as '1', indicating those with the presence of high cholesterol levels. All other participants, including those who did not answer will be categorized as '0', indicating those with the absence of high cholesterol levels.

The same will be done to determine the presence of hypertension. Those who responded 'yes' to the question if they ever had hypertension will be categorized as '1', indicating those with the presence of hypertension. All the others will be categorized as '0', indicating those without the presence of hypertension.

Below is an overview of all the variables I will include in my analysis.

Question	Variable	Coded in SPSS	Labelled in SPSS
Have you ever had a heart attack?	health19	heartattack_presence_adu_q_1	1= yes, 2 = no
“Do you have heart failure?”	health32	Heartfailure_presence_adu_q_1	1 = yes 2 = no 3 = I do not know
Have you ever had a stroke?	Health42	stroke_presence_adu_q_1	1 = yes 2 = no
“Have you ever been diagnosed with high cholesterol?” • Yes • No • I do not know	Health38	Highcholesterol_presence_adu_q_1	1 = yes 2 = no 3 = I do not know
“Have you ever had hypertension?” • Yes • No • I do not know	Health35	hypertension_presence_adu_q_1	1 = yes 2 = no 3 = I do not know

Table 7- Overview variables related to cardiovascular disease risk factors

Physical inactivity - DALAS

Participants were asked two questions with regard to their physical activity:

- 1) On an average of how many days a week, in total, do you spend at least half an hour cycling, chores, gardening, sports or other strenuous activities during the summer?
- 2) On an average of how many days a week, in total, do you spend at least half an hour cycling, chores, gardening, sports or other strenuous activities during the summer?

Respondents could answer between 0-7 days. According to the National Institute of Public Health and the Environment (RIVM), adults and the elderly should exercise 150 minutes of moderate or high intensity exercise per week, spread over several days. Thus, to meet these guidelines, they must be active for half an hour on average 5 days (Ministerie van Volksgezondheid, Welzijn en Sport, 2021)

So the average of both of these questions must be at least 5. A new continuous variable can be made that indicates the average of the scores. Additionally, a new dichotomous variable based on the other one can be made to indicate whether respondents meet the criteria of these guidelines. If participants have an average ≥ 5 they will be labelled with a 0. If participants have an average < 5 they are labelled by 1. However, I will first investigate the distribution of the data and see what the sample sizes will be for each groups.

Obesity – Lifelines 2A

As obesity is one of the modifiable risk factors for dementia (Livingston et al., 2020), I want use it in my analysis as potential confounder. Individuals height and weight was measured during Lifelines 1A and 2A. Measured of height and weight during Lifelines 2A are the most recent measures and also at the moment the Cogstate was administered. Therefore, I will use the measured of height and weight of participants from 2A.

Using the values of height and weight, the body mass index (BMI) can be calculated. The weight in kilos must be multiplied by the length in meters squared.

$$\text{Weight (in kg)} * (\text{length in meters} * \text{length in meters})$$

First I need to create a new variable that calculates the BMI. As length is used in centimeters, I need to divide it by 100 to calculate the BMI. As I want obesity to use in my analysis as a potential confounder, I need to use the official cutoff values of the BMI associated with obesity. Based on the Dutch Nutrition Centre, the BMI cutoff value to indicate obesity is ≥ 30 (Voedingscentrum, n.d.). I will create a new variable that defines those who have a BMI < 30 with '0', indicating those who do not have obesity. Those who have a BMI ≥ 30 with '1', indicating those who have obesity.

As I am not sure how many individuals will have obesity, so after making my variables I will decide whether or not to include this potential confounder in my analysis.

Depression - DALAS

For the DALAS questionnaire they used the PROMIS-29 (Patient-Reported Outcomes Measurement Information System) profile. This consists of a questionnaire with 29 items, incorporating short forms of 4, 6 or 8 items for depression, anxiety, fatigue, sleep disorders, physical functioning, impairments due to pain and satisfaction with social roles and activities (Dutch Flemish Promis Group, n.d.). As I am interested in depression, I will focus on the Emotional Distress-Depression short form 4a.

PROMIS Item Bank v1.0 – Emotional Distress-Depression – Short Form 4a

Emotional Distress-Depression – Short Form 4a

Please respond to each question or statement by marking one box per row.

In the past 7 days...		Never	Rarely	Sometimes	Often	Always
EDDEP04	I felt worthless	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
EDDEP06	I felt helpless	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
EDDEP29	I felt depressed	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
EDDEP41	I felt hopeless	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

Figure 3 - Emotional Distress-Depression Short form 4a and the assigned values for each question (Health Measures, 2016)

To score a PROMIS instrument, each question, typically with five response options, is assigned a numeric value. The total raw score is the sum of these values, producing a score between the minimum and maximum possible scores for that form. All questions must be answered for a valid score, and if a question is skipped, the Health Measures Scoring Service is recommended. The total raw score is then translated into a T-score using a conversion table, which can be seen in table 4. The T-score standardized the raw score with a mean of 50 and a standard deviation of 10. A T-score of 50 is the average for the U.S. general population, and higher T-scores indicate more of the measured concept. For negatively-worded concepts like depression, a T-score of 60 is one standard deviation worse than average, while a T-score of 40 is one standard deviation better. The T-score comes with a standard error (SE), which measures variability and is used to construct confidence intervals. A 95% confidence interval indicates the range within which the true T-score is likely to fall.

It's essential for all questions to be answered to generate a valid score. In case a question is skipped, the Health Measures Scoring Service can be used to compute the final score. However, a scoring service login is needed to do this and I do not have access to that. It is easier to just focus on the cases who will have full scores (Health Measures, 2021).

Depression 4a - Adult v1.0		
Short Form Conversion Table		
Raw Score	T-Score	SE*
4	41.0	6.2
5	49.0	3.2
6	51.8	2.7
7	53.9	2.4
8	55.7	2.3
9	57.3	2.3
10	58.9	2.3
11	60.5	2.3
12	62.2	2.3
13	63.9	2.3
14	65.7	2.3
15	67.5	2.3
16	69.4	2.3
17	71.2	2.4
18	73.3	2.4
19	75.7	2.6
20	79.4	2.6

SE* = Standard Error on T-Score d

Table 8 – Short form conversion table (Health Measures, 2021)

The four questions are coded as four different variables. In table 5 you see how they are named in SPSS.

Question	SPSS
Ik voelde mij alsof ik niets waard was	promis_depression_adu_q_09
Ik voelde me hulpeloos	promis_depression_adu_q_10
Ik voelde me depressief	promis_depression_adu_q_11
Ik voelde me zonder hoop	promis_depression_adu_q_12

Table 9 - Questions and associated variable names

I can recompute these variables to new ones where I will first assign the values never = 1, rarely = 2, sometimes = 3, often = 4 and always = 5. A new variable can be made that consists the sum of all of these values for each participant. Additionally, variables can be made that will contain the corresponding T-scores and standard error of the T-scores, based on the conversion table (table 4). The reference population is based on the US population, however they have also developed the mean scores for the Dutch population based on gender and age. The cut-off points for mild, moderate and severe scores for PROMOS Anxiety and Depression are the same as those established in America (Dutch-Flemish PROMIS National Center, 2023):

- < 55 = within normal limits
- 55 - 60 = mild
- 60 – 70 = moderate
- > 70 = severe

I will first investigate the distribution of the T-scores before making any categories. Additionally, it is easier to use dichotomous variables so I may make two categories where < 55 = no depressive symptoms and will be assigned to 0. A score ≥ 55 = presence of depressive symptoms and will be assigned to 1. However, I will first have to look into the distribution of the T-scores.

Loneliness - DALAS

For the presence of loneliness the De Jong Gierveld loneliness short scale is used. The original scale consists of eleven items, however for the DALAS questionnaire the short scale has been used with six items. The scale consists of six items; three are formulated negatively and the other three are formulated positively. Below there is an overview of the questions that have been included in the DALAS questionnaire.

The first three questions are negatively formulated. The neutral/positive answers must be counted. The answer ‘more or less’ can be used as an indicator of loneliness instead of a neutral answer. So for the questions 1-3 On items 1-3 “more or less”, “yes”, “yes!” can be counted and represents the emotional loneliness score. Count the missing values on these items. These are the missing emotional loneliness scores. On the items 4-6 count the negative answers (“No!”, “No”, “More or less”) can be counted (table 6). This represents the social loneliness score, the missing values are the missing social loneliness score (Van Tilburg, 2023).

The loneliness score will be the sum of the emotional loneliness score and the social loneliness score. Only if the missing emotional loneliness score equals 0, the score is valid. If the sum of the missing social loneliness score equals 0 or 1, the total loneliness score is valid. If a respondent has more missing values, it must be deleted from the analysis (Van Tilburg, 2023).

Using this procedure a respondent can score a total of 6 loneliness points in case of extreme loneliness. The minimum score is 0. If a respondent has scored one and only one missing value, the response is not considered to be a loneliness indicator. No scale point is given for the item (Van Tilburg, 2023).

A cut-off point for the six-item scale in the updated version is: 0-1 = not lonely; 2-4 = moderately lonely; 5-6 = strongly lonely. As mentioned, there are two short scales for emotional and social loneliness, with each having three items. However, the scale for emotional loneliness is more closely related to the direct question of loneliness than the scale for social loneliness. Therefore, cut-off point for both subscales does not result in a similar score on the shortened scale. Application of the prevalence criterion gives a dichotomy of 0 = not lonely and 1-3 = lonely for both short scales for emotional and social loneliness (Van Tilburg, 2023). However, in the previous version of this manual they use a combination of these short subscales and assign a dichotomy of 0-1 not lonely and 2-6 lonely, which can be seen in the table below (De Jong Gierveld & Van Tilburg, 2022). As I just want to use an indicator of loneliness as a potential confounder, it is easier to combine the emotional and loneliness scale. Therefore I will use the cut-off points indicated in the 2022 manual. I will create a new variable were the scores 0-1 will be assigned to 0 to represent “not lonely” and scores 2-6 will be assigned to 1 to represent “lonely”.

	Ja!	Ja	Min of meer	Nee	Nee!	<i>Variable code in SPSS</i>
1. Ik ervaar leegte om mij heen	1	1	1	0	0	loneliness_emptiness_adu_q_01
2. Ik mis mensen om mij heen	1	1	1	0	0	loneliness_people_adu_q_02

3. Vaak voel ik me in de steek gelaten	1	1	1	0	0	loneliness_abandoned_adu_q_03
4. Er zijn genoeg mensen op wie ik in geval van narigheid kan terugvallen	0	0	1	1	1	loneliness_fallback_adu_q_04
5. Ik heb veel mensen op wie ik volledig kan vertrouwen	0	0	1	1	1	loneliness_trust_adu_q_05
6. Er zijn voldoende mensen met wie ik me nauw verbonden voel	0	0	1	1	1	loneliness_connection_adu_q_06

Table 10 - Overview of counting the answers of the negatively and positively formulated questions of the De Jong Gierveld loneliness scale and their associated variable name in SPSS (Van Tilburg, 2023)

Scale	# items	Range	Cut-off
De Jong Gierveld loneliness scale	11	0-11	0-2 not lonely 3-8 moderately lonely 9-11 strongly lonely
Subscale for emotional loneliness	6	0-6	0-2 not emotionally lonely 3-6 emotionally lonely
Subscale for social loneliness	5	0-5	0-2 not socially lonely 3-5 socially lonely
Short scale (6 items)	6	0-6	0-1 not lonely 2-6 lonely
Short scale for emotional loneliness	3	0-3	0 not emotionally lonely 1-3 emotionally lonely
Short scale for social loneliness	3	0-3	0 not socially lonely 1-3 socially lonely

Table 11 - Summary table of the different scales and cut-off scores from the 2022 manual (De Jong Gierveld & Van Tilburg, 2022)

Social participation – DALAS (optional)

As it is known that being socially isolated is a risk factor for dementia (Livingston et al., 2020) and social network size a determinant is of participating in voluntary work (Niebuur et al., 2018), looking at someone's ability to participate in social roles and activities might be an interesting confounder. For the DALAS questionnaire they used the PROMIS-29 (Patient-Reported Outcomes Measurement Information System) profile. This consists of a questionnaire with 29 items, incorporating short forms of 4, 6 or 8 items for depression, anxiety, fatigue, sleep disorders, physical functioning, impairments due to pain and satisfaction with social roles and activities (Dutch Flemish Promis Group, n.d.).

I am interested in social participation, therefore I will focus on the "Ability to participate in social roles and activities" short form 4a. As mentioned in the PROMIS depression section, each question, typically with five response options, is assigned a numeric value (figure 4).

This score is then converted into a T-score with a mean of 50 and a standard deviation of 10, standardizing it against the U.S. general population. A higher PROMIS T-score represents more of the concept being measured. For positively-worded concepts like Ability to Participate in Social Roles and Activities, a T-score of 60 is one SD better than average. By comparison, an Ability to Participate in Social Roles and Activities T-score of 40 is one SD worse than average (Health Measures, 2018b).

Ability to Participate in Social Roles and Activities - Short Form 4a

Please respond to each item by marking one box per row.

		Never	Rarely	Sometimes	Usually	Always
SRPPER11_ CaPS	I have trouble doing all of my regular leisure activities with others	<input type="checkbox"/> 5	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1
SRPPER18_ CaPS	I have trouble doing all of the family activities that I want to do.....	<input type="checkbox"/> 5	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1
SRPPER23_ CaPS	I have trouble doing all of my usual work (include work at home)	<input type="checkbox"/> 5	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1
SRPPER46_ CaPS	I have trouble doing all of the activities with friends that I want to do	<input type="checkbox"/> 5	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1

Figure 4 - Questions used in the short form of the Ability to participate in social roles and activities item and the assigned values (Health Measures, 2018a)

This score is then converted into a T-score with a mean of 50 and a standard deviation of 10, standardizing it against the U.S. general population. Higher T-scores indicate greater social participation. A T-score of 60 signifies one standard deviation worse than average, while a T-score of 40 is one standard deviation better. The T-score comes with a standard error (SE) for constructing confidence intervals to indicate the range where the true T-score is likely to fall.

It's essential for all questions to be answered to generate a valid score. In case a question is skipped, the Health Measures Scoring Service can be used to compute the final score. However, a scoring service login is needed to do this and I do not have access to that. It is easier to just focus on the cases who will have full scores (Health Measures, 2018b).

Raw Score	Scale Score	SE [†]
4	27.5	4.1
5	31.8	2.5
6	34.0	2.3
7	35.7	2.2
8	37.3	2.1
9	38.8	2.2
10	40.5	2.3
11	42.3	2.3
12	44.2	2.3
13	46.2	2.3
14	48.1	2.2
15	50.0	2.2
16	51.9	2.2
17	53.7	2.3
18	55.8	2.3
19	58.3	2.7
20	64.2	5.1

[†]SE = Standard Error on T-score metric

Table 12 - Short form ability to participate social conversion table T-scores (Health Measures, 2018b)

The four questions are coded as four different variables. In table 5 you see how they are named in SPSS.

Question	SPSS
I have difficulty engaging in all my normal leisure pursuits with others	promis_social_adu_q_21
I have difficulty doing all the family activities that I want to	promis_social_adu_q_22
I have difficulty doing all my normal work (including work at home)	promis_social_adu_q_23
I have difficulty doing all my normal work (including work at home) / ability to participate in social roles and activities	promis_social_adu_q_24

Table 13 - Overview questions included and their assigned variable names

I can recompute these variables to new ones where I will first assign the values never = 1, rarely = 2, sometimes = 3, often = 4 and always = 5. A new variable can be made that consists the sum of all of these values for each participant. Additionally, variables can be made that will contain the corresponding T-scores and standard error of the T-scores, based on the conversion table (table 8).

The reference population is based on the US population, however they have also developed the mean scores for the Dutch population based on gender and age (Dutch-Flemish PROMIS National Center, 2023). It is recommended to use the distribution-based thresholds, the cut-off values for the Dutch populations will be:

- > 46 = within normal limits
- 41 – 46 = mild
- 32 – 41 = moderate
- < 32 = severe

I will first investigate the distribution of the T-scores before making any categories. Additionally, it is easier to use dichotomous variables so I might recode it into two categories. Individuals with a score > 46 are considered to be well able to participate socially and will be assigned to 0. Those with a score ≤ 46 are considered to be less able to participated socially and will be assigned to 1. However, I will first have to look into the distribution of the T-scores.

Appendix B: Overview of the variables

In binary outcome variables related to health, the reference category, usually coded as 0, indicates the healthiest group.

**Variables have been further explained in the exposure section of the analysis plan.*

***Variables have been further explained in the measure section of the analysis plan.*

****Variables have been further explained in appendix A.*

	Variable	Measurement level	Coding lifelines 1A / DALAS	Coding SPSS	Unit
1	“Have you ever participated in voluntary work?” (DALAS)	Categorical / dichotomous	UVVOLUNT_VOLUNT21 1 = yes 2 = no	<i>volunteering_lifetime_adu_q_1</i>	
2	“Are you currently participating in voluntary work?” (DALAS)	Categorical / dichotomous	UVVOLUNT_VOLUNT24 1 = yes 2 = no	<i>volunteering_now_adu_q_1</i>	

3	“Do you volunteer in an organised setting (such as a charity, sports club, senior citizens' organisation, church or musical society)?”(DALAS)	Categorical / nominal	UVVOLUNT_VOLUNT25 1 = yes, exclusively formal volunteering 2 = yes, combination of formal and informal volunteering 3 = no, only informal volunteering	<i>volunteernow_organization_adu_q_1</i>	
4	Participation in voluntary work* (recode into new variable)	Categorical / dichotomous	IF (volunteering_lifetime_adu_q_1 = "yes" AND volunteering_now_adu_q_1 = "yes") volunteer_status = 1. IF (volunteering_lifetime_adu_q_1 = "no" OR volunteering_now_adu_q_1 = "no") volunteer_status = 0.	<i>Volunteer_status</i> Non-volunteers → 0 Volunteers → 1	
5	“On average, how many hours did you spend each time on this volunteer work (1-5)?” (DALAS) (almost) daily: [xx] hours	Categorical / nominal	UVVOLUNT_VOLUNT33A1 - daily xx hours	<i>volunteernow_duration_adu_q_01_01t/m</i> <i>volunteernow_duration_adu_q_01_05</i>	
6	2 to 3 times per week: [xx] hours	Categorical / nominal	UVVOLUNT_VOLUNT33A2 - 2-3x a week xx hours	<i>Volunteernow_duration_adu_q_02_01 t/m</i> <i>volunteernow_duration_adu_q_02_05</i>	

7	1 time per week: [xx] hours	Categorical / nominal	UVVOLUNT_VOLUNT33A3 - 1x a week xx hours	<i>Volunteernow_duration_adu_q_0</i> 3_01 t/m <i>volunteernow_duration_adu_q_0</i> 3_05	
8	1 time per two weeks: [xx] hours	Categorical / nominal	UVVOLUNT_VOLUNT33A4 - 1x per 2 weeks xx hours	<i>Volunteernow_duration_adu_q_0</i> 4_01 t/m <i>volunteernow_duration_adu_q_0</i> 4_05	
9	1 time per month: [xx] hours	Categorical / nominal	UVVOLUNT_VOLUNT33A5 - 1x per month xx hours	<i>Volunteernow_duration_adu_q_0</i> 5_01 t/m <i>volunteernow_duration_adu_q_0</i> 5_05	
10	1 time per two months: [xx] hours	Categorical / nominal	UVVOLUNT_VOLUNT33A6 - 1x in 2 months xx hours	<i>Volunteernow_duration_adu_q_0</i> 6_01 t/m <i>volunteernow_duration_adu_q_0</i> 6_05	
11	1 time per three months: [xx] hours	Categorical / nominal	UVVOLUNT_VOLUNT33A7 - 1x in 3 months xx hours	<i>Volunteernow_duration_adu_q_0</i> 7_01 t/m <i>volunteernow_duration_adu_q_0</i> 7_05	
12	1 time per half year: [xx] hours	Categorical / nominal	UVVOLUNT_VOLUNT33A8 - Once every 6 months xx hours	<i>volunteernow_duration_adu_q_0</i> 8_01 t/m	
				<i>volunteernow_duration_adu_q_0</i> 8_05	

13	1 time per nine months: [xx] hours	Categorical / nominal	UVVOLUNT_VOLUNT33A9 - Once every 9 months xx hours	<i>Volunteernow_duration_adu_q_0 9_01 t/m volunteernow_duration_adu_q_0 9_05</i>	
14	1 time per year: [xx] hours	10_01 t/m 10_01 are scale (continuous) variables 10_04 and 10_05 are nominal (categorical) variables.	UVVOLUNT_VOLUNT33A7 - 1x in 3 months xx hours	<i>Volunteernow_duration_adu_q_1 0_01 t/m volunteernow_duration_adu_q_1 0_05</i>	
15	Total hours per week participating in voluntary work * (recode into new variable)	Continuous	<i>This analysis is optional and the conversion table is on page 4 to see how I will transform the hours to their belonging frequencies in total hours per week.</i>	<i>Total_Hours_Volunteering</i>	
16	Levels of participation in voluntary work* (recode into new variable if needed)	Categorical / Ordinal	<i>This analysis is optional and based on Total_Hours_Volunteering I possibly create categories. Depending on the distribution of the data.</i>	<i>Low_Participation Moderate_Participation High_Participation</i>	
17	Cogstate 2A and 3A (Lifelines 2A and 3A) - IDN	Continuous	LMNUNTRANS	<i>cogstate_speed_all_m_1_01_2a / cogstate_speed_all_m_1_01_3a</i>	milliseconds

				Label: speed measure - mean of reaction time for correct responses / test 1 (IDN)	
18	Cogstate 2A and 3A (Lifelines 2A and 3A) - DET	Continuous	LMNUNTRANS	<i>cogstate_speed_all_m_1_02_2a / cogstate_speed_all_m_1_02_3a</i> Label: speed measure - mean of reaction time for correct responses / test 2 (DET)	milliseconds
19	Cogstate 2A and 3A (lifelines 2A and 3A) - OCL	Continuous	ACCUNTRANS The primary outcome measure was the proportion of correct answers, which was normalised using arcsine transformation	<i>cogstate_hitrate_all_m_1_03_2a / cogstate_hitrate_all_m_1_03_3a</i> Label: accuracy measure - hit rate (correct responses divided by total responses (total includes tpos, tneg, fpos, fneg, anticipations and max outs) / test 3 (ocl)	Hitrate
20	Cogstate 2A and 3A (lifelines 2A and 3A) - OBK / ONB	Continuous	ACCUNTRANS	<i>cogstate_hitrate_all_m_1_04_2a / cogstate_hitrate_all_m_1_04_3a</i> Label: accuracy measure - hit rate (correct responses divided by total responses (total includes tpos, tneg, fpos, fneg,	hitrate

				anticipations and max outs) / test 4 (onb)	
21	Cognitive decline IDN ** (recode into new variable)	Continuous	COMPUTE IDNdecline = cogstate_speed_all_m_1_01_3a - cogstate_speed_all_m_1_01_2a.	<i>IDNdecline</i>	Milliseconds
22	Cognitive decline DET ** (recode into new variable)	Continuous	COMPUTE DETdecline = cogstate_speed_all_m_1_02_3a - cogstate_speed_all_m_1_02_2a	<i>DETdecline</i>	Milliseconds
23	Cognitive decline OCL ** (recode into new variable)	Continuous	COMPUTE OCLdecline = cogstate_hitrate_all_m_1_03_2a cogstate_hitrate_all_m_1_03_3a	<i>OCLdecline</i>	Hitrate
24	Cognitive decline ONB ** (recode into new variable)	Continuous	COMPUTE ONBdecline = cogstate_hitrate_all_m_1_04_2a cogstate_hitrate_all_m_1_04_3a	<i>ONBdecline</i>	Hitrate
25	Cogstate 2A (Lifelines 2A) - IDN	Continuous	LMNTRANS	<i>cogstate_speedtransf_all_m_1_01</i> speed measure - mean of the log10 transformed reaction time for correct responses / test 1 (idn)	Milliseconds

26	Cogstate 2A (Lifelines 2A) - DET	Continuous	LMNTRANS	<i>cogstate_speedtransf_all_m_1_02</i> speed measure - mean of the log10 transformed reaction time for correct responses / test 2 (det)	Milliseconds
27	Cogstate 2A (Lifelines 2A) - OCL	Continuous	ACCTRANS	<i>cogstate_hitratetransf_all_m_1_03</i> accuracy measure - arcsine square root transformed - hitrate (correct responses divided by total responses (total includes tpos, tneg, fpos, fneg, anticipations and max outs) / test 3 (ocl)	hitrate
28	Cogstate 2A (Lifelines 2A) - ONB	Continuous	ACCTRANS	<i>cogstate_hitratetransf_all_m_1_04</i> accuracy measure - arcsine square root transformed - hitrate (correct responses divided by total responses (total includes tpos, tneg, fpos, fneg, anticipations and max outs) / test 4 (onb)	hitrate
29	Age (DALAS)	Continuous	-	Age	Years

30	Age groups (Recode into new variable)	Categorical / dichotomous	IF (age >= 60 AND age <= 69) Age_Groups = 0. IF (age > 69 AND age <= 80) Age_Groups= 1.	<i>Age_Groups</i> 60 – 69 → 0 70 – 80 → 1	
31	Educational attainment (LifeLines 1A)	Categorical / nominal	<i>EDUCATION Degree_highest_adu_q_1</i> 1 = no education (did not finish primary school) 2 = primary education (primary school, special needs primary school)	<i>educational_attainment_adu_c_1</i> 1-4 = low → 1	

			<p>3 = lower or preparatory secondary vocational education (such as lts, leao, lhno, vmbo)</p> <p>4 = junior general secondary education (such as mavo, (m)ulo, mbo-short, vmbo-t)</p> <p>5 = secondary vocational education or work-based learning pathway (such as mbo-long, mts, meao, bol, bbl, inas)</p> <p>6 = senior general secondary education, pre-university secondary education (such as havo, vwo, atheneum, gymnasium, hbs, mms)</p> <p>7 = higher vocational education (such as hbo, hts, heao, 'kandidaats' university education (e.g. bachelor)</p> <p>8 = university education</p> <p>9 = other</p> <p>Participants who chose answer 9 ('other', n = ~4,300) were scored according to the specifications they provided in degree_highest_adu_q_1_a (EDUCOTHER). These specifications were manually checked for certain educational terms. Based on this they also got assigned to the low, middle or high group, or to no group. These next steps can be checked on https://wiki-lifelines.web.rug.nl/doku.php?id=educational_attainment&s[]=education</p>	<p>5-6 = middle → 2</p> <p>7-8 = high → 3</p>	
32	Gender (DALAS)	Categorical / dichotomous	<p><i>UVALGEMEEN_GENDER</i></p> <p>1 = male</p> <p>2 = female</p>	GENDER	
33	Gender (recode into new variable)	Categorical / dichotomous	<p>RECODE gender ('MALE'= 0) ('FEMALE'= 1) INTO Sex</p>	<p><i>Sex</i></p> <p>Male → 0</p> <p>Female → 1</p>	

34	Employment status (DALAS) “Have you engaged in the following activities in the past 12 months?”	Categorical / dichotomous	<i>UVTIJDBES_ACTIV1</i> – Paid Work 1= yes 2 = no	<i>activities_paidwork_adu_q_1</i>	
35	Employment status (recode into new variable)	Categorical/ dichotomous	RECODE activities_paidwork_adu_q_1 ('yes' = 1) ('no' = 0) (ELSE=SYSMIS) INTO Paid_Work.	<i>Paid_Work</i> Not employed → 0 employed → 1	
36	Marital Status (DALAS) ‘Do you have a partner now?’	Categorical	<i>UVDEMOG_DEMO7C</i> 1 = Yes, I am married. 2 = Yes, I live together/registered partnership 3 = Yes, I am in a LAT relationship/partner 4 = No, I am single/no partner 5 = other:	<i>partner_presence_adu_q_1</i>	
37	Living arrangement (recode marital status into new variable based on living arrangement)	Categorical	RECODE partner_presence_adu_q_1("Yes, I am married" = 1 "Yes, I live together/registered partnership" = 1 "Yes, I am in a LAT relationship/partner" = 0 "No, I am single/no partner" = 0 "other" = \$SYSMIS) INTO Living_Arrangement	<i>Living_Arrangement</i> 1 → 1 = cohabiting 2 → 1 = cohabiting 3 → 0 = not cohabiting 4 → 0 = not cohabiting	

38	Income (Lifelines 1A) “What is the net income per month?”	Categorical	<i>WORK4A</i> * less than € 750 * € 750 - € 1000 * € 1000 - € 1500 * € 1500 - € 2000 * € 2000 - € 2500 * € 2500 - € 3000 * € 3000 - € 3500 * more than € 3500 * i do not know this * i prefer not to answer that	<i>Income_net_adu_q_1_v2</i> 7 = i do not know this 8 = i prefer not to answer that 9 = less than € 750 10 = € 750 - € 1000 11 = € 1000 - € 1500 12 = € 1500 - € 2000 13 = € 2000 - € 2500 14 = € 2500 - € 3000 15 = € 3000 - € 3500 16 = € 3000 - € 3500 17 = more than € 3500	
39	Income (Lifelines 1A) “How many people live from this income?”	Categorical	Income (WORK4B) 1 2 3 4 5 6 more than 6	<i>income_net_adu_q_1_a</i> 1 = 1 2 = 2 3 = 3 4 = 4 5 = 5 6 = 6 7 = more than 6	
40	Midpoint income*** (Recode into new variable)	Continuous	* Calculate the midpoint of the answers of <i>Income_net_adu_q_1_v2</i> DO IF (Income_net_adu_q_1_v2 = 9). COMPUTE midpoint_income = 750. DO IF (Income_net_adu_q_1_v2 = 10). COMPUTE midpoint_income = (750 + 1000) / 2.	<i>Midpoint_Income</i>	

			<pre> ELSE IF (Income_net_adu_q_1_v2 = 11). COMPUTE midpoint_income = (1000 + 1500) / 2. ELSE IF (Income_net_adu_q_1_v2 = 12). COMPUTE midpoint_income = (1500 + 2000) / 2. ELSE IF (Income_net_adu_q_1_v2 = 13). COMPUTE midpoint_income = (2000 + 2500) / 2. ELSE IF (Income_net_adu_q_1_v2 = 14). COMPUTE midpoint_income = (2500 + 3000) / 2. ELSE IF (Income_net_adu_q_1_v2 = 15). COMPUTE midpoint_income = (3000 + 3500) / 2. ELSE IF (Income_net_adu_q_1_v2 = 16). COMPUTE midpoint_income = (3000 + 3500) / 2. ELSE IF (Income_net_adu_q_1_v2 = 17). COMPUTE midpoint_income = 3500. </pre>		
41	Equivalised income*** (recode into new variable)	Continuous	<pre> * Calculate equivalised income. COMPUTE equivalised_income = midpoint_income / SQRT(income_net_adu_q_1_a). EXECUTE. </pre>	<i>Equivalised_Income</i>	

42	Available equivalised income (recode into new variable)	Categorical / dichotomous	<p>From a lot of participants the equivalised income cannot be calculated, I will create a new variable that indicates the ones from who their equivalised income can be calculated and the ones who could not be calculated.</p> <p>*Define those who have missings or did not answer their income on Income_net_adu_q_1_v2</p> <p>RECODE Income_net_adu_q_1_v2 (7=0) (8=0) (Missing =0) (ELSE =1) INTO income_md</p>	<p><i>Available_equi_income</i></p> <p>0 → those without an equivalised income</p> <p>1 → those with an equivalised income</p>	
----	--	------------------------------	--	--	--

			<p>*Define those who have missings on household size RECODE income_net_adu_q_1_a (MISSING=0) (ELSE=1) INTO household_md.</p> <p>*Those who have on income_md a zero or on household_md a zero, their equivalised income cannot be calculated</p> <p>IF (income_md = 0 OR household_md = 0) available_equi_income = 0.</p> <p>ELSE IF (income_md = 1 AND household_md = 1) available_equi_income = 1.</p>		
--	--	--	---	--	--

43	Self-reported health (DALAS) “How would you rate your health, generally speaking?”	Categorical / nominal	<i>UVRAND36_RAND1</i> 1 = Excellent 2 = Very Good 3 = Good 4 = Moderate 5 = Poor	<i>rand_generalhealth_adu_q_01</i>	
44	Self-reported health (recode into new variable if needed)	Categorical / Ordinal	*Recode the variable in the right ordinal order and make mediocre and poor one category. RECODE rand_generalhealth_adu_q_01 (‘Excellent’ = 1) (‘Very good’ = 2) (‘Good’ = 3) (‘Mediocre’, ‘Poor’ = 4) (ELSE = COPY) INTO Self_Reported_Health	<i>Self_Reported_Health</i> 1 → excellent 2 → very good 3 → good 4 → mediocre / poor	
45	Cardiovascular disease (Lifelines 1A)	Categorical / dichotomous	health19 1= yes	<i>heartattack_presence_adu_q_1</i>	

	“Have you ever had a heart attack?”		2 = no		
46	Cardiovascular disease (Lifelines 1A) “Do you have heart failure (decreased pumping of the heart)?”	Categorical / dichotomous	health32 1= yes 2 = no 3 = I do not know	<i>Heartfailure_presence_adu_q_1</i>	

47	Cardiovascular disease (Lifelines 1A) “ Have you ever had a stroke?”	Categorical / dichotomous	Health42 1= yes 2 = no	<i>stroke_presence_adu_q_1</i>	
48	Presence of cardiovascular risk factors (recode into new variable)	Categorical / dichotomous	*If someone has answered ‘yes’ on one of the CVD related variables, CVD risk factors are present. IF (heartattack_presence_adu_q_1 = 1 OR Heartfailure_presence_adu_q_1 = 1 OR stroke_presence_adu_q_1 = 1) = 1) CHD_Risk_Presence = 1. CHD_Risk_Presence = 0. EXECUTE.	<i>CHD_Risk_Presence</i> 0 → no CHD risk 1 → CHD risk	
49	Cardiovascular disease (Lifelines 1A) “Have you ever been diagnosed with high cholesterol?”	Categorical / nominal	Health38 1= yes 2 = no 3= I do not know	<i>cholesterol_presence_adu_q_1</i>	
50	High cholesterol presence (Recode into new variable)	Categorical / dichotomous	IF (cholesterol_presence_adu_q_1 = 1) HighCholesterol_Presence = 1. ELSE HighCholesterol_Presence = 0.	<i>HighCholesterol_Presence</i> 0 → no presence of high cholesterol 1 → presence of high cholesterol	

51	Cardiovascular disease (Lifelines 1A) "Have you ever had hypertension"?	Categorical / nominal	Health35 1= yes 2= no 3= I do not know	<i>hypertension_presence_adu_q_1</i>	
52	Hypertension presence (recode into new variable)	Categorical / dichotomous	IF (hypertension_presence_adu_q_1 = 1) hypertension_presence = 1. ELSE hypertension_presence = 0.	<i>Hypertension_Presence</i> 0 → no presence of hypertension 1 → presence of hypertension	
53	Hearing problems (DALAS) 'Do you have hearing problems?'	Categorical / dichotomous	<i>UVHEALTH_HHIAS1</i> 1 = Yes 2 = No	<i>hhia_hearingproblems_adu_q_01</i>	
54	Hearing problems (recode into new variable)	Categorical / dichotomous	RECODE hhia_hearingproblems_adu_q_01 ('no' = 0) ('yes' = 1) INTO Hearing_Problems.	<i>Hearing_Problems</i> No → 0 Yes → 1	
55	Physical activity (DALAS)	Categorical / nominal	<i>UVSQUASH_SQUASH15A</i>	<i>squash_perweek_adu_1_15_a</i>	

	“On an average of how many days a week, in total, do you spend at least half an hour cycling, chores, gardening, sports or other strenuous activities during the summer? “	 days a week in summer on average active at least half an hour		
56	Physical activity (DALAS) “On an average of how many days a week, in total, do you spend at least half an hour cycling, chores, gardening, sports or other strenuous activities during the winter?”	Categorical / nominal	<i>UVSQUAS_SQUASH15B</i> Days a week in winter on average active at least half an hour	<i>squash_perweek_adu_1_15_b</i>	
57	Physical activity average*** (recode into new variable)	Continuous	*First calculate the average days of activity in summer and winter. COMPUTE squash_average = (squash_perweek_adu_1_15_a + squash_perweek_adu_1_15_b) / 2.	<i>Squash_average</i>	
58	Physical inactivity*** (recode into new variable)	Categorical / dichotomous	*Define those who do not meet the guidelines of 5 days IF (squash_average >= 5) Physical_inactivity = 0. IF (squash_average < 5) Physical_inactivity = 1.	<i>Physical_inactivity</i> Average of $\geq 5 = 0 \rightarrow$ these individuals meet the guidelines Average of $< 5 = 1 \rightarrow$ these individuals do not meet the guidelines	

59	Height (Lifelines 2A)	Continuous	Anthropometry - bodylength_cm_all_m_1	<i>bodylength_cm_all_m_1_2a</i>	cm
60	Weight (Lifelines 2A)	Continuous	Anthropometry - bodylength_cm_all_m_1	<i>bodyweight_kg_all_m_1_2a</i>	kg
61	Body Mass Index (BMI) (recode into new variable)	Continuous	*Calculate BMI COMPUTE BMI = bodyweight_kg_all_m_1_2a / ((bodylength_cm_all_m_1_2a/100) * (bodylength_cm_all_m_1_2a/100)).	<i>BMI</i>	Kg/m2
62	Obesity (Recode into new variable)	Categorical / dichotomous	*Indicate those with obesity IF (BMI >= 30) obesity = 1. IF (BMI < 30) obesity = 0.	<i>Obesity</i> BMI < 30 → 0 = no obesity BMI ≥ 30 → 1 = obesity <i>(Leave this out when there are too few cases of obesity)</i>	
63	Depressive symptoms (DALAS) “I felt worthless”	Categorical / nominal	promis9 <ul style="list-style-type: none">• Never• Rarely• Sometimes• Often• Always	<i>promis_depression_adu_q_09</i>	
64	Depressive symptoms (DALAS) “ I felt helpless”	Categorical / nominal	Promis10 <ul style="list-style-type: none">• Never• Rarely• Sometimes• Often	<i>promis_depression_adu_q_10</i>	

			<ul style="list-style-type: none"> • Always 		
65	Depressive symptoms (DALAS) “ I felt depressed”	Categorical / nominal	Promis11 <ul style="list-style-type: none"> • Never • Rarely • Sometimes • Often • Always 	<i>promis_depression_adu_q_11</i>	
66	Depressive symptoms (DALAS) “ I felt hopeless”	Categorical / nominal	Promis12 <ul style="list-style-type: none"> • Never • Rarely • Sometimes • Often • Always 	<i>promis_depression_adu_q_12</i>	
67	Assign scores to the PROMIS depression variables (recode into new variable)	Continuous	RECODE <i>promis_depression_adu_q_09</i> <i>promis_depression_adu_q_10</i> <i>promis_depression_adu_q_11</i> <i>promis_depression_adu_q_12</i> ('never'=1) ('rarely'=2) ('sometimes'=3) ('often'=4) ('always' = 5) INTO <i>Depression_09 Depression_10 Depression_11</i> <i>Depression_12.</i>	<i>Depression_09</i> <i>Depression_10</i> <i>Depression_11</i> <i>Depression_12</i> Never → 1 Rarely → 2 Sometimes → 3 Often → 4 Always → 5	
68	Take the sum of the depression scores (recode into new variable)	Continuous	COMPUTE <i>depression_score</i> = <i>Depression_09</i> + <i>Depression_10</i> + <i>Depression_11</i> + <i>Depression_12.</i>	<i>Depression_Score</i>	

69	Recode into T score and associated SE (Recode into new variable)	Continuous	<pre> DO IF (depression_score = 4). COMPUTE t_score_depression = 41.0. COMPUTE SE_score_depression = 6.2. ELSE IF (depression_score = 5). COMPUTE t_score_depression = 49.0. COMPUTE SE_score_depression = 3.2. ELSE IF (depression_score = 6). COMPUTE t_score_depression = 51.8. COMPUTE SE_score_depression = 2.7. ELSE IF (depression_score = 7). COMPUTE t_score_depression = 53.9. COMPUTE SE_score_depression = 2.4. ELSE IF (depression_score = 8). COMPUTE t_score_depression = 55.7. COMPUTE SE_score_depression = 2.3. ELSE IF (depression_score = 9). COMPUTE t_score_depression = 57.3. COMPUTE SE_score_depression = 2.3. ELSE IF (depression_score = 10). COMPUTE t_score_depression = 58.9. COMPUTE SE_score_depression = 2.3. ELSE IF (depression_score = 11). COMPUTE t_score_depression = 60.5. COMPUTE SE_score_depression = 2.3. ELSE IF (depression_score = 12). COMPUTE t_score_depression = 62.2. COMPUTE SE_score_depression = 2.3. ELSE IF (depression_score = 13). COMPUTE t_score_depression = 63.9. COMPUTE SE_score_depression = 2.3. ELSE IF (depression_score = 14). COMPUTE t_score_depression = 65.7. </pre>	<i>T_Score_Depression</i> <i>SE_Score_Depression</i>	
----	---	------------	--	---	--

			<pre> COMPUTE SE_score_depression = 2.3. ELSE IF (depression_score = 15). COMPUTE t_score_depression = 67.5. COMPUTE SE_score_depression = 2.3. ELSE IF (depression_score = 16). COMPUTE t_score_depression = 69.4. COMPUTE SE_score_depression = 2.3. ELSE IF (depression_score = 17). COMPUTE t_score_depression = 71.2. COMPUTE SE_score_depression = 2.4. ELSE IF (depression_score = 18). COMPUTE t_score_depression = 73.3. COMPUTE SE_score_depression = 2.4. ELSE IF (depression_score = 19). COMPUTE t_score_depression = 75.7. COMPUTE SE_score_depression = 2.6. ELSE IF (depression_score = 20). COMPUTE t_score_depression = 79.4. COMPUTE SE_score_depression = 2.6. END IF. </pre>		
70	Depressive symptoms (Recode into new variable)	Categorical / dichotomous	<pre> *Make categories to indicate those with depressive symptoms IF (t_score < 55) Depressive_Symptoms = 0. IF (t_score >= 55) Depressive_Symptoms = 1. </pre>	<i>Depressive_Symptoms</i>	<pre> 0 → no depressive symptoms 1 → depressive symptoms </pre>
71	Social network size (DALAS) 'On average, with how many different people do you have contact with in which you discuss personal matters over a period of two weeks?'	Continuous	UVSOCIAL_CONTACTSA	<i>contacts_number_adu_q_2</i>	

72	Social network size (Recode into new variable based on the mean or median)	Categorical / dichotomous	IF (contacts_number_adu_q_2 >= MEAN) Social_Network_Size = 0. IF (contacts_number_adu_q_2 < MEAN) Social_Network_Size = 1.	<i>Social_Network_Size</i> 0 → large social network size 1 → small social network size <i>Based on the distribution of the data I might use another approach or make an additional category, or use quartiles to make categories.</i>	
73	Loneliness (DALAS) “ I experience an emptiness around me”	Categorical / nominal	LONELYSC1 • yes! • yes • more or less • no • no!	<i>loneliness_emptiness_adu_q_01</i>	
74	Loneliness (DALAS) “I miss people around me”	Categorical / nominal	LONELYSC2 • yes! • yes • more or less • no • no!	<i>loneliness_people_adu_q_02</i>	
75	Loneliness (DALAS) “I often feel abandoned”	Categorical / nominal	LONELYSC3 • yes! • yes • more or less • no • no!	<i>loneliness_abandoned_adu_q_03</i>	

76	Loneliness (DALAS) “I have enough people to fall back on in case of adversity”	Categorical / nominal	LONELYSC4 <ul style="list-style-type: none"> • yes! • yes • more or less • no • no! 	<i>loneliness_fallback_adu_q_04</i>	
77	Loneliness (DALAS) “I have many people that I can fully trust”	Categorical / nominal	LONELYSC5 <ul style="list-style-type: none"> • yes! • yes • more or less • no • no! 	<i>loneliness_trust_adu_q_05</i>	
78	Loneliness (DALAS) “There are enough people that I feel closely connected to”	Categorical / nominal	LONELYSC6 <ul style="list-style-type: none"> • yes! • yes • more or less • no • no! 	<i>loneliness_connection_adu_q_06</i>	
79	Assign scores for emotional loneliness (recode into new variable)	Continuous	RECODE loneliness_emptiness_adu_q_01 loneliness_people_adu_q_02 loneliness_abandoned_adu_q_03 ('yes!'=1) (‘yes’=1) (‘more or less’=1) (‘no’=0) (‘no!’=0) (ELSE=SYSMIS) INTO Emotional_Loneliness_01 Emotional_Loneliness_02 Emotional_Loneliness_03.	<i>Emotional_Loneliness_01</i> <i>Emotional_Loneliness_02</i> <i>Emotional_Loneliness_03</i> ‘Yes!’, ‘yes’ and ‘more or less’ = 1 ‘No’ and ‘no!’ = 0	

80	Emotional loneliness score (recode into new variable)	Continuous	* Calculate the sum and define missings COMPUTE Emotional_Loneliness_Score = Emotional_Loneliness_01 + Emotional_Loneliness_02 + Emotional_Loneliness_03.	<i>Emotional_Loneliness_Score</i>	
			DO IF MISSING(Emotional_Loneliness_01) MISSING(Emotional_Loneliness_02) MISSING(Emotional_Loneliness_03). COMPUTE Emotional_Loneliness_Score = \$SYSMIS. END IF. EXECUTE.		
81	Assign scores for social loneliness (recode into new variable)	Continuous	RECODE loneliness_fallback_adu_q_04 loneliness_trust_adu_q_05 loneliness_connection_adu_q_06 (‘more or less’=1) (‘no’=1) (‘no!’=1) (‘yes!’=0) (‘yes’=0) (ELSE=SYSMIS) INTO Social_Loneliness_04 Social_Loneliness_05 Social_Loneliness_06.	<i>Social_Loneliness_04</i> <i>Social_Loneliness_05</i> <i>Social_Loneliness_06.</i> ‘no!’, ‘no’ and ‘more or less’ = 1 ‘yes’ and ‘yes!’ = 0	
82	Social Loneliness Score (Recode into new variable)		COMPUTE Social_Loneliness_Score = Social_Loneliness_04 + Social_Loneliness_05 + Social_Loneliness_06. DO IF MISSING(Social_Loneliness_04) MISSING(Social_Loneliness_05) MISSING(Social_Loneliness_06). COMPUTE Social_Loneliness_Score = \$SYSMIS. END IF.	<i>Social_Loneliness_Score</i>	

83	Total Loneliness Score (recode into new variable)	Continuous	COMPUTE Total_Loneliness_Score = Emotional_Loneliness_Score + Social_Loneliness_Score.	<i>Total_Loneliness_Score</i>	
84	Loneliness presence	Categorical / dichotomous	DO IF Total_Loneliness_Score >= 0 AND Total_Loneliness_Score <= 1. COMPUTE Loneliness_Presence = 0. /* Not lonely ELSE IF Total_Loneliness_Score >= 2 AND Total_Loneliness_Score <= 6. COMPUTE Loneliness_Presence = 1. /* Lonely ELSE. COMPUTE Loneliness_Presence = \$SYSMIS. /* Missing END IF.	<i>Loneliness_Presence</i> 0-1 not lonely = no → 0 2-6 lonely = yes → 1	
85	Informal caregiving (DALAS) 'providing informal care / have you engaged in the following activities in the past 12 months?'	Categorical / dichotomous	UVTIJDBES_ACTIV2 Informal caregiving 1 = yes 2 = no	<i>activities_informalcare_adu_q_1</i>	
86	Informal caregiving (recode into new variable)	Categorical/ dichotomous	RECODE activities_informalcare_adu_q_1 ('yes'=1) ('no'= 0) (ELSE=0) INTO Informal_Caregiving.	<i>Informal_Caregiving</i> No → 0 Yes → 1	
87	Grandchildren (DALAS) "Do you have grandchildren?"	Categorical / dichotomous	UVTIJDBES_GRANDCH1 1 = yes 2 = No	<i>grandchildren_current_adu_q_1</i>	

88	Taking care of grandchildren (DALAS)	Categorical / dichotomous	UVTIJDBES_GRANDCH1 1 = yes 2 = no	<i>grandchildren_care_adu_q_1</i>	
	'have you cared for your grandchildren in the past 12 months?'				
89	Taking care of grandchildren (recode into new variable)	Categorical / dichotomous	RECODE grandchildren_care_adu_q_1 ('yes'=1) ('no'=0) (ELSE=0) INTO Taking_Care_Grandchildren	Taking_Care_Grandchildren No → 0 Yes → 1	
90	Social roles and activities (DALAS) "I have difficulty engaging in all my normal leisure pursuits with others"	Categorical / nominal	Promis21 • Never • Rarely • Sometimes • Usually • Always	<i>promis_social_adu_q_21</i>	
91	Social roles and activities (DALAS) "I have difficulty doing all the family activities that I want to"	Categorical / nominal	Promis22 • Never • Rarely • Sometimes • Usually • Always	<i>promis_social_adu_q_22</i>	

92	Social roles and activities (DALAS) “I have difficulty doing all my normal work (including work at home)”	Categorical / nominal	Promis23 <ul style="list-style-type: none"> • Never • Rarely • Sometimes • Usually • Always 	promis_social_adu_q_23	
93	Social roles and activities (DALAS)	Categorical / nominal	Promis24	promis_social_adu_q_24	

	“I have difficulty doing all the activities with friends that I want to”		<ul style="list-style-type: none"> • Never • Rarely • Sometimes • Usually • Always 		
94	Assign scores to the PROMIS social roles and activities variables (recode into new variable)	Continuous	RECODE promis_social_adu_q_21 promis_social_adu_q_22 promis_social_adu_q_23 promis_social_adu_q_24 ('never'=5) ('rarely'=4) ('sometimes'=3) ('usually'=2) ('always'=1) INTO Social_21 Social_22 Social_23 Social_24.	<i>Social_21</i> <i>Social_22</i> <i>Social_23</i> <i>Social_24</i> Always = 1 Usually = 2 Sometimes = 3 Rarely = 4 Never = 5	
95	Take the sum of the depression scores (recode into new variable)	Continuous	COMPUTE Social_Score = Social_21 + Social_22 + Social_23 + Social_24.	<i>Social_Participation_Score</i>	

96	Recode into T score and associated SE (Recode into new variable)	Continuous	<pre> DO IF (Social_Score = 4). COMPUTE T_Score_Social = 27.5. COMPUTE SE_Score_Social = 4.1. ELSE IF (Social_Score = 5). COMPUTE T_Score_Social = 31.8. COMPUTE SE_Score_Social = 2.5. ELSE IF (Social_Score = 6). COMPUTE T_Score_Social = 34.0. COMPUTE SE_Score_Social = 2.3. ELSE IF (Social_Score = 7). COMPUTE T_Score_Social = 35.7. </pre>	<i>T_Score_Social</i> <i>SE_Score_Social</i>	
----	---	------------	---	---	--

			<p> COMPUTE SE_Score_Social = 2.2. ELSE IF (Social_Score = 8). COMPUTE T_Score_Social = 37.3. COMPUTE SE_Score_Social = 2.1. ELSE IF (Social_Score = 9). COMPUTE T_Score_Social = 38.8. COMPUTE SE_Score_Social = 2.2. ELSE IF (Social_Score = 10). COMPUTE T_Score_Social = 40.5. COMPUTE SE_Score_Social = 2.3. ELSE IF (Social_Score = 11). COMPUTE T_Score_Social = 42.3. COMPUTE SE_Score_Social = 2.3. ELSE IF (Social_Score = 12). COMPUTE T_Score_Social = 44.2. COMPUTE SE_Score_Social = 2.3. ELSE IF (Social_Score = 13). COMPUTE T_Score_Social = 46.2. COMPUTE SE_Score_Social = 2.3. ELSE IF (Social_Score = 14). COMPUTE T_Score_Social = 48.1. COMPUTE SE_Score_Social = 2.2. ELSE IF (Social_Score = 15). COMPUTE T_Score_Social = 50.0. COMPUTE SE_Score_Social = 2.2. ELSE IF (Social_Score = 16). COMPUTE T_Score_Social = 51.9. COMPUTE SE_Score_Social = 2.3. ELSE IF (Social_Score = 17). COMPUTE T_Score_Social = 53.7. COMPUTE SE_Score_Social = 2.3. ELSE IF (Social_Score = 18). COMPUTE T_Score_Social = 55.8. </p>	
--	--	--	---	--

			<pre> COMPUTE SE_Score_Social = 2.3. ELSE IF (Social_Score = 19). COMPUTE T_Score_Social = 58.3. COMPUTE SE_Score_Social = 2.7. ELSE IF (Social_Score = 20). COMPUTE T_Score_Social = 64.2. COMPUTE SE_Score_Social = 5.1. END IF. EXECUTE. </pre>		
97	Ability to participate in social roles and activities (recode into new variable)	Categorical / dichotomous	<pre> IF (T_Score_Social > 46) Social_Participation = 0. IF (T_Score_Social <= 46) Social_Participation = 1. </pre>	<i>Social_Participation</i> 0 = able to participate socially 1 = less able to participate socially	
98	“Have you ever participated in voluntary work?” (DALAS)	Categorical / dichotomous	<pre> UVVOLUNT_VOLUNT21 1 = yes 2 = no </pre>	<i>volunteering_lifetime_adu_q_1</i>	
99	History of voluntary work *** (recode into new variable)	Categorical / dichotomous	<pre> RECODE volunteering_lifetime_adu_q_1 (1 = 1) (2 = 0) INTO History_Volunteering </pre>	<i>Volunteering_History</i> 0 → those without a history of volunteering 1 → those with a history of volunteering	
100	Church attendance (DALAS) ‘How often have you attended a religious gathering in the past 12 months?’	Categorical / nominal	<pre> 1 = multiple times per week 2 = 1 time per week 3 = 1 time per 2 weeks 4 = 1 time per month 5 = less than 1 time per month </pre>	<i>religion_current_adu_q_1_c</i>	

			6 = 7 = 1 time 8 = not a single time		
101	Church attendance (recode into new variable)	Categorical / dichotomous	RECODE religion_current_adu_q_1_c ('multiple times per week' = 1) ('1 time per week' = 1) ('1 time per 2 weeks' = 1) ('1 time per month' = 1) ('less than 1 time per month' = 0) ('1 time' = 0) ('not a single time' = 0) INTO Church_Attendance.	<i>Church_Attendance</i> 8-5 → no = 0 (less than 1x per month) 4-1 → yes = 1 (1x per month or more)	

References

- International Labour Office Geneva. (2011). Manual on the measurement of volunteer work.
- LifeLines Wiki. (n.d.). Cogstate. <http://wiki-lifelines.web.rug.nl/doku.php?id=cogstate&do=>
- Turney, S. (2023, November 10). Skewness | Definition, Examples & Formula. Scribbr. Retrieved January 22, 2024, from <https://www.scribbr.com/statistics/skewness/>
- Turney, S. (2023a, June 22). What is Kurtosis? | Definition, Examples & Formula. Scribbr. <https://www.scribbr.com/statistics/kurtosis/>
- Bhandari, P. (2023, June 21). Missing data | Types, explanation, & imputation. Scribbr. <https://www.scribbr.com/statistics/missing-data/>
- OECD (2018). Compare your income – Methodology and conceptual issues. 2018. Available from: <https://www.oecd.org/statistics/Compare-your-incomemethodology.pdf>.
- Schiepers, O. J. G., Köhler, S., Deckers, K., Irving, K., O'Donnell, C. A., van den Akker, M., Verhey, F. R. J., Vos, S. J. B., de Vugt, M. E., & van Boxtel, M. P. J. (2018). Lifestyle for Brain Health (LIBRA): a new model for dementia prevention. *International journal of geriatric psychiatry*, 33(1), 167–175.
<https://doi.org/10.1002/gps.4700>
- National Heart, Lung, and Blood Institute, National Institutes of Health. (2023, December 20). What Is Coronary Heart Disease?
<https://www.nhlbi.nih.gov/health/coronary-heart-disease>
- National Health Service. (2023, March 10). Coronary Heart Disease.
<https://www.nhs.uk/conditions/coronary-heart-disease/treatment/>
- Ministerie van Volksgezondheid, Welzijn en Sport. (2021, November 29). Beweegrichtlijnen. Publicatie | Rijksoverheid.nl.
<https://www.rijksoverheid.nl/documenten/publicaties/2020/07/07/beweegrichtlijnen>
- Voedingscentrum. (n.d.). BMI berekenen. <https://www.voedingscentrum.nl/bmi>
- Van Tilburg. (2023, June 22). Manual De Jong Gierveld Loneliness Scale. OSF Home.
<https://doi.org/10.17605/OSF.IO/U6GCK>
- Dutch Flemish Promis National Center (n.d.). Short Forms. https://dutchflemishpromis.nl/short-forms_19_16.html
- Health Measures. (2021, 30 July). Depression Scoring Manual.
https://www.healthmeasures.net/images/PROMIS/manuals/Scoring_Manuals/PROMIS_Depression_Scoring_Manual.pdf
- Health Measures. (2016, June 26). PROMIS Item Bank v1.0 – Emotional Distress-Depression – Short Form 4a.
https://www.healthmeasures.net/administrator/components/com_instruments/uploads/PROMIS%20SF%20v1.0-%20ED-Depression%204a%206-262016.pdf

Dutch-Flemish PROMIS National Center. (2023, May 1). Nederlandse PROMIS Referentie Waarde. <https://www.dutchflemishpromis.nl/images/upload/files/Nederlandse%20PROMIS%20referentiewaarden%20v7.pdf>

De Jong Gierveld, J., & Van Tilburg, T. (2022, February 24). Manual of the Loneliness Scale. Vrije Universiteit, Faculty of Social Sciences, Department of Sociology. https://home.fsw.vu.nl/TG.van.Tilburg/manual_loneliness_scale_1999.html

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N. C., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., . . . Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, 396(10248), 413–446. [https://doi.org/10.1016/s0140-6736\(20\)30367-6](https://doi.org/10.1016/s0140-6736(20)30367-6)

Niebuur, J., Van Lenté, L., Liefbroer, A. C., Steverink, N., & Smidt, N. (2018). Determinants of participation in voluntary work: a systematic review and metaanalysis of longitudinal cohort studies. *BMC Public Health*, 18(1). <https://doi.org/10.1186/s12889-018-6077-2>

Health Measures. (2018a, July 10). PROMIS® Item Bank v2.0 - Ability to Participate in Social Roles and Activities- Short Form 4a. https://www.healthmeasures.net/index.php?option=com_instruments&view=measure&id=192&Itemid=992

Health Measures. (2018b, July 23). Ability to Participate in Social Roles and Activities Scoring Manual. https://www.healthmeasures.net/images/promis/manuals/PROMIS_Ability_to_Participate_in_Social_Roles_and_Activities_Scoring_Manual.pdf