TREBALL DE FI DE MÀSTER

# Màster en Estadística i Investigació Operativa

**Títol: Comparison of structural equation models with observed and latent variables: An application to the mediating role of disability in the impact of common conditions on perceived health.**

**Autora: Núria Duran Adroher**

**Director: Carlos García Forero**
**Ponent: Klaus Gerhard Langohr**

**Institució: IMIM-Institut Hospital del Mar d'Investigacions Mèdiques**

**Departament: Estadística i Investigació Operativa**

**Convocatòria: Gener 2014**



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

**Universitat Politècnica de Catalunya**
**Facultat de Matemàtiques i Estadística**

Treball de Fi de Màster
Màster en Estadística i Investigació Operativa

# Comparison of structural equation models with observed and latent variables: An application to the mediating role of disability in the impact of common conditions on perceived health

Núria Duran Adroher

Director: Carlos García Forero
Ponent: Klaus Gerhard Langohr

IMIM-Insitut Hospital del Mar d'Investigacions Mèdiques

Departament d'Estadística i Investigació Operativa

Gener 2014

# Abstract

**Background**: Structural Equation Modeling (SEM) allows to study the simultaneous relationships among chronic conditions and perceived health mediated by disability dimensions. We hypothesized that considering some items as indicator variables of the underlying concept they describe (a latent variable) would provide more accurate estimates and better fit than using only observed scores.

**Methods**: Two Complex Disability Mediated Models —CDMM-O (with all the variables Observed) and CDMM-L (with some Latent variables)— were fitted in a sample of the WHO World Mental Health (WMH) Surveys including 11 countries (n=24,797), and taking into account the complex sampling design. A visual analog scale (VAS) measured perceived health and disability was assessed using a modified version of the WHO Disability Assessment Schedule (WHODAS). Nine common mental and ten common physical conditions were considered. SEM was used to estimate total effects of conditions on perceived health, their separate direct and indirect effects, and their specific indirect effects. Before comparing CDMM-O and CDMM-L in terms of parameter estimates, standard errors and model fit, a Confirmatory Factor Analysis (CFA) for the indicators of the latent variables was conducted.

**Results**: The CFA presented excellent fit (RMSEA=0.011, CFI=TLI=0.999). A better fit was observed for CDMM-L. CFI and TLI were not acceptable for CDMM-O. Standard errors were lower for CDMM-L, and parameter estimates were more distinct among CDMM-O and CDMM-L than expected. CDMM-O presented inconsistent estimates: a negative proportion of indirect over total effect for Drug abuse, and positive direct and specific indirect effects for Getting along. Cognition was the third most important dimension for CDMM-L, while it occupied the fifth position for CDMM-O.

**Conclusions**: A model with latent variables is preferred; benefits of assessing pure relations, without measurement error, were observed even treating a few number of variables as latent: CDMM-L corrected the inconsistencies present in CDMM-O, and more precise estimates were obtained.

**Keywords**: Structural Equation Modeling, latent variables, mediation, Patient-Reported Outcomes

# Acknowledgements

I would like to express my deep gratitude to Carlos García Forero, director of this master thesis, for his patient and effective guidance, and his valuable and constructive suggestions. I would also like to thank Gemma Vilagut and Jordi Alonso, for their useful critiques. My grateful thanks are also extended to the members of the Health Services Group, specially to Stefi, Elisabeth, Gabi, Oleguer, Pere, Elena, Roser, Carme, Albert, Mónica, Àngels, and Olatz.

I am particularly grateful to Klaus Langohr for having accepted to tutor this project and for his dedicated support.

I would like to give special thanks to Rosa, Jaume, and Joan.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Motivation

Patient-reported outcomes (PROs) are reports coming directly from patients about how they feel or function in relation to a health condition and its therapy without interpretation by healthcare professionals or anyone else. PROs include any treatment or outcome evaluation obtained directly from patients through interviews, self-completed questionnaires, diaries or other data collection tools such as hand-held devices and web-based forms.

PROs provide patients' perspective on treatment benefit. They directly measure treatment benefit beyond survival, disease, and physiologic markers. Besides, they are often the outcomes of greatest importance to patients. Reports from patients may include the signs and symptoms reported in diaries, the evaluation of sensations (most commonly classified as symptoms), reports of behaviours and abilities (most commonly classified as functional status), and general perceptions or feelings of well-being. Other reports including satisfaction with treatment, general or health-related quality of life, and adherence to treatments are also considered PROs by the Food and Drug Administration (FDA).

PROs are gaining more importance in medical evaluation studies. For instance, when assessing Health-Related Quality of Life outcomes, both objective components such as income, and subjective components such as the health state perceived and reported by the patient, take part into the evaluation. However, these subjective dimensions are becoming more and more relevant, but they are not directly observable. In many situations, these PRO measures are referred to as *latent traits*.

Hence, in PRO assessment, one has to deal with indirectly observable concepts obtained from self-reports, such as quality of life, welfare, perceived functioning and disability. Such concepts are referred to as **latent variables** (Borsboom et al., 2003) because they are

inferred from other **observed** variables that are assummed to be dependent (i.e. caused) of an unobservable dimension. Thus, observed variables are indicators of the concept they represent. Under standard quantitative models of person-response, their appearance is caused by the individual quantity in the latent trait. Moreover, an observed variable contains random or systematic measurement error, but the latent trait is free of these sources of error (Bollen, 1989) and only contains estimation error. Therefore, studying the relations of concepts using latent variables is more accurate than only using observed variables.

A situation in which the use of latent variables seems appropiate is when assessing the effects of common conditions on perceived health. In fact, my collegues from IMIM-Institut Hospital del Mar d'Investigacions Mèdiques (Health Services Group) and I suspected that disability dimensions *mediate* those effects; for instance, when assessing the total effect of arthritis on perceived health, we guessed that part of the effect of the condition passed through the disability in mobility caused by it. If mediation actually exists, the intervention is possible not only through the trigger (arthritis) but through the mediator variable (mobility) as well; thus, a window of opportunity on intervention that may be useful in the context of clinical attention and health services and policies, and offer causal explanation, is opened.

A structural equation model (SEM) allows to study the simultaneous relationships among the variables by defining multi-equation regression models. With a database containing 19 predictors (common conditions), 8 mediators (disability dimensions) and an outcome (perceived health), SEM is an appropriate tool to describe such relationships. Note that there are $19 \cdot 8 = 152$ simple mediation models, and considered together make up multiple mediation paths which enable to explain the simultaneous effects of the predictors on the outcome. Thus, a Complex Disability Mediated Model (**CDMM**) will be defined in order to take into account the indirect effects through the disability dimensions in two different scenarios: (1) The CDMM will only contain observed variables, and this model will be denoted as **CDMM-O**; (2) the CDMM will contain both observed and latent variables, and this model will be denoted as **CDMM-L**. We hypothesize that considering some of the items describing the disability dimensions as indicator variables of the underlying concept they describe (a latent variable) would provide more accurate estimates than using only observed scores.

## 1.2 Background

Perceived health is widely recognized as an important indicator of health, and is often used to monitor health trends in the general population (Rohrer et al., 2007; Perruccio et al.,

2007) as well as to assess patient-centered outcomes in clinical studies (Alonso, 2000).

Chronic conditions are among the most important predictors of perceived health (Saarni et al., 2006; Schultz and Kopec, 2003; Alonso et al., 2004). Some conditions, such as those causing pain, are known to be associated with substantial decrements in perceived health (van Dijk et al., 2008); important decrements associated with neurological conditions, depression and arthritis —once the presence of other conditions had been taken into account (Alonso et al., 2011)—, and a higher impact of mental conditions (as compared to other medical conditions) have also been reported (Ormel et al., 1998).

A number of conceptual frameworks and models of health propose that disability mediates the effects of chronic disorders on perceived health (Wilson and Cleary, 1995). Mediation models explain *how* an effect occurred by hypothesizing a causal sequence: the independent variable $x$ (condition) causes the mediator $m$ (disability dimension) which in turn causes the dependent variable $y$ (perceived health), therefore explaining how $x$ had its effect on $y$. In general, a given variable may be said to function as a mediator to the extent that it accounts for the relation between the predictor and the outcome (Baron and Kenny, 1986).

In the literature there is evidence showing that disability is significantly associated with perceived health both cross-sectionally (Lee et al., 2008) and longitudinally (Leinonen et al., 2001); and that chronic conditions are significantly associated with disability (Ormel et al., 2008). Based on this evidence, my collegues and I carried out a multidimensional assessment to explore the extent to which disability mediates the associations of 19 chronic contitions (9 mental, 10 physical) on perceived health in the epidemiological sample of the World Health Organization World Mental Health Survey Initiative (wmh). A paper describing these associations was published (Alonso et al., 2013), using observed variables, and the purpose of this master thesis is going a step further examining the model in more detail and using latent variables.

# Chapter 2

# Objectives

### General objective

The aim of the present master thesis is to compare the results of the Complex Disability Mediated Model (CDMM) using observed variables (CDMM-O), and using latent variables (CDMM-L).

### Specific objectives

- To analyze to what extent the extraction of the measurement error, as a consequence of using latent variables, has an influence on effect estimates and standard errors.
- To compare the fit of the CDMM-O and the CDMM-L.

### Hypothesis

- The effect estimates will be similar for both the CDMM-O and the CDMM-L.
- The standard errors will be lower for the CDMM-L than for the CDMM-O.
- The fit of the CDMM-L will be better than the fit of the CDMM-O.

# Chapter 3

# Materials and Methods

## 3.1 Sample

The World Mental Health (WMH) Survey Initiative is a World Health Organization (WHO) initiative designed to help countries carry out and analyze epidemiological surveys of the burden of mental disorders in their populations (wmh). The sample analyzed in this report consists of a total of 11 nationally representative surveys classified as high income countries by the World Bank (2009) at the time of data collection: Belgium, France, Germany, Israel, Italy, Japan, Netherlands, Northern Ireland, Portugal, Spain, and the United States of America.

The weighted average response rate across countries was 63.5% with country-specific response rates ranging from 45.9% (France) to 78.5% (Spain). The minimum age was 18 years, and the upper age was unrestricted.

The WMH surveys required collaborating countries to employ probability sample designs to select nationally or regionally representative samples of adults for the survey interview. The aim of sampling in the WMH surveys was to obtain a representative sample of the general population in the country or region under study. This usually involved drawing a multistage (generally a three-stage or four-stage) clustered area probability sample of households in the population and then selecting one, or in some cases two, respondents from each sampled household using probability methods without replacement. These sample designs were standardized across countries based on the principles of probability sampling, but with less emphasis placed on the specific probability sample design features employed across countries in recognition of the fact that countries varied widely in the information available to develop a sample frame from which the WMH sample could be selected.

Except for Israel, each interview had two parts. All respondents completed Part 1; the interview began with a series of basic descriptive warm-up questions and then evaluated lifetime presence of a wide range of core mental disorders. All the respondents who met criteria for any of these disorders were continued with Part 2, which included questions about a wide range of correlates of the core disorders and also assessed mental disorders of secondary interest. In addition, a probability sub-sample of other Part 1 respondents (i.e., those who did not meet criteria for any core disorder) were also selected to complete Part 2 while interviews with the remaining non-cases were ended after the completion of the Part 1 questions. In Israel, all individuals completed Part 2. Data were weighted to adjust for differential probabilities of selection and to match population distributions on socio-demographic and geographic data. An additional weight was used for the over sampling of respondents for the Part 2 sample.

A total 24,797 respondents (Part 2 respondents) were assessed in the present analysis.

## 3.2 Variables

Table 3.1 summarizes the role, type, status, and metric of the variables present in the models. Below they are grouped by type and described with more detail:

**Mental disorders** All WMH surveys use the same standardized procedures for sampling, interviewing, and data analysis. They also use the same diagnostic interview for mental disorders, the WHO Composite International Diagnostic Interview (CIDI) Version 3.0. The CIDI is a fully-structured research diagnostic interview designed for use by trained lay interviewers who do not have clinical experience. It generates diagnoses of mental disorders according to the definitions and criteria of the Diagnostic and Statistical Manual of the American Psychiatric Association, IV$^{th}$ edition (DSM-IV, APA (2000)). Consistent WHO translation, back-translation, and harmonization procedures were used to modify the CIDI for use in each WMH country. The same interviewer training materials, training programs, and quality control monitoring procedures were also used across WMH surveys to guarantee cross-survey comparability of data.

The nine mental conditions included in the analysis are: Alcohol abuse with or without dependence, Bipolar disorder (mania, hypomania, bipolar I, bipolar II), Major depressive disorder, Drug abuse with or without dependence, Generalized anxiety disorder (GAD), Panic disorder (panic disorder, agoraphobia without panic), Posttraumatic stress disorder (PTSD), Social phobia, and Specific phobia.

**Chronic Physical conditions** Physical conditions were assessed with a standard chronic condition checklist that asked respondents if they had ever suffered from the given physical

health condition, if they had the condition in the past 12 months and if they had received any treatment. Checklists like this have been shown to yield more complete and accurate reports than estimates derived from responses to open-ended questions. Methodological studies have documented a moderate to good concordance between such condition reports and medical records (Baumeister et al., 2010).

The ten physical conditions included in the analysis are: Arthritis, Cancer, Cardiovascular (heart attack, heart disease, hypertension, and stroke), Chronic pain (chronic back or neck pain, and other chronic pain), Diabetes, Digestive disorders (stomach or intestinal ulcer, irritable bowel condition), frequent or severe Headaches or Migraines, Insomnia, Neurological (multiple sclerosis, Parkinson's, epilepsy, or seizures), and Respiratory (seasonal allergies, asthma, or COPD or emphysema).

**Disability** Disability consists of eight dimensions and was assessed with a modified version of the World Health Organization Disability Assessment Schedule (WHODAS) 2.0 (who). Seven dimensions —Cognition, Getting along, Mobility, Self-care, Discrimination, Family burden and Stigma— were measured through a series of ordinal items with a 5 Likert-type scale: the scores 1 (no disability), 2, 3, 4, and 5 (extreme disability) were rescaled to 0, 1, 2, 3, and 4 in order to make them more comparable when assesing the subsample of individuals with difficulties; in that way, a score above 0 implies having disability problems. The respondents were asked about how much difficulty they had had during the past 30 days. A different number of items was used to describe each dimension:

- Cognition: 4 ordinal items concerning difficulties in concentrating, understanding, remembering, and learning a new task. The sum score ranges from 0 to 16.
- Getting along: 5 ordinal items concerning difficulties in conversing with people, dealing with unknown people, maintaining and making friends, and controlling emotions. The sum score ranges from 0 to 20.
- Mobility: 3 ordinal items concerning difficulties in standing, getting around, and walking. The sum score ranges from 0 to 12.
- Self-care: 3 ordinal items concerning difficulties in attending personal hygiene, dressing, and staying alone. The sum score ranges from 0 to 12.
- Discrimination: 1 ordinal item concerning the discrimination or unfair treatment experienced due to the health condition. The sum score ranges from 0 to 4.
- Family burden: 1 ordinal item concerning the interference of the health condition on the day to day activities of their family members. The sum score ranges from 0 to 4.

- Stigma: 1 ordinal item concerning the extent of embarrassment experienced due to the heath condition. The sum score ranges from 0 to 4.

The remaining dimension, Role functioning, consists of a set of 4 item questions asking about the number of days with activity limitation in the last 30 days. A weighted score ranging from 0 to 30 was obtained from the 4 questions.

In all the dimensions, the minimum score of 0 implies no difficulty, while the maximum (4, 12, 16, 20 or 30) implies complete difficulty.

Cognition, Getting along, Mobility, and Self-care are the only variables that will be treated both as observed and latent.

**Perceived health** Perceived health was assessed using a visual analog scale (VAS) approach (Paul-Dauphin et al., 1999). Respondents were asked to use a 0 to 100 scale, where 0 represents the worst possible health a person can have and 100 perfect health, to describe their own overall physical and mental health during the past 30 days.

The items corresponding to disability and perceived health are presented in Table A.1.

**Sociodemographics** The five sociodemographic variables (covariates) included in the model are the following (the bold categories are the reference ones):

- Age, continuous.
- Country, categorical: 11 countries mentioned in Section 3.1. United states is the reference one.
- Employment status, categorical:
  - **Working**
  - Student
  - Homemaker
  - Retired
  - Other
- Marital status, categorical:
  - **Never married**
  - Married/Cohabiting
  - Separated/Widowed/Divorced
- Sex, dichotomous:

- **Female**
- Male

| Variable | Role | Type | Status | Metric | Observations |
|---|---|---|---|---|---|
| Alcohol abuse | Predictor | Dichotomous | Observed | CIDI | Onset in the last 12 months |
| Bipolar | Predictor | Dichotomous | Observed | CIDI | Onset in the last 12 months |
| Depression | Predictor | Dichotomous | Observed | CIDI | Onset in the last 12 months |
| Drug abuse | Predictor | Dichotomous | Observed | CIDI | Onset in the last 12 months |
| Generalized anxiety | Predictor | Dichotomous | Observed | CIDI | Onset in the last 12 months |
| Panic disorder | Predictor | Dichotomous | Observed | CIDI | Onset in the last 12 months |
| Posttraumatic stress | Predictor | Dichotomous | Observed | CIDI | Onset in the last 12 months |
| Social phobia | Predictor | Dichotomous | Observed | CIDI | Onset in the last 12 months |
| Specific phobia | Predictor | Dichotomous | Observed | CIDI | Onset in the last 12 months |
| Arthritis | Predictor | Dichotomous | Observed | Checklists | |
| Cancer | Predictor | Dichotomous | Observed | Checklists | |
| Cardiovascular | Predictor | Dichotomous | Observed | Checklists | |
| Chronic pain | Predictor | Dichotomous | Observed | Checklists | |
| Diabetes | Predictor | Dichotomous | Observed | Checklists | |
| Digestive disorders | Predictor | Dichotomous | Observed | Checklists | |
| Headaches/Migraines | Predictor | Dichotomous | Observed | Checklists | |
| Insomnia | Predictor | Dichotomous | Observed | Checklists | |
| Neurological | Predictor | Dichotomous | Observed | Checklists | |
| Respiratory | Predictor | Dichotomous | Observed | Checklists | |
| Cognition | Mediator | Continuous | Obs/Lat | WHODAS | Indicators are 4 ordinal items |
| Getting along | Mediator | Continuous | Obs/Lat | WHODAS | Indicators are 5 ordinal items |
| Mobility | Mediator | Continuous | Obs/Lat | WHODAS | Indicators are 3 ordinal items |
| Self-care | Mediator | Continuous | Obs/Lat | WHODAS | Indicators are 3 ordinal items |
| Role functioning | Mediator | Continuous | Observed | WHODAS | |
| Discrimination | Mediator | Continuous | Observed | WHODAS | |
| Family Burden | Mediator | Continuous | Observed | WHODAS | |
| Stigma | Mediator | Continuous | Observed | WHODAS | |
| Perceived health | Outcome | Continuous | Observed | VAS | |
| Age | Covariate | Continuous | Observed | | |
| Country | Covariate | Categorical | Observed | | 11 categories |
| Employment status | Covariate | Categorical | Observed | | 5 categories |
| Marital status | Covariate | Categorical | Observed | | 3 categories |
| Sex | Covariate | Dichotomous | Observed | | 2 categories |

Table 3.1: Role, type, status and metric of the variables included in the models

## 3.3 Statistical software

SUDAAN V11.0 (RTI International, USA) was used to generate estimates of condition prevalence and descriptive statistics for the distributions of the variables. It is an internationally recognized statistical software package that specializes in providing efficient and accurate analysis of data from complex studies; it is ideal for the proper analysis of data from surveys and experimental studies, since its procedures properly account for complex design features, such as clustering, weighting, and stratification. Its procedures were implemented in a SAS V9.2 program.

Mplus 7.1 (Muthén and Muthén, Los Angeles, CA) was used to adjust the Structural Equation Models, with and without latent variables. It is a latent variable modeling program with a wide variety of analysis capabilities, including Structural Equation Modeling.

Both SUDAAN and Mplus use a Taylor series linearization method to estimate variances in complex sample surveys (Binder, 1983).

R (R Foundation for Statistical Computing, Vienna, Austria), version 3.0.0, was used to produce graphics.

# Chapter 4

# Statistical Methods

In the previous chapters I mentioned the keywords observed and latent variables, total, direct and indirect effects, mediation models, and structural equation models (SEM). In fact, SEM encompasses all of those keywords, and in this Section we will have a closer look to all of them.

## 4.1 Definition of SEM and Covariance role

Structural equation models (SEM) are multi-equation regression models. Unlike the more traditional multivariate linear model, however, the response variable in one regression equation in SEM may appear as a predictor in another equation; indeed, variables in SEM may influence one-another reciprocally, either directly or through other variables as intermediaries. These structural equations are meant to represent causal relationships among the variables in the model (Fox and Weisberg, 2011).

In structural equation modeling, instead of minimizing functions of observed and predicted individual values (as it is done in multiple regression or ANOVA), **the difference between the sample covariances and the covariances predicted by the model is minimized**. The observed covariances minus the predicted covariances form the residuals, and the fundamental hypothesis for these structural equation procedures is that the covariance matrix of the observed variables is a function of a set of parameters. If the model was correct and the parameters known, the population covariance matrix would be exactly reproduced. Hence, the fundamental hypothesis in SEM is

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) \tag{4.1}$$

$\boldsymbol{\Sigma}$ is the population covariance matrix of observed variables, $\boldsymbol{\theta}$ is a vector containing the model parameters, and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the covariance matrix written as a function of $\boldsymbol{\theta}$.

For instance, in a simple regression equation $y = \gamma x + \zeta$ considered in deviation form (i.e., $E(x) = 0$), in terms of (4.1), and assuming $\text{cor}(x, \zeta) = 0$, $E(\zeta) = 0$, we would write

$$\begin{bmatrix} \text{var}(y) & \\ \text{cov}(x, y) & \text{var}(x) \end{bmatrix} = \begin{bmatrix} \gamma^2 \text{var}(x) + \text{var}(\zeta) & \\ \gamma \text{var}(x) & \text{var}(x) \end{bmatrix} \tag{4.2}$$

In (4.2), $\boldsymbol{\theta} = (\gamma, \text{var}(\boldsymbol{\zeta}))$. Equation (4.1) implies that each element on the left-hand side equals its corresponding element on the right-hand side. Therefore, $\gamma = \frac{\text{cov}(x,y)}{\text{var}(x)}$, and $\text{var}(\zeta) = \text{var}(y) - \frac{\text{cov}(x,y)^2}{\text{var}(x)}$.

We could also consider the following system of equations:

$$\begin{aligned} y &= \gamma\xi + \zeta \\ x_1 &= \xi + \delta_1 \\ x_2 &= \xi + \delta_2 \end{aligned} \tag{4.3}$$

In the first equation, the predictor $\xi$ is unobserved, and $x_1$ and $x_2$ are indicators of the factor or latent variable $\xi$. Here it is assumed that $\zeta, \delta_1$ and $\delta_2$ are uncorrelated with $\xi$ and with each other, and that each has an expected value of zero. If we put both variance-covariance matrices equal as in (4.1) it results in

$$\begin{bmatrix} \text{var}(y) & & \\ \text{cov}(x_1, y) & \text{var}(x_1) & \\ \text{cov}(x_2, y) & \text{cov}(x_2, x_1) & \text{var}(x_2) \end{bmatrix} = \begin{bmatrix} \gamma^2 \text{var}(\xi) + \text{var}(\zeta) & & \\ \gamma \text{var}(\xi) & \text{var}(\xi) + \text{var}(\delta_1) & \\ \gamma \text{var}(\xi) & \text{var}(\xi) & \text{var}(\xi) + \text{var}(\delta_2) \end{bmatrix} \tag{4.4}$$

In (4.4), $\boldsymbol{\theta} = (\gamma, \text{var}(\xi), \text{var}(\zeta), \text{var}(\delta_1), \text{var}(\delta_2))$.

The system of simultaneous linear equations in (4.3) is a structural equation model.

## 4.2 Notation

Table 4.1 describes the notation used in this report. In the system of equations (4.3), the first equation would correspond to the **Latent variable model** and the other two to the **Measurement model**. The assumptions for the parameters of each model are also found in this table.

| Symbol | Dimension | Definition |
|---|---|---|
| $\boldsymbol{S}$ | $(p+q)\times(p+q)$ | Sample covariance matrix |
| $\boldsymbol{\Sigma}$ | $(p+q)\times(p+q)$ | Population covariance matrix |
| | | **Latent Variable model:** $\boldsymbol{\eta} = \boldsymbol{B\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\zeta}$ |
| $\boldsymbol{\eta}$ | $m \times 1$ | Latent endogenous[1] variables |
| $\boldsymbol{\xi}$ | $n \times 1$ | Latent exogenous[1] variables |
| $\boldsymbol{\zeta}$ | $m \times 1$ | Latent errors in equations |
| $\boldsymbol{B}$ | $m \times m$ | Coefficient matrix for latent endogenous[1] variables |
| $\boldsymbol{\Gamma}$ | $m \times n$ | Coefficient matrix for latent exogenous[1] variables |
| $\boldsymbol{\Phi}$ | $n \times n$ | Covariance matrix of $\boldsymbol{\xi} = E(\boldsymbol{\xi\xi}')$ |
| $\boldsymbol{\Psi}$ | $m \times m$ | Covariance matrix of $\boldsymbol{\zeta} = E(\boldsymbol{\zeta\zeta}')$ |
| **Assumptions**: $E(\boldsymbol{\eta}) = 0, E(\boldsymbol{\xi}) = 0, E(\boldsymbol{\zeta}) = 0, \boldsymbol{\zeta}$ uncorrelated with $\boldsymbol{\xi}, (\boldsymbol{I} - \boldsymbol{B})$ nonsingular. | | |
| | | **Mesurement model:** $\boldsymbol{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}; \boldsymbol{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}$ |
| $\boldsymbol{y}$ | $p \times 1$ | Observed indicators of $\boldsymbol{\eta}$ |
| $\boldsymbol{x}$ | $q \times 1$ | Observed indicators of $\boldsymbol{\xi}$ |
| $\boldsymbol{\epsilon}$ | $p \times 1$ | Measurement errors of $\boldsymbol{y}$ |
| $\boldsymbol{\delta}$ | $q \times 1$ | Measurement errors of $\boldsymbol{x}$ |
| $\boldsymbol{\Lambda}_y$ | $p \times m$ | Coefficients of the regression of $\boldsymbol{y}$ on $\boldsymbol{\eta}$ or factor loadings |
| $\boldsymbol{\Lambda}_x$ | $q \times n$ | Coefficients of the regression of $\boldsymbol{x}$ on $\boldsymbol{\xi}$ or factor loadings |
| $\boldsymbol{\Theta}_\epsilon$ | $p \times p$ | Covariance matrix of $\boldsymbol{\epsilon} = E(\boldsymbol{\epsilon\epsilon}')$ |
| $\boldsymbol{\Theta}_\delta$ | $q \times q$ | Covariance matrix of $\boldsymbol{\delta} = E(\boldsymbol{\delta\delta}')$ |
| **Assumptions**: $E(\boldsymbol{\eta}) = 0, E(\boldsymbol{\xi}) = 0, E(\boldsymbol{\epsilon}) = 0, E(\boldsymbol{\delta}) = 0,$ $\boldsymbol{\epsilon}$ uncorrelated with $\boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\zeta}$, and $\boldsymbol{\delta}$, $\boldsymbol{\delta}$ uncorrelated with $\boldsymbol{\eta}, \boldsymbol{\xi}$, and $\boldsymbol{\zeta}$ | | |

[1] These terms will be defined in Section 4.5. Briefly, an endogenous variable is determined by the model (i.e., by the relationships among the variables), while an exogenous variable is determined by factors lying outside the model.

Table 4.1: Notation and assumptions for model parameters

## 4.3 Model specification

The first component of the structural equations is the latent variable model:

$$\text{Latent variable model: } \boldsymbol{\eta} = \boldsymbol{B\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\zeta} \tag{4.5}$$

The second component of the general system is the measurement model:

$$\text{Measurement model: } \begin{aligned} \boldsymbol{y} &= \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon} \\ \boldsymbol{x} &= \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \end{aligned} \tag{4.6}$$

Notice that no intercept terms appear in the equations: it is assumed that explantory variables are deviated from their means in order to simplify the discussion. In the whole project this assumption will be done; in Appendix B the model with intercepts is presented.

The measurement error is the deviation of the outcome of a measurement from the true value. In the case of Patient Reported Outcomes, it refers to all that detaches the true response from the observed one. The vectors $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ contain the measurement errors of the indicators $\boldsymbol{y}$ and $\boldsymbol{x}$, respectively.

Generally, using observed variables leads to inconsistent estimators and to inaccurate assessments of the relation between the underlying latent variables defined in (4.5) (Bollen, 1989). To correct these problems, we need to understand the process of measurement by incorporating the relation between the observed variables and latent variables into structural equation models.

Measurement is the process by which a concept is linked to one or more latent variables; in fact, the latter are linked to observed variables by means of the equations in (4.6). The concept can vary from a highly abstract one (intelligence) to a more concrete one (age). One or several latent variables may be needed to represent the concept. The observed variables can be responses to questionnaire items, census figures, meter readings, etc. A concept is an idea that unites phenomena under a single term. Anger, for instance, gathers characteristics such as screaming, throwing objects, having a blood-flushed face, among others. The concept of anger acts as a summarizing device to replace a list of specific traits that an individual may exhibit. Latent variables are the representations of concepts. Once the concept is devised, the four steps in the measurement process are:

1. Give the meaning of/Define theoretically the concept and limit its dimensions.
2. Identify the dimensions and latent variables to represent it. Dimensions are the distinct aspects of a concept, components that cannot easily be subdivided into additional components.
3. Form measures of the latent variables (for instance, responses to questionnaire items).
4. Specify the relation between the indicators and the latent variables (i.e., to construct the measurement model).

## 4.4 The General SEM Model

As defined in Section 4.3, a General SEM model consists of the latent and measurement models:

$$\begin{aligned} \boldsymbol{\eta} &= \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \\ \boldsymbol{y} &= \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \epsilon \\ \boldsymbol{x} &= \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \end{aligned}$$

The implied covariance matrix is

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta}) = E(\boldsymbol{y}\boldsymbol{y}') & \boldsymbol{\Sigma}_{yx}(\boldsymbol{\theta}) = E(\boldsymbol{y}\boldsymbol{x}') \\ \boldsymbol{\Sigma}_{xy}(\boldsymbol{\theta}) = E(\boldsymbol{x}\boldsymbol{y}') & \boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}) = E(\boldsymbol{x}\boldsymbol{x}') \end{bmatrix}$$

Therefore, $\boldsymbol{\Sigma}_{yy}$, $\boldsymbol{\Sigma}_{yx}$, and $\boldsymbol{\Sigma}_{xx}$, the components of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, must be computed. Taking into account the assumptions in Table 4.1 —$\boldsymbol{\zeta}$ uncorrelated with $\boldsymbol{\xi}$, $\boldsymbol{\epsilon}$, and $\boldsymbol{\delta}$, $\boldsymbol{\epsilon}$ uncorrelated with $\boldsymbol{\eta}$, $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$, and $\boldsymbol{\delta}$ uncorrelated with $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$— the components of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be deduced as follows:

First,

$$\begin{aligned} E(\boldsymbol{y}\boldsymbol{y}') &= E\left((\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\eta}'\boldsymbol{\Lambda}_y' + \boldsymbol{\epsilon}')\right) = \boldsymbol{\Lambda}_y E(\boldsymbol{\eta}\boldsymbol{\eta}')\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon \underbrace{=}_{\boldsymbol{\eta}=(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})} \\ &= \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}E\left((\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})(\boldsymbol{\xi}'\boldsymbol{\Gamma}' + \boldsymbol{\zeta}')\right)[(\boldsymbol{I}-\boldsymbol{B})^{-1}]'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon \\ &= \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\left(\boldsymbol{\Gamma}E(\boldsymbol{\xi}\boldsymbol{\xi}')\boldsymbol{\Gamma}' + E(\boldsymbol{\zeta}\boldsymbol{\zeta}')\right)[(\boldsymbol{I}-\boldsymbol{B})^{-1}]'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon \\ &= \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})[(\boldsymbol{I}-\boldsymbol{B})^{-1}]'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon. \end{aligned}$$

Second,

$$E(\boldsymbol{y}\boldsymbol{x}') = E\left((\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\xi}'\boldsymbol{\Lambda}_x' + \boldsymbol{\delta}')\right) = \boldsymbol{\Lambda}_y E(\boldsymbol{\eta}\boldsymbol{\xi}')\boldsymbol{\Lambda}_x' \underbrace{=}_{\boldsymbol{\eta}=(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})} \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x'.$$

Third,

$$E(\boldsymbol{x}\boldsymbol{x}') = E((\boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\xi}'\boldsymbol{\Lambda}_x' + \boldsymbol{\delta}')) = \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta.$$

Therefore,

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})[(\boldsymbol{I}-\boldsymbol{B})^{-1}]'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon & \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' \\ \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Gamma}'\left[(\boldsymbol{I}-\boldsymbol{B})^{-1}\right]'\boldsymbol{\Lambda}_y' & \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta \end{bmatrix} \quad (4.7)$$

Once the covariance matrix is specified, the next step is to identify the model, namely to write the unknown parameters of $\boldsymbol{\theta}$ as a function of one or more known elements in $\boldsymbol{\Sigma}$. A model is identified if all unknown parameters in $\boldsymbol{\theta}$ are identified, or, alternatively, if no vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ exist such that $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$, unless $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. As should be evident, model identifcation in SEM with observed variables is not possible without placing restrictions on model parameters.

In practice we do not know either the population covariances and variances or the parameters. The task is to form sample estimates of the unknown parameters based on sample estimates of the covariance matrix. Let $\boldsymbol{S}$ be the sample estimate of the covariance matrix and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ the implied covariance matrix. The unknown parameters in $\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}_y, \boldsymbol{\Lambda}_x, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Theta}_\epsilon$, and $\boldsymbol{\Theta}_\delta$ are estimated so that $\hat{\boldsymbol{\Sigma}}$ is close to the sample covariance matrix $\boldsymbol{S}$. *Close* must be defined —that is, we require a function to be minimized (fitting function). In Section 4.8 the fitting functions used in the analyses are defined.

### 4.4.1 SEM with observed variables

A special case of the general structural equation procedures with latent variables is SEM with only observed variables. Performing SEM with observed variables assumes that each variable is a perfect measure of its corresponding latent variable, i.e., $\boldsymbol{y} = \boldsymbol{\eta}$ (and consequently $\boldsymbol{\epsilon} = 0$), and $\boldsymbol{x} = \boldsymbol{\xi}$ (and consequently $\boldsymbol{\delta} = 0$). In that way, according to Table 4.1, $\boldsymbol{\Phi} = E(\boldsymbol{x}\boldsymbol{x}')$, $\boldsymbol{\Lambda}_y = \boldsymbol{I}_m$, $\boldsymbol{\Lambda}_x = \boldsymbol{I}_n$, $\boldsymbol{\Theta}_\epsilon = 0$, and $\boldsymbol{\Theta}_\delta = 0$. Therefore, there is no measurement model and the structural equation model will correspond to the latent equation (4.5):

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{y} + \boldsymbol{\Gamma}\boldsymbol{x} + \boldsymbol{\zeta} \tag{4.8}$$

Hence, taking into account (4.7), the implied covariance matrix for SEM with observed variables is:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} (\boldsymbol{I} - \boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})[(\boldsymbol{I} - \boldsymbol{B})^{-1}]' & (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}'[(\boldsymbol{I} - \boldsymbol{B})^{-1}]' & \boldsymbol{\Phi} \end{bmatrix} \tag{4.9}$$

## 4.5 Path diagrams

A *Path Diagram* is a pictoral representation of a system of simultaneous equations: it represents the relationships that the equations are assumed to hold. As it is shown in Figure 4.1, the observed variables are enclosed in boxes; the latent variables are circled, with the exception of the disturbance terms which are not enclosed. Straight single-headed arrows represent causal relations between the variables connected by the arrows. A curved two-headed arrow indicates an association between two variables.

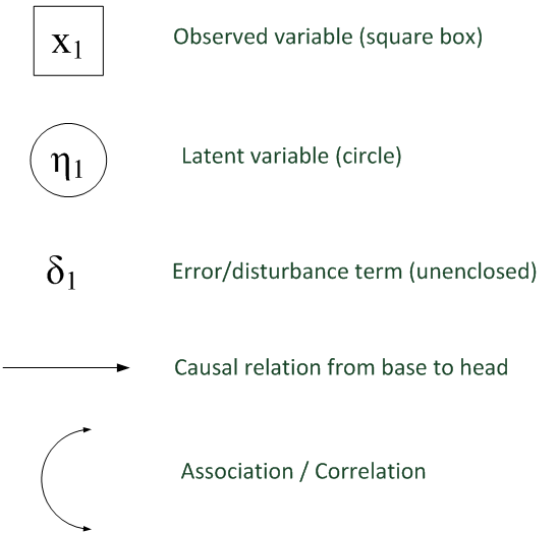



Figure 4.1: Symbols used in Path Diagrams

In fact, path diagrammatic notation has a one-to-one equivalence with the matrix formulation by means of the so-called Reticular Action Model (RAM) as it is explained in McArdle and McDonald (1984). The graphs which will be represented corresponding to the matricial models are RAM notation conventions. Indeed, a model is equivalent to a graph; the nodes represent the variables, and there are two types of them: observed (indicated with a square) and latent (indicated with a circle). The edges represent causal relationships or associations.

### 4.5.1 Case 1: Pure measurement model (Confirmatory Factor Analysis, CFA)

An example of a confirmatory factor analysis is shown in Figure 4.2. It corresponds to a two-factor model. It is strictly a measurement model because the paths only describe the link between the latent variables ($\xi_1$ and $\xi_2$) and their indicators ($x_1, x_2$ and $x_3$ for $\xi_1$, and $x_4$, $x_5$, $x_6$, and $x_7$ for $\xi_2$), and there are no causal relationships between latent variables (the curved arrow representing an association is not a causal relationship). Note

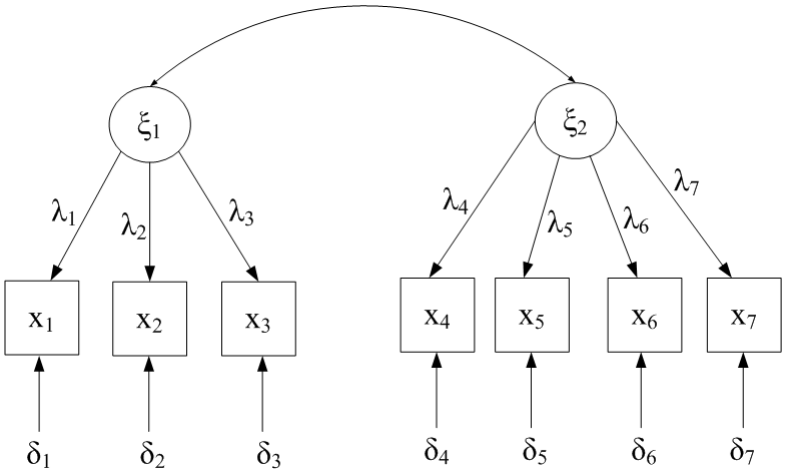



Figure 4.2: Pure measurement model or CFA

that this path diagram corresponds to the second equation in (4.6), and the equivalent matrix formulation is the following:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \xi_1 + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}$$

$$\begin{bmatrix} x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \end{bmatrix} \xi_2 + \begin{bmatrix} \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \end{bmatrix}$$

### 4.5.2 Case 2: Pure structural model

In contrast, Figure 4.3 describes a pure structural model because only the relationships among variables are displayed; indeed, there is no measurement model because no latent variables appear. It is considered that the observed variables are measured without error as in (4.8).

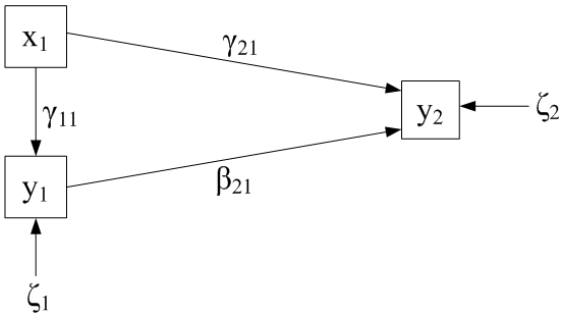



Figure 4.3: Pure structural model

The equivalence matrix formulation would be:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} x_1 + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

### 4.5.3 Case 3: Full structural model

Figure 4.4 shows an example of a full structural model representation, because it includes both measurement and latent models. It is found in Bollen (1989) and it shows the relationship of political democracy ($\eta$) to industrialization ($\xi_1$) in developing countries.

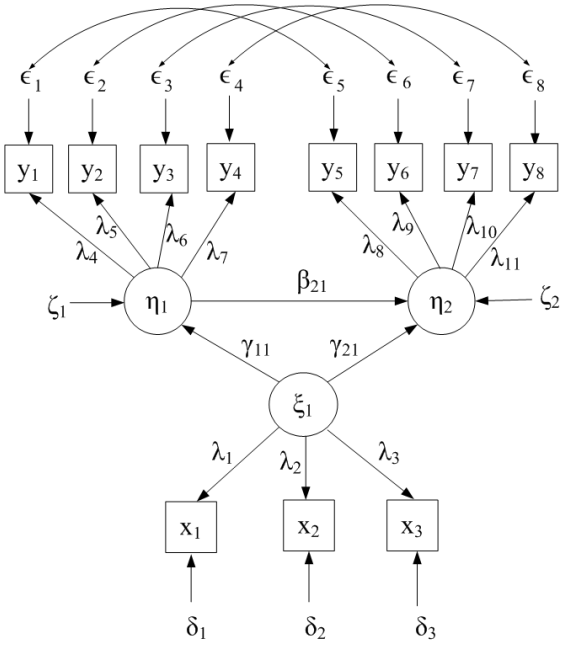



Figure 4.4: Full structural model

Political democracy refers to the extent of political rights and political liberties in a country. Industrialization is the degree to which a society's economy is characterized by mechanized

manufacturing processes. In this model there are three latent random variables: political democracy in 1960 ($\eta_1$) and in 1965 ($\eta_2$), and industrialization in 1960 ($\xi_1$); it is assumed that political democracy in 1965 ($\eta_2$) is a function of 1960 political democracy ($\eta_1$) and industrialization ($\xi_1$). The 1960 industrialization level also affects the 1960 political democracy.

In SEM, there are not *independent* and *dependent* variables, but *exogenous* and *endogenous* variables. Industrialization ($\xi_1$) is exogenous because its causes lie outside the model; thus, $\xi_1$ has paths coming from it and none leading to it (we don't count the curved arrows because they're simply describing correlations among the variables and aren't considered to be paths). The political democracy variables ($\eta_1$ and $\eta_2$) are endogenous because they are determined by the variables within the model; thus, $\eta_1$ and $\eta_2$ have at least one path leading to them. Note also that all endogenous variables have an error/disturbance term tacked on ($\zeta_1$ and $\zeta_2$), which corresponds to the unexplained part of the model.

The previous paragraphs refer to the latent model. The measurement model has equations representing the link between the latent and observed variables. There are three indicators of industrialization in 1960: gross national product per capita ($x_1$), inanimate energy consumption ($x_2$), and the percentage of the labor force in industry ($x_3$). For political democracy there are the same four indicators for 1960 and 1965: expert ratings of the freedom of the press ($y_1$ in 1960, $y_5$ in 1965), the freedom of political opposition ($y_2$ and $y_6$), the fairness of elections ($y_3$ and $y_7$), and the effectiveness of the elected legislature ($y_4$ and $y_8$). Hence, each latent variable is measured with several observed variables.

This structural model corresponds to the following simultaneous system of equations:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} \xi_1 + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} \lambda_4 & 0 \\ \lambda_5 & 0 \\ \lambda_6 & 0 \\ \lambda_7 & 0 \\ 0 & \lambda_8 \\ 0 & \lambda_9 \\ 0 & \lambda_{10} \\ 0 & \lambda_{11} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \xi_1 + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}$$

The $\beta_{21}$ coefficient indicates the change in the expected value of $\eta_2$ after a one-unit increase in $\eta_1$ holding $\xi_1$ constant. The $\gamma_{11}$ and $\gamma_{21}$ coefficients have analogous interpretations. $\beta_{21}$ is associated with the latent endogenous variable $\eta_1$, whereas $\gamma_{11}$ and $\gamma_{21}$ refer to the latent exogenous variable $\xi_1$.

The $\lambda_i$ coefficients are the magnitude of the expected change in the observed variable for a one unit change in the latent variable, namely the effects of the latent variables on the observed variables. They are called *factor loadings*. The $\delta_i$ and $\epsilon_i$ variables are the errors of measurement for $x_i$ and $y_i$, respectively. They are disturbances that disrupt the relation between the latent and observed variables. They are assumed to be uncorrelated with each other, to have an expected value of zero, and to be uncorrelated with all $\xi$'s, $\eta$'s, and $\zeta$'s. Similarly, $\zeta$'s have an expected value of zero and are uncorrelated with the exogenous variable $\xi$. As mentioned before, $\zeta_i$ includes those variables that influence $\eta_i$ but are excluded from the $\eta_i$ equation.

It is assumed that the errors of measurement for the indicators of industrialization ($x_1$ to $x_3$) are uncorrelated. However, as we have the same set of indicators at two points in time for political democracy, it is likely that the error in measuring an indicator in 1960 is correlated with the error in measuring the same indicator in 1965. That is why there are curved arrows from $\epsilon_i$ to $\epsilon_{i+4}$.

Note that all the observed variables depend on the latent variables, so it is assumed that latent variables *cause* indicators.

This example reveals some of the major features of structural equations with latent variables that are distinct from the standard regression approach. The models are more realistic in their allowance for measurement error in the observed variables. They allow random measurement error in $\epsilon$ and $\delta$, and systematic differences in scale are introduced with the $\lambda$ coefficients. The error in measuring one variable can correlate with that of another. Multiple indicators can measure one latent variable. Furthermore, the relation between latent variables can be analyzed unobscured by measurement error. All of these features bring us closer to testing the hypotheses set forth in theories.

## 4.6 Mediation

As mentioned in Chapter 1, a given variable may be said to function as a mediator to the extent that it accounts for the relation between a predictor and an outcome (Baron and Kenny, 1986).

### 4.6.1 Decomposition of the causal effects

Mediation models distinguish three types of effects: direct, indirect and total effects. The **direct** effect is that influence of one variable on another that is unmediated by any other variable in a path model, i.e., the effect going from node A to node B without passing through any other node. The **indirect** effects of a variable are mediated by at least one intervening variable, i.e, they are the effects going from A to B passing through at least one another node C. The sum of the direct and indirect effects are the **total** effects:

$$\text{Total effects=Direct effect + Indirect effects} \tag{4.10}$$

To illustrate these types of effects, consider again Figure 4.4. An example of a direct effect is the effect of $\eta_1$ on $\eta_2$, that is, $\beta_{21}$. There are no mediating variables between $\eta_1$ and $\eta_2$. The direct effect of $\xi_1$ on $\eta_2$ is $\gamma_{21}$, and $\lambda_8$ is the direct effect of $\eta_2$ on $y_5$.

Regarding indirect effects, consider the influence of $\xi_1$ on $\eta_2$. The intervening variable in this case is $\eta_1$. A one unit change in $\xi_1$ leads to an expected $\gamma_{11}$ change in $\eta_1$. This $\gamma_{11}$ change in $\eta_1$ leads to an expected $\beta_{21}$ change in $\eta_2$. Thus the indirect effect of $\xi_1$ on $\eta_2$ is $\gamma_{11}\beta_{21}$. Following a similar procedure the indirect effect of $\eta_1$ on $y_7$ is $\beta_{21}\lambda_{10}$.

Concerning total effects, from (4.10), we deduce that the total effect of $\xi_1$ on $\eta_2$ is $\gamma_{21}$ (direct effect) $+\gamma_{11}\beta_{21}$ (indirect effects). On the other hand, the total effect of $\xi_1$ on $y_8$ is $0 + (\gamma_{21}\lambda_{11} + \gamma_{11}\beta_{21}\lambda_{11})$. Note that $\xi_1$ has no direct effect on $y_8$.

The coefficient matrices in the structural equations (4.5) and (4.6) are the direct effects. For instance, (4.5) shows the direct effects of $\boldsymbol{\xi}$ on $\boldsymbol{\eta}$ as $\boldsymbol{\Gamma}$.

In order to obtain indirect and total effects, we write equations (4.5) and (4.6) in *reduced form*, i.e., all endogenous variables are written as functions of only exogenous variables (Bollen, 1987). The coefficients of the exogenous variables correspond to the total effect on the endogenous variable.

$$\boldsymbol{\eta} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\xi} + (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\zeta} \tag{4.11}$$

The total effects of $\boldsymbol{\xi}$ on $\boldsymbol{\eta}$ ($T_{\boldsymbol{\eta\xi}}$) are $(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}$. Although in Figure 4.4 there are only direct effects between $\eta$'s, there are situations where there can also be indirect effects, as it shows Figure 4.5, which displays a simple model consisting of three endogenous variables.

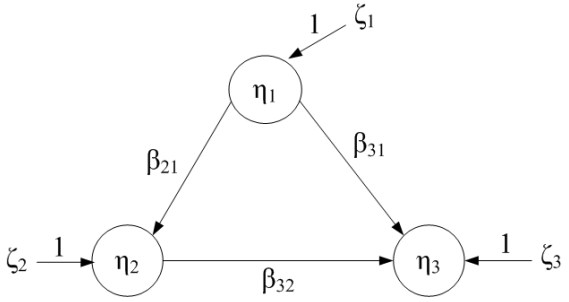



Figure 4.5: Structural model for three endogenous variables

Hence, total effects of $\boldsymbol{\eta}$ on $\boldsymbol{\eta}$ are not just the direct effects $\boldsymbol{B}$. To obtain them, consider the total effects of $\boldsymbol{\zeta}$, $(\boldsymbol{I}-\boldsymbol{B})^{-1}$. From (4.5) we deduce that the direct effect of $\boldsymbol{\zeta}$ on $\boldsymbol{\eta}$ is $\boldsymbol{I}$. Since $\boldsymbol{\zeta}$ affects directly only variables comprising $\boldsymbol{\eta}$, all its indirect effects must be mediated by $\boldsymbol{\eta}$. Note that all the effects of $\boldsymbol{\eta}$ on $\boldsymbol{\eta}$ are included in the indirect effects of $\boldsymbol{\zeta}$ on $\boldsymbol{\eta}$. For instance, in Figure 4.5, the total effect of $\eta_1$ on $\eta_3$ equals $T_{\eta_3\eta_1} = \beta_{31} + \beta_{21}\beta_{32} = I_{\eta_3\eta_1\zeta_1} + I_{\eta_3\eta_2\eta_1\zeta_1}$, where $I_{v_1\ldots v_2}$ represents the indirect effect of $v_2$ on $v_1$ through the ... intervening variables. Thus, the indirect effects of $\boldsymbol{\zeta}$ on $\boldsymbol{\eta}$ equal the total effects of $\boldsymbol{\eta}$ on $\boldsymbol{\eta}$.

As shown in (4.5), the direct effects of $\boldsymbol{\zeta}$ on $\boldsymbol{\eta}$ equal $I$. As the indirect effects correspond to the difference between the total and direct effects, the indirect effects of $\boldsymbol{\zeta}$ on $\boldsymbol{\eta}$ (i.e., the total effects of $\boldsymbol{\eta}$ on $\boldsymbol{\eta}$) equal $(\boldsymbol{I}-\boldsymbol{B})^{-1} - \boldsymbol{I}$.

The reduced form of the measurement model for $\boldsymbol{y}$ is

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\zeta} + \boldsymbol{\epsilon} \tag{4.12}$$

The total effects $T_{\boldsymbol{y\xi}}$ correspond to the coefficient for $\boldsymbol{\xi}$, $\boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}$. As $\boldsymbol{\xi}$ has no direct effect on $\boldsymbol{y}$, $T_{\boldsymbol{y\xi}} = I_{\boldsymbol{y\xi}}$.

The total effects of $\boldsymbol{\eta}$ on $\boldsymbol{y}$ are also deduced through $\boldsymbol{\zeta}$. Since $\boldsymbol{\zeta}$ has no influence on $\boldsymbol{y}$ unmediated by $\boldsymbol{\eta}$, the direct effects of $\boldsymbol{\zeta}$ on $\boldsymbol{y}$ are 0. The indirect effects of $\boldsymbol{\zeta}$ on $\boldsymbol{y}$ (which equal $T_{\boldsymbol{y\zeta}}$) are all the influences that $\boldsymbol{\eta}$ exerts on $\boldsymbol{y}$ (see Figure 4.4). Therefore, $T_{\boldsymbol{y\eta}} = T_{\boldsymbol{y\zeta}} = \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}$, the coefficient for $\boldsymbol{\zeta}$ in (4.12).

Nevertheless, in order to be able to estimate $T_{\boldsymbol{\eta\eta}}$, the modulus or absolute value of the largest eigenvalue of $\boldsymbol{B}$ must be less than 1 (Bentler and Freeman, 1983).

Table 4.2 contains the expressions deduced above for the decomposition of effects.

| | $\boldsymbol{\xi} \longrightarrow \boldsymbol{\eta}$ | $\boldsymbol{\eta} \longrightarrow \boldsymbol{\eta}$ |
|---|---|---|
| Direct | $\boldsymbol{\Gamma}$ | $\boldsymbol{B}$ |
| Indirect | $(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}-\boldsymbol{\Gamma}$ | $(\boldsymbol{I}-\boldsymbol{B})^{-1}-\boldsymbol{I}-\boldsymbol{B}$ |
| Total | $(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}$ | $(\boldsymbol{I}-\boldsymbol{B})^{-1}-\boldsymbol{I}$ |
| | $\boldsymbol{\xi} \longrightarrow \boldsymbol{y}$ | $\boldsymbol{\eta} \longrightarrow \boldsymbol{y}$ |
| Direct | 0 | $\boldsymbol{\Lambda}_y$ |
| Indirect | $\boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}$ | $\boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}-\boldsymbol{\Lambda}_y$ |
| Total | $\boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}$ | $\boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}$ |

Table 4.2: Formula to obtain Direct, Indirect and Total Effects

### 4.6.2 Specific indirect effects

The indirect effects comprise all of the indirect paths from one variable to another. Sometimes it can be of interest to analyze those effects transmitted by a particular variable: the **specific indirect effects**. For example, suppose that we want to estimate all of the specific indirect effects of $\boldsymbol{x}$ on $\boldsymbol{y}$ through $y_1$ in Figure 4.6 (Bollen, 1987).

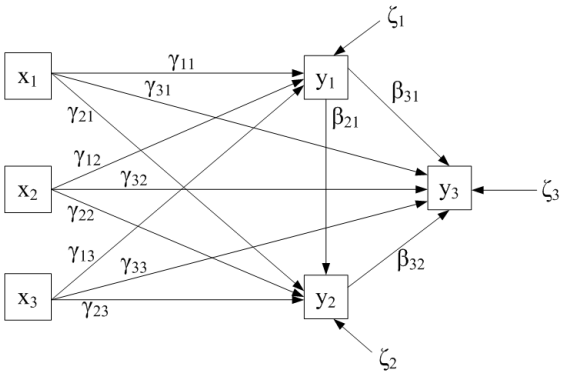



Figure 4.6: Structural model to illustrate specific indirect effects extracted from Alwin and Hauser (1975)

The coefficient matrices $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ for this example are

$$\boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \quad \boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \tag{4.13}$$

If paths through $y_1$ (i.e., paths involving $\gamma_{1i}$ and $\beta_{i1}$) were eliminated and if we calculated the resulting decomposition, we would know the decomposition of effect *not* due to $y_1$ but to the remaining variables ($y_2$). Matrices $\boldsymbol{B}_{c_1}$ and $\boldsymbol{\Gamma}_{r_1}$ in (4.14) correspond to coefficient

matrices from (4.13) but with coefficients involving paths through $y_1$ set to 0. Subindices $c_1$ and $r_1$ stand for *zeros in column 1*, and *zeros in row 1*, respectively.

$$\boldsymbol{B}_{c_1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \beta_{32} & 0 \end{bmatrix} \quad \boldsymbol{\Gamma}_{r_1} = \begin{bmatrix} 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \tag{4.14}$$

From Table 4.2 we know that the indirect effects of $\boldsymbol{x}$ on $\boldsymbol{y}$ are $I_{\boldsymbol{\eta\xi}} = I_{\boldsymbol{yx}} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma} - \boldsymbol{\Gamma}$. If we substract the modified indirect effects $I_{\boldsymbol{yx}_{y_1}} = (\boldsymbol{I} - \boldsymbol{B}_{c_1})^{-1}\boldsymbol{\Gamma}_{r_1} - \boldsymbol{\Gamma}_{r_1}$ from original ones ($I_{\boldsymbol{yx}}$), we would know the specific indirect effects through $y_1$, $SI_{y_1}$, the quantity desired:

$$I_{\boldsymbol{yx}} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21}\gamma_{11} & \beta_{21}\gamma_{12} & \beta_{21}\gamma_{13} \\ (\beta_{31} + \beta_{21}\beta_{32})\gamma_{11} + \beta_{32}\gamma_{21} & (\beta_{31} + \beta_{21}\beta_{32})\gamma_{12} + \beta_{32}\gamma_{22} & (\beta_{31} + \beta_{21}\beta_{32})\gamma_{13} + \beta_{32}\gamma_{23} \end{bmatrix}$$

$$I_{\boldsymbol{yx}_{y_1}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{32}\gamma_{21} & \beta_{32}\gamma_{22} & \beta_{32}\gamma_{23} \end{bmatrix}$$

$$SI_{y_1} = I_{\boldsymbol{yx}} - I_{\boldsymbol{yx}_{y_1}} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21}\gamma_{11} & \beta_{21}\gamma_{12} & \beta_{21}\gamma_{13} \\ (\beta_{31} + \beta_{21}\beta_{32})\gamma_{11} & (\beta_{31} + \beta_{21}\beta_{32})\gamma_{12} & (\beta_{31} + \beta_{21}\beta_{32})\gamma_{13} \end{bmatrix}$$

Considering each type of effects leads to a more complete understanding of the relation between variables than if these distinctions are not made. In the typical regression analysis, the regression coefficient is an estimate of the direct effect of a variable. If we ignore the indirect effects that a variable may have through other variables, we may be grossly off in the assessment of its overall effect.

## 4.7 Complex Disability Mediated Models (CDMM) on WMH Data

We would like to assess the effect of 19 common conditions on perceived health and to what extent the 8 disability dimensions mediate these effects. As mentioned in Chapter 1, there are $19 \cdot 8 = 152$ simple mediation models, which considered together make up multiple mediation paths enabling to explain the simultaneous effects of the predictors on the outcome. In this Section, once explained the SEM notation, model specification, path diagrams, and mediation, we are well placed to present formally the structural equation model applied to our WMH database. In the following sections I will define, with path diagrams and matrices, the Complex Disability Mediated Models CDMM-O and CDMM-L.

### 4.7.1 CDMM-O

I first consider the situation in which it is assumed that each variable is a perfect measure of its corresponding latent variable, i.e., $\boldsymbol{y} = \boldsymbol{\eta}$ and $\boldsymbol{x} = \boldsymbol{\xi}$. Therefore, the model fitted is the one from (4.8), $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{y} + \boldsymbol{\Gamma}\boldsymbol{x} + \boldsymbol{\zeta}$, and the errors of measurement $\boldsymbol{\delta}$ and $\boldsymbol{\epsilon}$ are 0. The factor loadings $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are $\boldsymbol{I}_m$ and $\boldsymbol{I}_n$, respectively.

Table 3.1 contains the variables included in the model. Figure 4.7 displays the model with only observed variables.

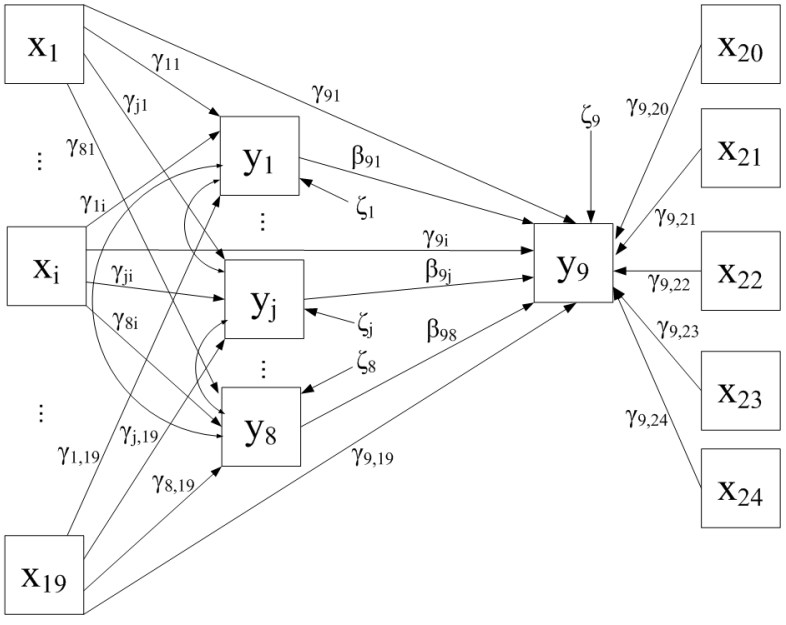



Figure 4.7: SEM application with observed variables

The variables appearing in the analysis are described as follows:

- $x_i$ are exogenous variables. $x_1$ to $x_{19}$ represent the 19 conditions, and each of them has a direct effect on all the $y$'s. $x_{20}$ to $x_{24}$ are the covariates age, country, employment status, marital status, and sex, and they only have a direct effect on $y_9$, the perceived health. Note that all the $x$'s have paths going out from them and none leading to them. According to Table 4.1, $n = q = 24$.

- $y_j$ are endogenous variables. $y_1$ to $y_8$ represent the 8 disability dimensions; they are the mediators of the conditions. Note that the correlation among them is taken into account. They have paths leading to them (from $\{x_1 \dots x_{19}\}$) and also coming from them and leading to $y_9$, the perceived health. In fact, they are mediators between $\{x_1 \dots x_{19}\}$ and $y_9$ because of having both types of paths. In turn, $y_9$ has only paths leading to it from all of the variables. According to Table 4.1, $m = p = 9$.

The model equations would be:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ \beta_{91} & \dots & \beta_{98} & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \dots & \gamma_{119} & 0 & \dots & 0 \\ \gamma_{21} & \dots & \gamma_{219} & 0 & \dots & 0 \\ \gamma_{31} & \dots & \gamma_{319} & 0 & \dots & 0 \\ \gamma_{41} & \dots & \gamma_{419} & 0 & \dots & 0 \\ \gamma_{51} & \dots & \gamma_{519} & 0 & \dots & 0 \\ \gamma_{61} & \dots & \gamma_{619} & 0 & \dots & 0 \\ \gamma_{71} & \dots & \gamma_{719} & 0 & \dots & 0 \\ \gamma_{81} & \dots & \gamma_{819} & 0 & \dots & 0 \\ \gamma_{91} & \dots & \gamma_{919} & \gamma_{920} & \dots & \gamma_{924} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{19} \\ x_{20} \\ \vdots \\ x_{24} \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \\ \zeta_6 \\ \zeta_7 \\ \zeta_8 \\ \zeta_9 \end{bmatrix}$$

### 4.7.2 CDMM-L

In this Section, variables $y_1$ to $y_4$ are considered to be measured with measurement error. Therefore, 4 latent variables $\eta_1$ to $\eta_4$ are also included in the model. The structural equation model fitted in this Section corresponds to equations (4.5) and (4.6).

$$\begin{aligned} \boldsymbol{\eta} &= B\boldsymbol{\eta} + \Gamma\boldsymbol{\xi} + \boldsymbol{\zeta} \\ \boldsymbol{y} &= \Lambda_y\boldsymbol{\eta} + \epsilon \\ \boldsymbol{x} &= \Lambda_x\boldsymbol{\xi} + \boldsymbol{\delta} \end{aligned}$$

All the $x$'s are considered to be perfect measures of the underlying concept, so $\boldsymbol{\delta} = 0$. In turn, $\Lambda_x = I_n = I_q = I_{24}$.

Figure 4.8 displays the model with latent variables. All the $x$'s and $\{y_5, y_6, y_7, y_8, y_9\}$ are defined as in the previous Section. $\{y_{11}, \dots y_{14}\}$ are indicators of $\eta_1$, $\{y_{21} \dots y_{25}\}$ of $\eta_2$,

$\{y_{31}, \ldots y_{33}\}$ of $\eta_3$, and $\{y_{41} \ldots y_{43}\}$ of $\eta_4$. Thus, the mediating variables are not the previous $y_1$ to $y_4$ but $\eta_1$ to $\eta_4$. Therefore, in this case $m = 9$ but $p = 20$. Note that the correlation among mediators is also taken into account.

However, before building CDMM-L, the measurement model for the latent variables and their indicators must be tested. Hence, a Confirmatory Factor Analysis with the variables $\eta_1, y_{11}, \ldots, y_{14}, \eta_2, y_{21}, \ldots, y_{25}, \eta_3, y_{31}, \ldots, y_{33}$, and $\eta_4, y_{41}, \ldots, y_{43}$ must be carried out.

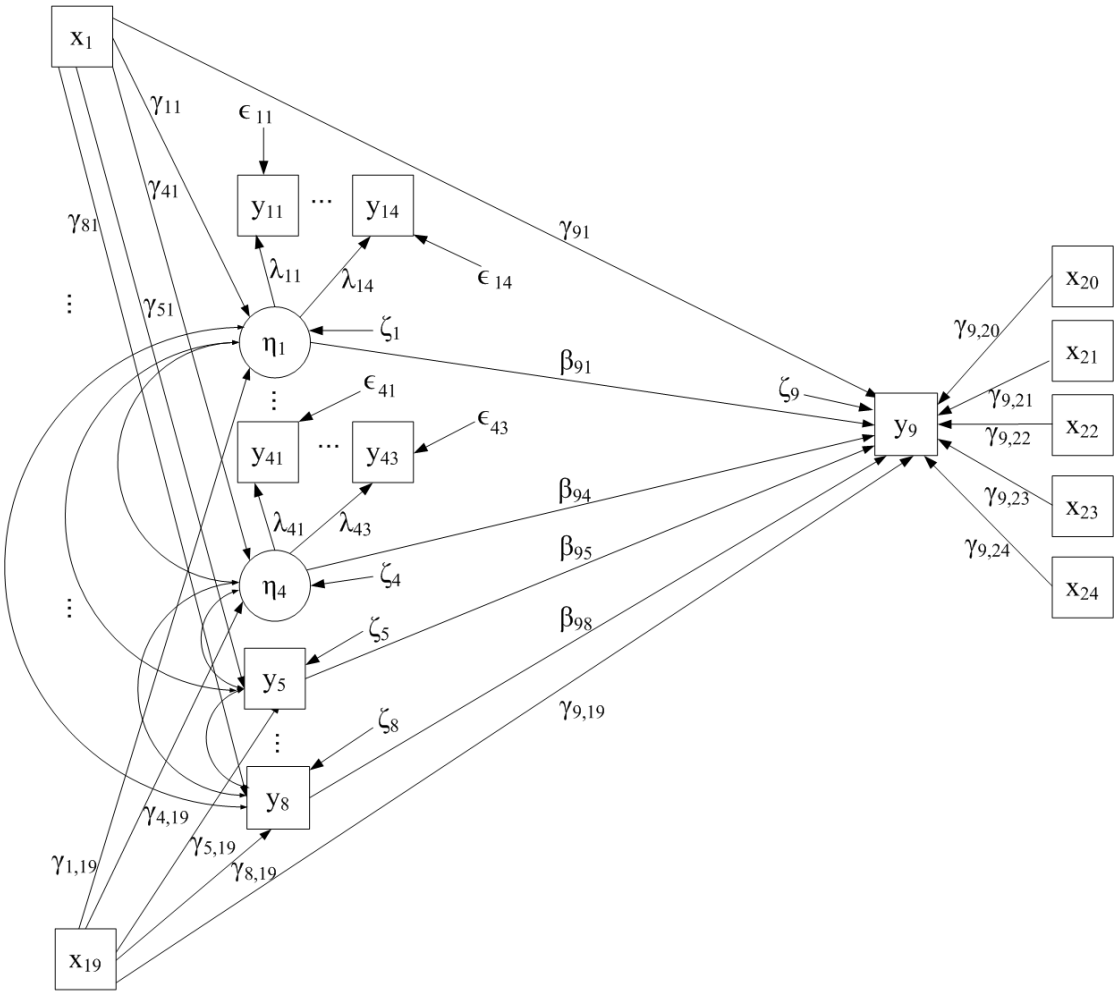



Figure 4.8: SEM application with latent variables

The model equations would be:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \\ \eta_9 \end{bmatrix} = \begin{bmatrix} 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ \beta_{91} & \dots & \beta_{98} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \\ \eta_9 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \dots & \gamma_{119} & 0 & \dots & 0 \\ \gamma_{21} & \dots & \gamma_{219} & 0 & \dots & 0 \\ \gamma_{31} & \dots & \gamma_{319} & 0 & \dots & 0 \\ \gamma_{41} & \dots & \gamma_{419} & 0 & \dots & 0 \\ \gamma_{51} & \dots & \gamma_{519} & 0 & \dots & 0 \\ \gamma_{61} & \dots & \gamma_{619} & 0 & \dots & 0 \\ \gamma_{71} & \dots & \gamma_{719} & 0 & \dots & 0 \\ \gamma_{81} & \dots & \gamma_{819} & 0 & \dots & 0 \\ \gamma_{91} & \dots & \gamma_{919} & \gamma_{920} & \dots & \gamma_{924} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_{19} \\ \xi_{20} \\ \vdots \\ \xi_{24} \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \\ \zeta_6 \\ \zeta_7 \\ \zeta_8 \\ \zeta_9 \end{bmatrix}$$

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{22} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{25} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{32} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{33} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{41} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{42} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{43} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \\ \eta_9 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{25} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_{19} \\ x_{20} \\ \vdots \\ x_{24} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & & & \dots & 0 \\ 0 & \ddots & & & & & 0 \\ \vdots & & \ddots & & & & \vdots \\ & & & 1 & & & \\ & & & & 1 & & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \dots & & & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_{19} \\ \xi_{20} \\ \vdots \\ \xi_{24} \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Notice that the Identity matrix in this last system of equations involves that the explanatory variables are measured without error.

#### 4.7.2.1 Polychoric correlation

The observed variables are continuous, but $y_{1j}$, $y_{2j}$, $y_{3j}$, and $y_{4j}$ are ordinal variables, i.e., responses are classified into different ordered categories. Hence, the covariance matrix of continuous variables cannot be used in the SEM model. Instead, polychoric correlation should be employed.

An ordinal variable $z$ may be regarded as a crude measurement of an underlying unobserved or unobservable continuous variable $z^*$. For example, a four-point ordinal scale may be conceived as:

$$\begin{aligned} z \text{ is scored } 1 \quad &\text{if} \quad z^* \leq \alpha_1 \\ z \text{ is scored } 2 \quad &\text{if} \quad \alpha_1 < z^* \leq \alpha_2 \\ z \text{ is scored } 3 \quad &\text{if} \quad \alpha_2 < z^* \leq \alpha_3 \\ z \text{ is scored } 4 \quad &\text{if} \quad \alpha_3 < z^* \end{aligned}$$

where $\alpha_1 < \alpha_2 < \alpha_3$ are threshold values for $z^*$. It is often assumed that $z^*$ has a standard normal distribution, in which case the thresholds can be estimated from the inverse of the normal distribution function.

Let $z_1$ and $z_2$ be two ordinal variables with underlying continuous variables $z_1^*$ and $z_2^*$, respectively. Assuming that $z_1^*$ and $z_2^*$ have a bivariate normal distribution, their correlation is called the *polychoric correlation coefficient*.

An ordinal variable $z$ does not have a metric scale. To use such a variable in a linear relationship we use the corresponding underlying variable $z^*$ instead. The polychoric correlation

is not computed from actual scores but are rather theoretical correlations of the underlying $z^*$ variables. These correlations are estimated from the observed pairwise contingency tables of the ordinal variables in a two step process (Olsson, 1979):

1. Maximum Likelihood (ML) estimation of variable thresholds in $z^*$ implied by the observed ordinal categories.
2. ML estimation of the correlations between the $z^*$ (assumed to be normally distributed).

## 4.8 Model estimation and Goodness Of Fit

### 4.8.1 Model estimation

A fitting function $F$ must satisfy the following conditions to obtain consistent estimators of $\boldsymbol{\theta}$ (Browne, 1984):

1. $F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is a scalar.
2. $F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) \geq 0$.
3. $F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = 0 \iff \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{S}$.
4. $F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is continuous in $\boldsymbol{S}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

#### 4.8.1.1 Maximum Likelihood

The fitting function used for CDMM-O is the Maximum Likelihood (ML). The minimized function is the following

$$F_{ML} = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \operatorname{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) - \log|\boldsymbol{S}| - (p+q) \tag{4.15}$$

Generally it is assumed that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{S}$ are positive definite which means that they are not singular. Note that when $\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}$, $F_{ML} = 0$.

The ML estimators possess these important properties:

- They are asymptotically unbiased.

- They are consistent (they converge in probability to the population parameters $\boldsymbol{\theta}$: $\lim_{N\to\infty} P[|\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}| < \delta] = 1$, for any $\delta > 0$, where $\hat{\boldsymbol{\theta}}_N$ referes to a random vector coming from a sample with $N$ observations).
- They are asymptotically efficient (among consistent estimators, none has a smaller asymptotic variance).
- They are asymptotically normally distributed.

#### 4.8.1.2 Weighted Least Squares

When dealing with ordinal indicators, Muthén (1984) suggested a Weighted Least Squares (WLS) estimation in a 3-stage approach. The first two steps correspond to the ones mentioned in the Polychoric correlation Section, and the third consists in fitting a WLS function:

$$F_{WLS} = (\boldsymbol{s} - \boldsymbol{\sigma})' \boldsymbol{W}^{-1} (\boldsymbol{s} - \boldsymbol{\sigma}) \tag{4.16}$$

where

$$\boldsymbol{s}' = (s_{11}, s_{21}, s_{22}, s_{31}, \ldots, s_{p+q,p+q})$$

is a vector of the elements in the lower half, including the diagonal, of the covariance matrix $\boldsymbol{S}$ of order $(p+q) \times (p+q)$ used to fit the model to the data;

$$\boldsymbol{\sigma}' = (\boldsymbol{\sigma}_{11}, \boldsymbol{\sigma}_{21}, \boldsymbol{\sigma}_{22}, \boldsymbol{\sigma}_{31}, \ldots, \boldsymbol{\sigma}_{p+q,p+q})$$

is a vector of the corresponding elements of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ reproduced from the model parameters $\boldsymbol{\theta}$;

$\boldsymbol{W}$ is a weighted matrix of order $(p+q)(p+q+1)/2$, and $\boldsymbol{W}^{-1}$ must be a positive definite matrix. There are three main $\boldsymbol{W}$ choices:

1. $\boldsymbol{W} = \boldsymbol{I}$, leading to an Unweighted Least Squares estimation, ULS.
2. $\boldsymbol{W} = \text{diag}(\boldsymbol{S})$ (i.e., a diagonal matrix containing the variances of the observed variables), leading to a Mean and Variance adjusted Weigted Least Squares (WLSMV).
3. $\boldsymbol{W} = \boldsymbol{S}$, leading to a Fully Weighted Least Squares (FWLS).

The WLS approach is compatible with a complex sampling design with cluster and stratification variables in MPLUS.

The fitting function used for CDMM-L is the WLSMV, which is appropriate for latent variables with ordinal indicators (Flora and Curran, 2004). Moreover, as discussed in Yang-Wallentin et al. (2010), ML and WLSMV provide very similar results.

### 4.8.2 Goodness Of Fit

Fit refers to the ability of a model to reproduce the data (i.e., the variance-covariance matrix in SEM). A good fitting model is one that is reasonably consistent with the data and so it matches the observed data.

$\chi^2$ is a classic goodness-of-fit measure to determine overall model fit. The null hypothesis is that the implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is equivalent to the observed sample covariance matrix $\boldsymbol{S}$. A large $\chi^2$ and rejection of the null hypothesis means that the model estimates do not sufficiently reproduce sample covariance; the model does not fit the data well. By contrast, a small $\chi^2$ and failure to reject the null hypothesis is a sign of a good model fit. Nevertheless, this test is widely recognized to be problematic (Jöreskog, 1969), being one of its drawbacks to be sensitive to sample size: it becomes more and more difficult to retain the null hypothesis as the number of cases increases.

To cope with $\chi^2$ test problems, several goodness of fit indices have been developed. An absolute and two incremental fit indices are presented as follows, and they all are based on the $\chi^2$ statistic, the degrees of freedom of the model $df$, and the sample size $N$.

#### 4.8.2.1 Root Mean Squared Error of Approximation, RMSEA

The formula of this absolute measure is:

$$\text{RMSEA} = \sqrt{\frac{\chi^2 - df}{df(N-1)}} \tag{4.17}$$

If $\chi^2$ is less than $df$, then the RMSEA is set to zero, indicating perfect fit. By dividing by $df$, RMSEA penalizes free parameters. It also rewards a large sample size because $N$ is in the denominator. Hu and Bentler (1999) suggested values below 0.06 as a cut-off value for a good fit.

#### 4.8.2.2 Tucker-Lewis Index, TLI

The formula of this incremental fit measure is:

$$\text{TLI} = \frac{\frac{\chi_b^2}{df_b} - \frac{\chi_p^2}{df_p}}{\frac{\chi_b^2}{df_b} - 1} \tag{4.18}$$

where the subscript $_b$ refers to a baseline model and the subscript $_p$ to the proposed model.

Dividing by *df* penalizes free parameters to some degree. A value of 1 indicates perfect fit. TLI is also called non-normed because it may assume values $< 0$ and $> 1$, which are raised to 0 and truncated to 1, respectively. Hu and Bentler (1999) proposed $\geq 0.95$ as a cut-off value for a good fit.

#### 4.8.2.3 Comparative Fit Index, CFI

The formula of this incremental fit measure is:

$$\text{CFI} = 1 - \frac{\chi_p^2 - df_p}{\chi_b^2 - df_b} \tag{4.19}$$

where the subscript $_b$ refers to a baseline model and the subscript $_p$ to the proposed model.

Here, subtracting *df* from $\chi^2$ provides some penalty for free parameters. As before, values $> 1$ are truncated to 1, and values $< 0$ are raised to 0, and Hu and Bentler (1999) proposed $\geq 0.95$ as a cut-off value for a good fit.

To sum up, when RMSEA values are close to 0.06 or below, and CFI and TLI are close to 0.95 or greater, the model may have a reasonably good fit.

# Chapter 5

# Results

The presentation of results includes first a description of the variables in the dataset, continues with the results of the measurement model fit, and finalizes with the comparison of CDMM-O and CDMM-L by means of fit, parameters and standard errors.

## 5.1 Dataset description

### 5.1.1 Sociodemographics

Table 5.1 contains the description of the sociodemographic variables by country. Individuals had an average of 46 years of age and more than half (52.1%) were females. 60.7% of the sample was working, the next more prevalent employment category was retired (18.3%), and there were very few students (3.5%). Almost two thirds of the sample was married or cohabiting (64.7%). The countries with more Part 2 respondents were United States and Israel, while Belgium and the Netherlands were the countries with the least individuals. Except sex, there were significant differences among countries for the sociodemographic variables.

| **Country** | **N** | **Age** | **Employment status** | | | | | **Marital status** | | | **Sex** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Working | Student | Homemaker | Retired | Other | Married/ Cohabiting | Separated/ Widowed/ Divorced | Never married | Females |
| | | Mean (SE) | % (SE) | % (SE) | % (SE) | % (SE) | % (SE) | % (SE) | % (SE) | % (SE) | % (SE) |
| Belgium | 1043 | 46.9 (0.7) | 57.8 (1.4) | 3.2 (1.0) | 5.4 (1.0) | 24.1 (1.5) | 9.5 (1.0) | 69.8 (1.7) | 11.1 (1.2) | 19.1 (1.7) | 51.7 (2.4) |
| France | 1436 | 46.3 (0.7) | 62.1 (1.8) | 1.8 (0.4) | 4.6 (0.7) | 25.0 (1.6) | 6.6 (0.8) | 71.1 (1.8) | 11.2 (1.2) | 17.8 (1.3) | 52.2 (1.8) |
| Germany | 1323 | 48.2 (0.8) | 56.5 (2.1) | 1.7 (0.5) | 6.5 (0.8) | 26.7 (1.7) | 8.6 (1.1) | 63.3 (1.7) | 14.0 (1.0) | 22.7 (1.4) | 51.7 (1.4) |
| Israel | 4859 | 44.4 (0.2) | 60.2 (0.8) | 3.4 (0.3) | 6.1 (0.4) | 15.3 (0.4) | 15.1 (0.7) | 67.8 (0.7) | 13.4 (0.5) | 18.7 (0.5) | 51.9 (0.4) |
| Italy | 1779 | 47.7 (0.6) | 53.9 (1.7) | 1.6 (0.4) | 12.0 (1.0) | 24.0 (1.3) | 8.5 (0.8) | 66.7 (1.6) | 8.3 (1.0) | 25.0 (1.4) | 52.0 (1.5) |
| Japan | 1682 | 51.2 (0.7) | 63.5 (1.8) | 1.7 (0.6) | 16.7 (1.3) | 12.9 (1.3) | 5.2 (0.7) | 68.8 (1.4) | 13.0 (1.0) | 18.2 (1.6) | 53.0 (1.9) |
| Netherlands | 1094 | 45.0 (0.8) | 62.3 (2.6) | 5.2 (1.6) | 10.6 (1.3) | 14.5 (1.5) | 7.4 (1.1) | 72.1 (2.6) | 11.5 (1.1) | 16.4 (2.3) | 50.9 (2.2) |
| N.Ireland | 1708 | 45.3 (0.6) | 62.6 (1.9) | 4.2 (0.7) | 6.9 (0.8) | 20.0 (1.5) | 6.3 (0.7) | 59.6 (1.8) | 12.9 (1.1) | 27.6 (1.6) | 51.0 (1.4) |
| Portugal | 2060 | 46.5 (0.6) | 59.7 (1.5) | 5.5 (0.7) | 4.4 (0.7) | 21.7 (1.6) | 8.8 (0.7) | 69.6 (1.4) | 9.3 (0.6) | 21.2 (1.3) | 51.9 (1.5) |
| Spain | 2121 | 45.5 (0.6) | 50.4 (1.8) | 6.7 (1.1) | 16.6 (1.3) | 18.0 (1.1) | 8.2 (1.0) | 65.3 (1.5) | 9.3 (0.7) | 25.4 (1.5) | 51.4 (1.7) |
| United States | 5692 | 45.0 (0.4) | 66.8 (1.1) | 3.0 (0.5) | 5.6 (0.5) | 15.0 (0.8) | 9.6 (0.7) | 56.0 (1.2) | 20.8 (0.7) | 23.2 (1.2) | 53.0 (1.0) |
| All countries | 24797 | 46.0 (0.2) | 60.7 (0.5) | 3.5 (0.2) | 8.0 (0.3) | 18.3 (0.3) | 9.5 (0.3) | 64.7 (0.5) | 13.7 (0.3) | 21.6 (0.4) | 52.1 (0.4) |
| Comparison among countries[1] | | 13.3 (<0.0001) | | | 13.3 (<0.0001) | | | | 14.3 (<0.0001) | | 0.2 (1.0) |

[1] Wald statistic (p.value).

Table 5.1: Sociodemographic variables description

### 5.1.2 Disability dimensions and Perceived health

Figure 5.1 displays the proportions of indicator categories for the four latent variables.

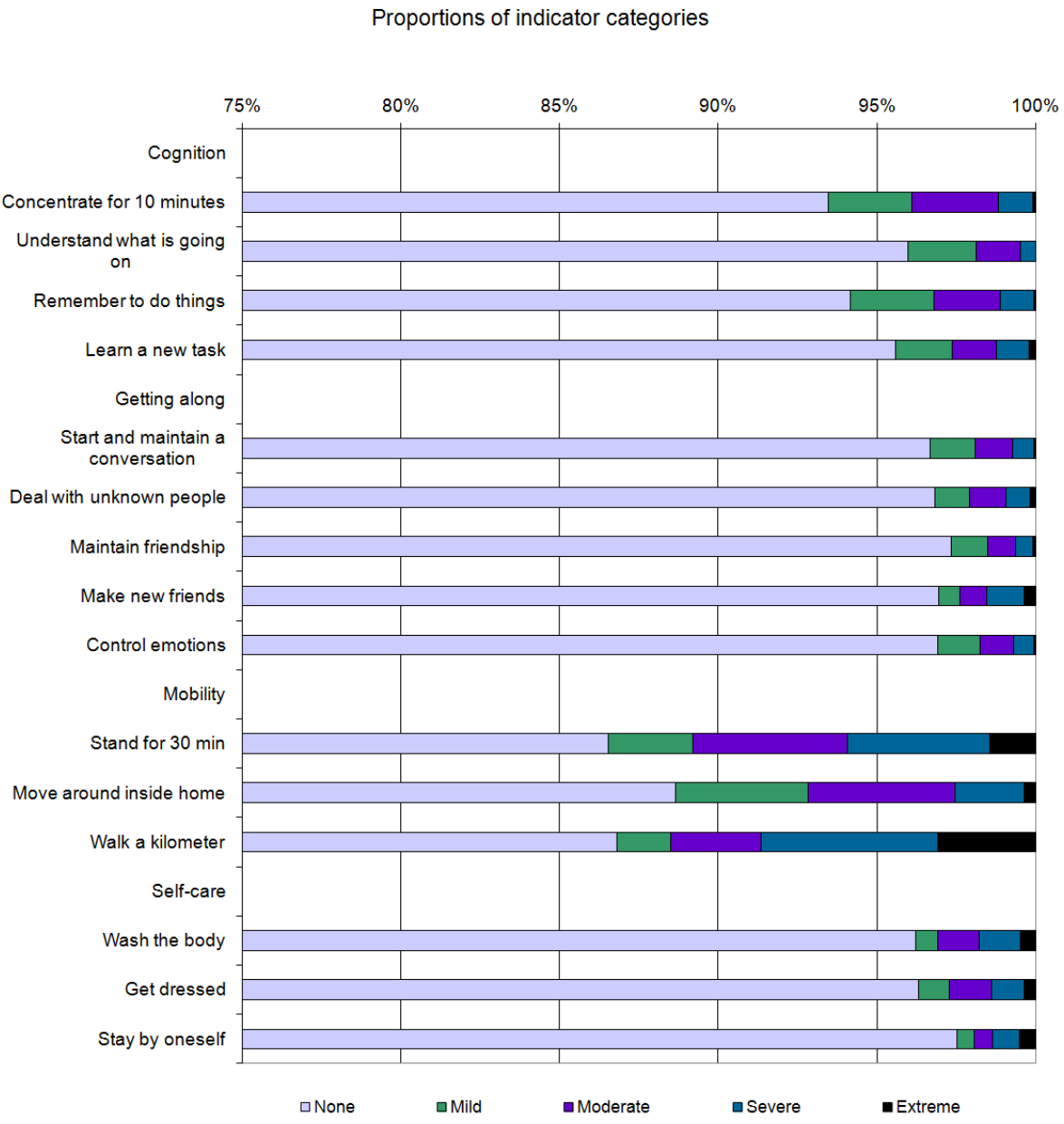



Figure 5.1: Proportions of indicator categories separated by disability dimensions
As there is a clear predominant category, the percentages are presented from 75% on.

The items showed an important asymetric distribution, and there was a substantial floor effect. Taken the 15 indicators together, a 80.2% of the individuals scored 0 (None) in all of them. The None category ranged from 86.6% (Mobility) to 97.5% (Self-care), so there was a clear predominance of this category; Mild from 0.6% (Self-care) to 4.2% (Mobility);

Moderate from 0.6% (Self-care) to 4.9% (Mobility); Severe from 0.4% (Cognition) to 5.6% (Mobility); and Extreme from 0.03% (Cognition) to 3.1% (Mobility).

Mobility was the dimension with the highest proportion of the categories other than None, followed by Cognition. Cognition had virtually no Extreme values, and Getting along, save for Make new friends, neither. Self-care was the dimension with the most balanced categories other than None.

Table 5.2 describes the observed disability dimensions and the outcome. Both the subsample concerning individuals with disability (affected individuals, i.e., those with a score $> 0$) and the whole sample are considered.

| Endogenous variable | Sum score range | All individuals | Individuals with disability | | |
|---|---|---|---|---|---|
| | | | Disability Dimension prevalence | Disability Dimension | Perceived health |
| | | Mean (SE) | % (SE) | Mean (SE) | Mean (SE) |
| Cognition | 0-16 | 0.37 (0.01) | 7.91 (0.20) | 4.63 (0.06) | 63.12 (0.62) |
| Getting along | 0-20 | 0.30 (0.01) | 4.79 (0.17) | 6.26 (0.13) | 58.30 (0.90) |
| Mobility | 0-12 | 0.89 (0.02) | 14.58 (0.30) | 6.13 (0.06) | 61.41 (0.52) |
| Self-care | 0-12 | 0.23 (0.01) | 4.13 (0.15) | 5.68 (0.14) | 51.96 (1.01) |
| Role functioning | 0-30 | 3.21 (0.06) | 42.04 (0.43) | 7.64 (0.13) | 76.16 (0.29) |
| Discrimination | 0-4 | 0.05 (0.00) | 2.76 (0.11) | 1.93 (0.04) | 50.26 (1.11) |
| Family burden | 0-4 | 0.16 (0.00) | 8.71 (0.22) | 1.81 (0.03) | 56.00 (0.66) |
| Stigma | 0-4 | 0.14 (0.00) | 7.60 (0.19) | 1.87 (0.03) | 55.43 (0.74) |
| Perceived health (outcome) | 0-100 | 80.65 (0.16) | | | |

Table 5.2: Observed disability dimensions and Perceived health description

Role functioning was the most frequently affected dimension (42.0%), and Mobility and Family burden showed the second most frequent difficulties (14.6% and 8.7%, respectively), while Discrimination was the least frequently affected (2.8%). Across the three directly comparable dimensions (Discrimination, Family Burden, and Stigma), the last two had a very similar distribution, while Discrimination presented lower scores. In fact, individuals with discrimination difficulties were the ones with the lowest perceived health mean (50.3). Besides, the eight mean VAS scores among individuals with any kind of difficulties were lower than the overall mean value (80.7). Among individuals with disability, those with Role functioning problems had the highest VAS scores.

Table 5.3 shows the correlation among the observed endogenous variables. All the disability dimensions were negatively correlated to Perceived health. Role functioning and Mobility were the dimensions most correlated with the outcome, with a value slightly below 0.5 in absolute value. Discrimination was the least correlated dimension (-0.25).

All the mediators were positively related: the values ranged from 0.22 (Self-care and Discrimination) to 0.59 (Family burden and Stigma), and most of them were between 0.3 and 0.5. In fact, both the mean and median values equaled 0.4, suggesting that mediators were moderately correlated, according to Dancey and Reidy (2004).

| | Perceived health | Cognition | Getting along | Mobility | Self-care | Role functioning | Discri-mination | Family burden | Stigma |
|---|---|---|---|---|---|---|---|---|---|
| Perceived health | 1.00 | | | | | | | | |
| Cognition | -0.29 | 1.00 | | | | | | | |
| Getting along | -0.26 | 0.46 | 1.00 | | | | | | |
| Mobility | -0.46 | 0.30 | 0.33 | 1.00 | | | | | |
| Self-care | -0.33 | 0.27 | 0.32 | 0.51 | 1.00 | | | | |
| Role functioning | -0.48 | 0.35 | 0.33 | 0.58 | 0.39 | 1.00 | | | |
| Discrimination | -0.25 | 0.30 | 0.31 | 0.27 | 0.22 | 0.28 | 1.00 | | |
| Family burden | -0.41 | 0.41 | 0.45 | 0.52 | 0.43 | 0.53 | 0.44 | 1.00 | |
| Stigma | -0.39 | 0.40 | 0.40 | 0.45 | 0.39 | 0.45 | 0.46 | 0.59 | 1.00 |

Table 5.3: Correlations among observed endogenous variables

### 5.1.3 Chronic conditions

The first column of Table 5.4 shows the prevalence of the 19 chronic conditions. Mental disorders (from Alcohol abuse to Specific phobia) were less prevalent than physical conditions (from Arthritis to Respiratory). Chronic pain (21.6%), Cardiovascular (19.3%), and Respiratory (19.2%) were the most prevalent conditions, while Neurological (1.1%) and Digestive (2.7%) were the least prevalent physical conditions. In turn, Depression (6.1%) and Specific phobia (5.5%) were the most prevailing mental disorders, while Drug abuse (0.6%) and Bipolar (0.9%) were the least prevalent conditions.

The remaining columns show the mean score of each disability dimension and perceived health among individuals with the condition in question. Save for Drug abuse in Mobility, individuals suffering from a condition had a higher mean disability score in all the dimensions, as well as a lower mean perceived health score, as compared to the overall respective means shown in Table 5.2. Note that each dimension had its scale (see Table 5.2), so the table is not comparable column-wise but it is row-wise. However, the highest (worst) disability scores always corresponded to Bipolar, Neurological or Posttraumatic stress. Similarly, the lowest scores always corresponded to Respiratory or Alcohol abuse (except for Cognition, which corresponded to Cardiovascular immediately followed by Respiratory).

The conditions with the lowest (worst) scores on perceived health were Neurological, Digestive, Panic disorder, Posttraumatic Stress, and Bipolar. In turn, the conditions with the highest scores on perceived health were Respiratory, Alcohol Abuse, and Drug abuse. The former includes the three conditions with most disability, and the latter includes the two conditions with less disability.

| Condition | Prevalence | Cognition | Getting along | Mobility | Self-care | Role functioning | Discrimination | Family burden | Stigma | Perceived health |
|---|---|---|---|---|---|---|---|---|---|---|
| | % (SE) | Mean (SE) | Mean (SE) | Mean (SE) | Mean (SE) | Mean (SE) | Mean (SE) | Mean (SE) | Mean (SE) | Mean (SE) |
| Alcohol abuse | 1.61 (0.1) | 1.00 (0.1) | 0.79 (0.1) | 0.90 (0.1) | 0.24 (0.0) | 4.76 (0.5) | 0.10 (0.0) | 0.24 (0.0) | 0.25 (0.0) | 78.92 (0.8) |
| Bipolar | 0.85 (0.1) | 2.56 (0.2) | 2.43 (0.3) | 2.31 (0.2) | 0.85 (0.1) | 9.79 (0.8) | 0.39 (0.1) | 0.69 (0.1) | 0.63 (0.1) | 66.73 (1.6) |
| Depression | 6.12 (0.2) | 1.45 (0.1) | 1.27 (0.1) | 1.67 (0.1) | 0.53 (0.0) | 7.80 (0.2) | 0.20 (0.0) | 0.48 (0.0) | 0.47 (0.0) | 68.43 (0.6) |
| Drug abuse | 0.57 (0.1) | 1.09 (0.2) | 1.20 (0.2) | 0.85 (0.1) | 0.34 (0.1) | 5.52 (0.7) | 0.18 (0.1) | 0.33 (0.1) | 0.25 (0.1) | 75.81 (1.7) |
| Generalized anxiety | 1.59 (0.1) | 1.78 (0.2) | 1.70 (0.2) | 2.20 (0.2) | 0.73 (0.1) | 8.75 (0.6) | 0.24 (0.0) | 0.62 (0.1) | 0.56 (0.1) | 66.76 (1.1) |
| Panic disorder | 1.85 (0.1) | 1.97 (0.1) | 2.01 (0.2) | 2.14 (0.1) | 0.78 (0.1) | 9.73 (0.5) | 0.24 (0.0) | 0.63 (0.0) | 0.63 (0.1) | 65.30 (1.1) |
| Posttraumatic stress | 1.85 (0.1) | 2.12 (0.2) | 1.99 (0.2) | 2.59 (0.2) | 0.86 (0.1) | 10.44 (0.6) | 0.26 (0.0) | 0.67 (0.1) | 0.60 (0.1) | 66.39 (1.1) |
| Social Phobia | 2.70 (0.1) | 1.66 (0.1) | 1.52 (0.1) | 1.72 (0.1) | 0.51 (0.1) | 7.31 (0.3) | 0.21 (0.0) | 0.51 (0.0) | 0.47 (0.0) | 71.63 (0.8) |
| Specific phobia | 5.45 (0.2) | 0.99 (0.1) | 0.82 (0.1) | 1.66 (0.1) | 0.43 (0.0) | 6.40 (0.3) | 0.14 (0.0) | 0.39 (0.0) | 0.36 (0.0) | 73.46 (0.7) |
| Arthritis | 15.84 (0.4) | 0.69 (0.0) | 0.58 (0.0) | 2.22 (0.1) | 0.61 (0.0) | 6.53 (0.2) | 0.12 (0.0) | 0.33 (0.0) | 0.33 (0.0) | 69.53 (0.5) |
| Cancer | 3.11 (0.2) | 0.63 (0.1) | 0.65 (0.1) | 2.04 (0.1) | 0.55 (0.1) | 6.55 (0.4) | 0.11 (0.0) | 0.37 (0.0) | 0.30 (0.0) | 71.82 (1.0) |
| Cardiovascular | 19.31 (0.3) | 0.59 (0.0) | 0.51 (0.0) | 1.88 (0.1) | 0.57 (0.0) | 5.79 (0.2) | 0.11 (0.0) | 0.32 (0.0) | 0.30 (0.0) | 70.21 (0.4) |
| Chronic pain | 21.55 (0.4) | 0.80 (0.0) | 0.68 (0.0) | 2.19 (0.1) | 0.60 (0.0) | 7.09 (0.2) | 0.13 (0.0) | 0.37 (0.0) | 0.35 (0.0) | 70.62 (0.4) |
| Diabetes | 5.12 (0.2) | 0.68 (0.1) | 0.57 (0.1) | 2.06 (0.1) | 0.67 (0.1) | 6.73 (0.4) | 0.13 (0.0) | 0.32 (0.0) | 0.31 (0.0) | 67.91 (1.0) |
| Digestive | 2.74 (0.1) | 1.14 (0.1) | 1.08 (0.1) | 2.64 (0.2) | 0.80 (0.1) | 9.40 (0.5) | 0.21 (0.0) | 0.55 (0.0) | 0.54 (0.0) | 64.85 (1.2) |
| Headaches/ Migraines | 11.01 (0.3) | 1.15 (0.1) | 0.93 (0.1) | 1.82 (0.1) | 0.54 (0.0) | 6.47 (0.2) | 0.15 (0.0) | 0.39 (0.0) | 0.34 (0.0) | 72.65 (0.5) |
| Insomnia | 7.42 (0.2) | 1.43 (0.1) | 1.24 (0.1) | 2.43 (0.1) | 0.80 (0.1) | 8.42 (0.3) | 0.23 (0.0) | 0.54 (0.0) | 0.47 (0.0) | 67.71 (0.7) |
| Neurological | 1.10 (0.1) | 1.43 (0.2) | 1.38 (0.2) | 2.97 (0.3) | 1.28 (0.2) | 9.55 (0.9) | 0.35 (0.1) | 0.57 (0.1) | 0.63 (0.1) | 62.81 (1.8) |
| Respiratory | 19.22 (0.5) | 0.62 (0.0) | 0.48 (0.0) | 1.31 (0.1) | 0.32 (0.0) | 4.40 (0.1) | 0.08 (0.0) | 0.22 (0.0) | 0.20 (0.0) | 78.32 (0.4) |

Table 5.4: Prevalence of Conditions and Mean of Disability and Perceived health among individuals with the condition

## 5.2 Measurement model for latent disability dimensions

Before building the structural model, the measurement model for the latent variables and their indicators must be tested.

Figure 5.2 shows the confirmatory factor analysis carried out with the four latent variables —Cognition, Getting along, Mobility and Self-care— and their respective indicators. The model presented excellent fit (RMSEA=0.011, and CFI=TLI=0.999).

We observe that all the standardized factor loadings were above 0.93, which translates into the fact that all indicators had an important *loading* on the latent factor they represent. Self-care was the factor with the highest loadings, and Getting along the one with the lowest. The correlations among factors ranged between 0.56 and 0.83, so factors were from moderate to strongly correlated (Dancey and Reidy, 2004).

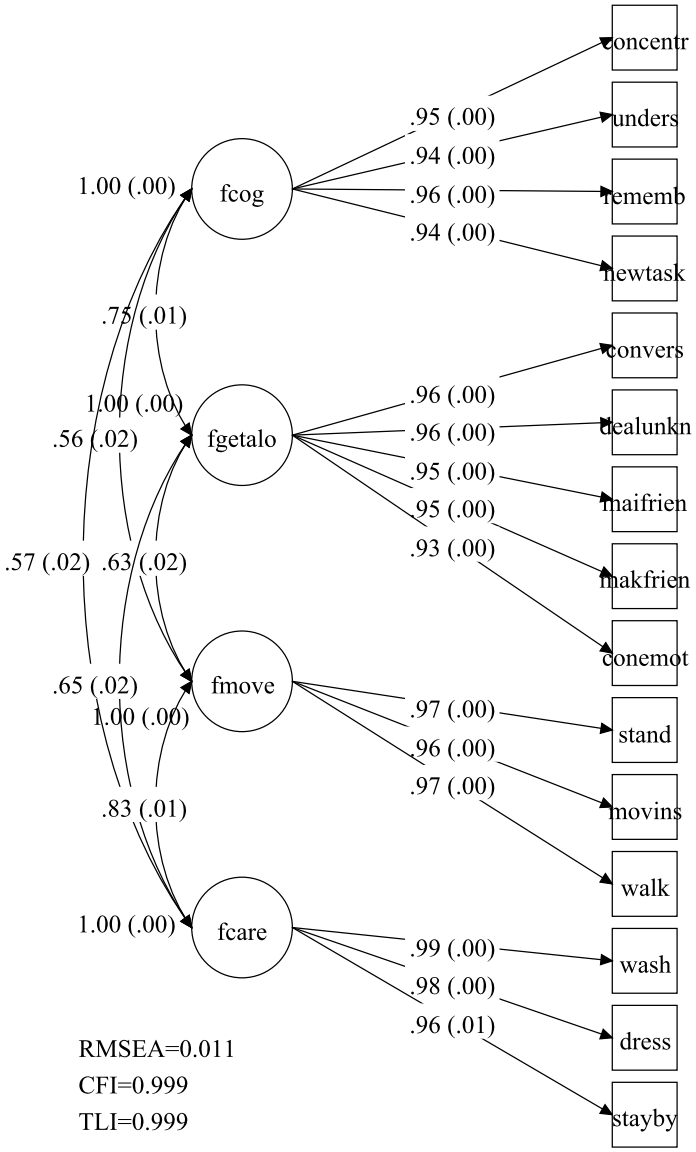



Figure 5.2: Measurement model for latent disability dimensions
Standardized coefficients are presented.

From top to bottom, the names of each set of variables are the following:
Cognition (*fcog*): Concentrate (*concentr*), Understand (*unders*), Remember (*rememb*), Learn a new task (*newtask*).
Getting along (*fgetalo*): Converse (*convers*), Deal with unknown people (*dealunkn*), Maintain friendship (*maifrien*), Make new friends (*makfrien*), Control emotions (*conemot*).
Mobility (*fmove*): Stand for 30 min (*stand*), Move around inside (*movins*), Walk a kilometer (*walk*).
Self-care (*fcare*): Wash the body (*wash*), Get dressed (*dress*), Stay by oneself (*stayby*).

## 5.3 CDMM-O and CDMM-L: Model comparison

### 5.3.1 Model specification

In this Section I compare the Complex Disability Mediated Models with observed (CDMM-O) and latent (CDMM-L) variables. The observed variables —the outcome (Perceived health), the 19 chronic conditions, the sociodemographic variables and the four mediators Role functioning, Discrimination, Family buren and Stigma— were the same for both models. The remaining mediators —Cognition, Getting along, Mobility, and Self-care— were also observed for CDMM-O but latent for CDMM-L.

It is important to highlight that correlations among dimensions were taken into account; in Table 5.3 we observed that the mean and median correlation values among dimensions were 0.4. In Figure 5.2 we saw that the correlation among latent factors ranged between 0.56 and 0.83. This sizeable figures suggest that correlation among factors cannot be ignored.

### 5.3.2 Goodness of fit

Figures 5.3 and 5.4 depict the path diagrams for CDMM-O and CDMM-L, respectively.

All the fit measures were better for CDMM-L. Although both RMSEA values were acceptable, this was not the case for CFI and TLI. According to Hu and Bentler (1999), they should be above the cutoff 0.95, and the CDMM-O values were both below. In contrast, the CDMM-L were both above, suggesting a satisfactory fit.

Regarding the proportion of variance explained for the outcome ($R^2$), both values were very similar: 0.390 for CDMM-O and 0.425 for CDMM-L, slightly higher for CDMM-L, but significantly different as their Confidence Intervals (CI) did not overlap: The 99% CI for $R^2$ (observed) was 0.378-0.402, while for $R^2$ (latent) it was 0.413-0.437.

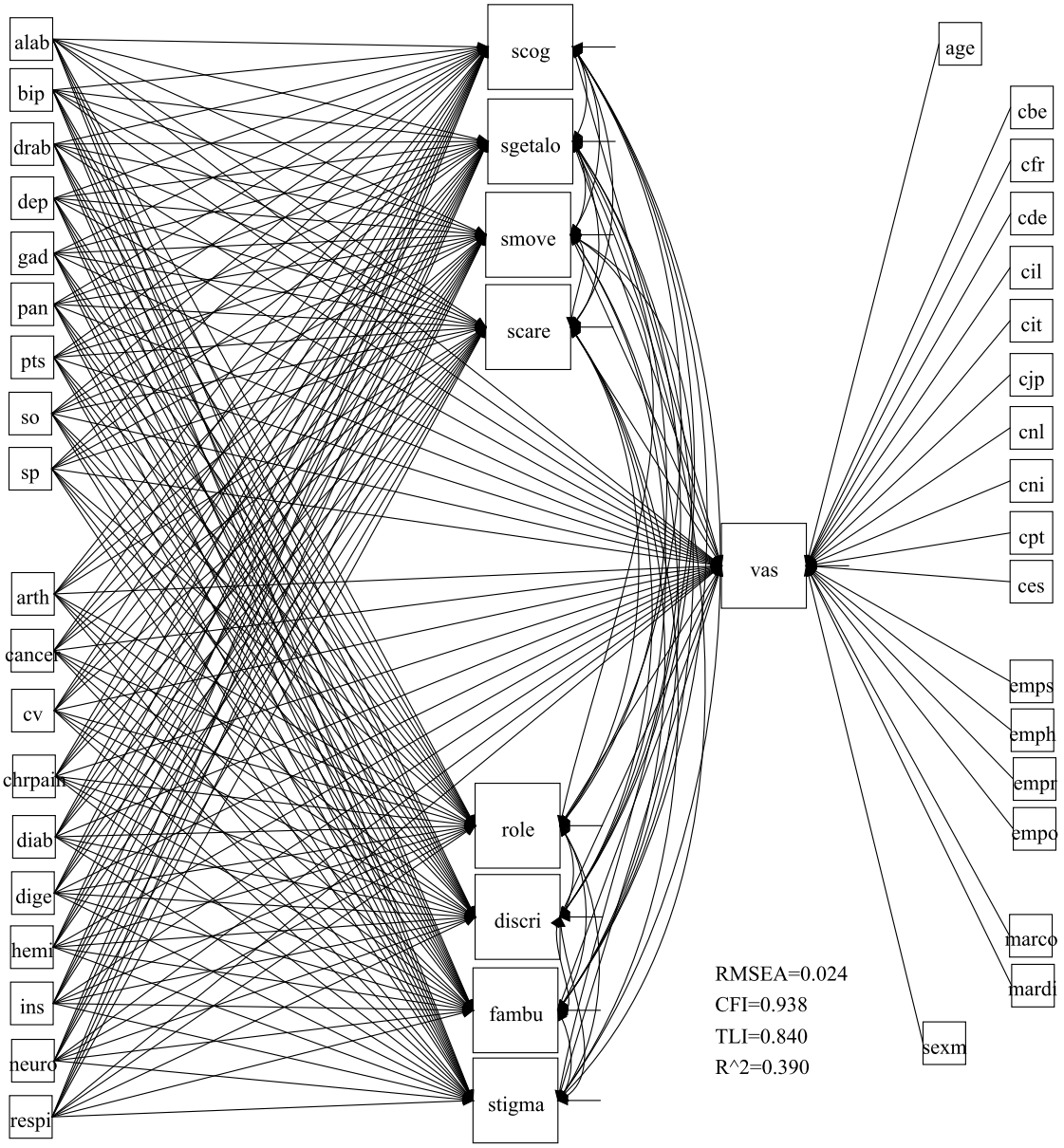



Figure 5.3: Structural model for CDMM-O obtained from MPLUS

From left to right, the names of each set of variables are the following:
Alcohol abuse (*alab*), Bipolar (*bip*), Drug abuse (*drab*), Depression (*dep*), Generalized anxiety (*gad*),
Panic disorder (*pan*), Posttraumatic stress (*pts*), Social phobia (*so*), Specific phobia (*sp*).
Arthritis (*arth*), Cancer (*cancer*), Cardiovascular (*cv*), Chronic pain (*chrpain*), Diabetes (*diab*), Digestive (*dige*),
Headaches/Migraines (*hemi*), Insomnia (*ins*), Neurological (*neuro*), Respiratory (*respi*).
Cognition (*scog*), Getting along (*sgetalo*), Mobility (*smove*), Self-care (*scare*)
Role functioning (*role*), Discrimination (*discri*), Family burden (*fambu*), Stigma (*stigma*).
Perceived health (*vas*).
Age (*age*), Belgium (*cbe*), France (*cfr*), Germany (*cde*), Israel (*cil*), Italy (*cit*), Japan (*cjp*),
Netherlands (*cnl*), Northern Ireland (*cni*), Portugal (*cpt*), Spain (*ces*),
Student (*emps*), Homemaker (*emph*), Retired (*empr*), Other (*empo*).
Married/Cohabiting (*marco*), Separated/Widowed/Divorced (*mardi*), Male (*sexm*).

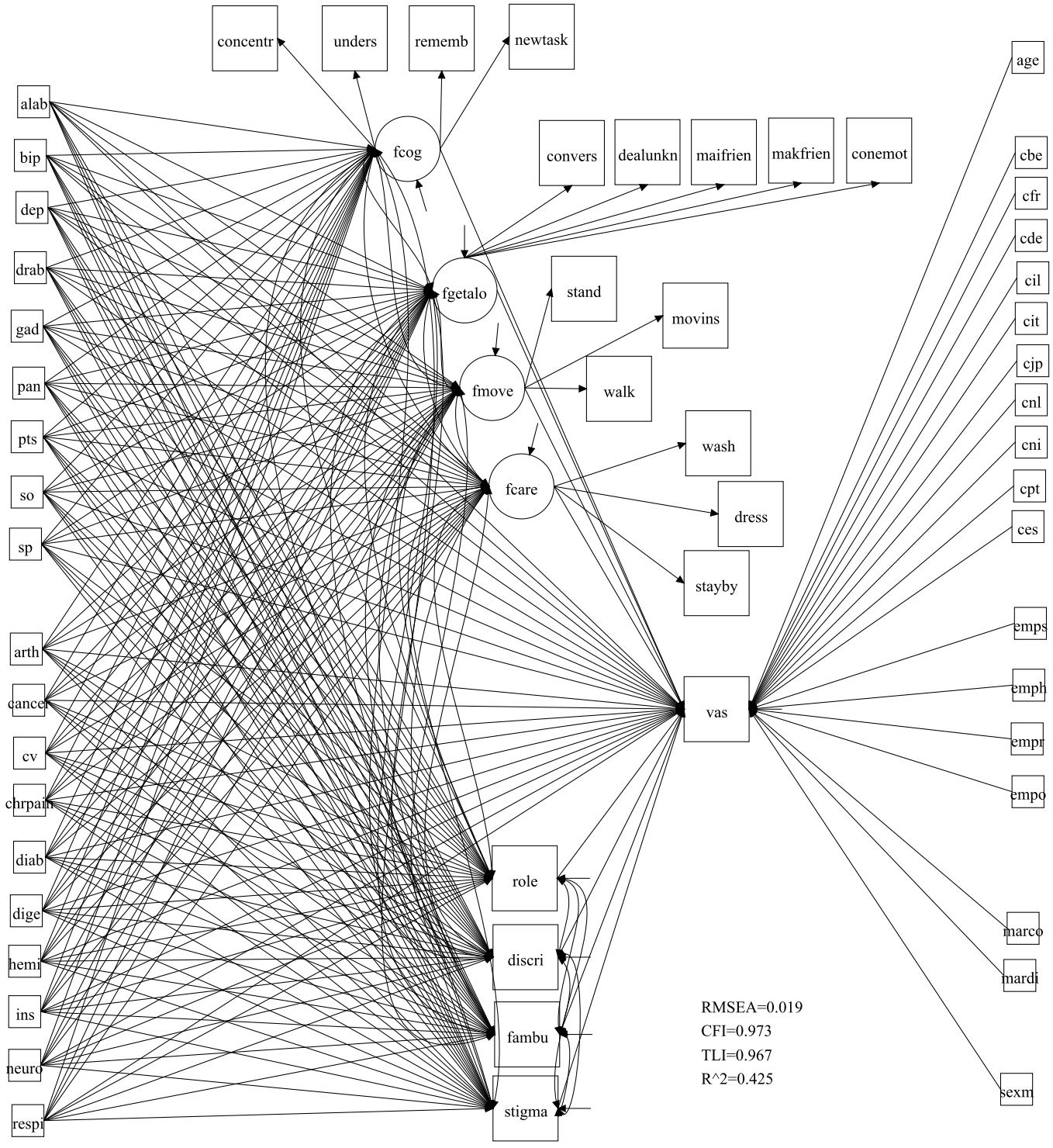



Figure 5.4: Structural model for CDMM-L obtained from MPLUS

From left to right, the names of each set of variables are the following:
Alcohol abuse (*alab*), Bipolar (*bip*), Drug abuse (*drab*), Depression (*dep*), Generalized anxiety (*gad*),
Panic disorder (*pan*), Posttraumatic stress (*pts*), Social phobia (*so*), Specific phobia (*sp*).
Arthritis (*arth*), Cancer (*cancer*), Cardiovascular (*cv*), Chronic pain (*chrpain*), Diabetes (*diab*), Digestive (*dige*),
Headaches/Migraines (*hemi*), Insomnia (*ins*), Neurological (*neuro*), Respiratory (*respi*).
Cognition (*fcog*): Concentrate (*concentr*), Understand (*unders*), Remember (*rememb*), Learn a new task (*newtask*).
Getting along (*fgetalo*): Converse (*convers*), Deal with unknown people (*dealunkn*), Maintain friendship (*maifrien*),
Make new friends (*makfrien*), Control emotions (*conemot*).
Mobility (*fmove*): Stand for 30 min (*stand*), Move around inside (*movins*), Walk a kilometer (*walk*).
Self-care (*fcare*): Wash the body (*wash*), Get dressed (*dress*), Stay by oneself (*stayby*).
Role functioning (*role*), Discrimination (*discri*), Family burden (*fambu*), Stigma (*stigma*).
Perceived health (*vas*).
Age (*age*), Belgium (*cbe*), France (*cfr*), Germany (*cde*), Israel (*cil*), Italy (*cit*), Japan (*cjp*),
Netherlands (*cnl*), Northern Ireland (*cni*), Portugal (*cpt*), Spain (*ces*),
Student (*emps*), Homemaker (*emph*), Retired (*empr*), Other (*empo*).
Married/Cohabiting (*marco*), Separated/Widowed/Divorced (*mardi*), Male (*sexm*).

### 5.3.3 Parameters and Standard Errors evaluation

Before tackling the decomposition of causal effects, regarding sociodemographic variables, all of them except Sex were significant. The significant coefficients had the same sign for both CDMM-O and CDMM-L. Coefficients were higher in absolute value for CDMM-L, while the standard errors were similar for CDMM-L and CDMM-O.

More details can be found in the Appendix Table C.1.

#### 5.3.3.1 Decomposition of the causal effects

**Mediators on perceived health**

Table 5.5 shows the direct effects of mediators on perceived health.

| Mediator | **Direct effects for CDMM-O** | **Direct effects for CDMM-L** | **Standardized Direct effects for CDMM-O** | **Standardized Direct effects for CDMM-L** |
|---|---|---|---|---|
| | Coef (SE) | Coef (SE) | Coef (SE) | Coef (SE) |
| Cognition[1] | -0.50 (0.11)* | -1.25 (0.26)* | -0.039 (0.008)* | -0.066 (0.014)* |
| Getting along[1] | 0.08 (0.11) | -0.23 (0.35) | 0.006 (0.009) | -0.012 (0.018) |
| Mobility[1] | -0.99 (0.09)* | -2.09 (0.33)* | -0.125 (0.012)* | -0.115 (0.018)* |
| Self-care[1] | -0.58 (0.17)* | -1.28 (0.38)* | -0.039 (0.012)* | -0.069 (0.020)* |
| Role functioning[2] | -0.47 (0.03)* | -0.45 (0.01)* | | |
| Discrimination[2] | -1.16 (0.59)*† | -3.72 (0.55)*† | | |
| Family Burden[2] | -2.44 (0.42)* | -2.29 (0.17)* | | |
| Stigma[2] | -2.34 (0.44)* | -2.09 (0.19)* | | |

[1] Direct effects were not directly comparable because they have different scales in CDMM-O and CDMM-L. Standardized effects should be regarded in terms of comparison.
[2] Direct effects were directly comparable because the variables are the same in both models.
$*$ p.value $< 0.05$.
† Observed and latent coefficients were significantly different at 5% level.

Table 5.5: Direct effects of mediators on perceived health for CDMM-O and CDMM-L The models are adjusted for Age, Country, Employment status, Marital status, Sex, and the 19 chronic conditions.

Regarding standardized coefficients, with the exception of Mobility, they were higher in absolute value for CDMM-L. The standard errors were also higher for CDMM-L. Getting along was the only non-significant disability dimension, and it had a positive value.

The coefficients for the remaining dimensions, which are directly comparable, were very similar for CDMM-O and CDMM-L, with the exception of Discrimination, which was more than three times higher for CDMM-L (-3.72) than for CDMM-O (-1.16). The standard errors for these coefficients were lower for CDMM-L.

Across the eight dimensions, Discrimination was the only one with a direct effect significantly different for CDMM-O and CDMM-L.

**Chronic conditions on perceived health**

Figure 5.5 depicts the total effects of chronic conditions on VAS score broken down into direct and indirect effects, according to both modelling strategies.

The total effects of CDMM-O and CDMM-L were perfectly correlated ($r = 0.98$). All of the total effects were significant for CDMM-L, while the total effects for Drug Abuse, Alcohol abuse, and Respiratory were not significant in the CDMM-O model.

Neurological, Depression, and Bipolar presented the highest total effects. The decrement in perceived health was 11.1 (CDMM-O) and 10.5 (CDMM-L) for Neurological conditions, 8.2/7.9 for Depression and 7.6/7.4 for Bipolar.

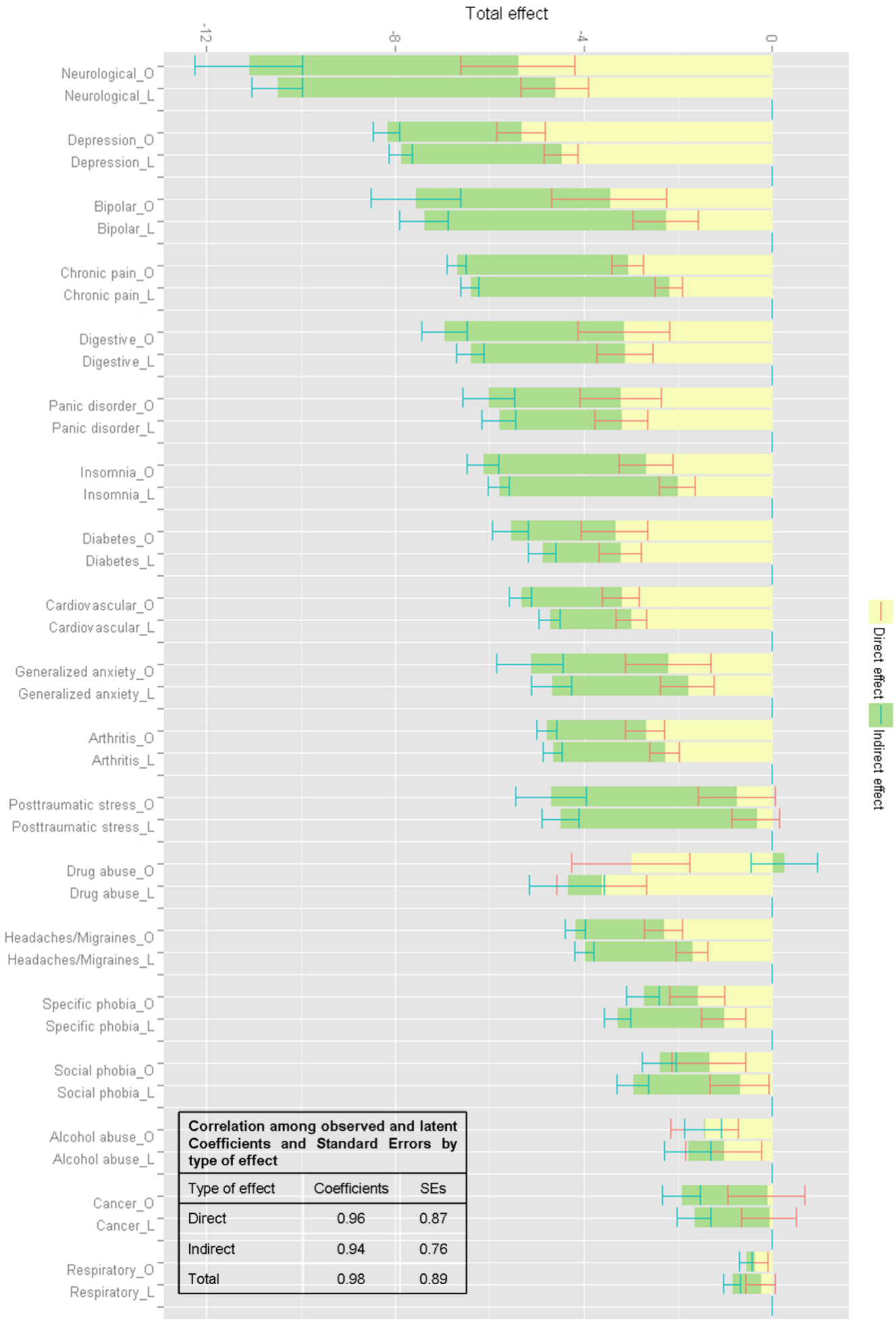
Direct effects were perfectly correlated, too ($r = 0.96$). They were not significant for Posttraumatic stress, Social phobia, Cancer, and Respiratory for both models; Alcohol abuse was neither significant for CDMM-L.

Indirect effects, although highly correlated ($r = 0.94$), had the lowest correlation coefficient among observed and latent models. They were not significant for Alcohol and Drug abuse in both models; they were neither significant for Respiratory in CDMM-O. Cancer had virtually all its effect mediated. The indirect effect was also much more important than the direct effect for Posttraumatic stress.

The standard errors of all type of effects were strongly correlated among both models: they were around 0.90 with the exception of Indirect effects with a value of 0.76. For the three types of effects, the standard errors were lower for CDMM-L than for CDMM-O, even though the metric for latent variables was higher than for the observed ones, and therefore it intrinsically led to higher standard errors for CDMM-L; despite this detriment, standard errors of CDMM-L kept lower.

With standardized coefficients, both the coefficients and the standard errors can be compared: Direct effects were higher in absolute value for CDMM-O (Mean value=-0.029) than for CDMM-L (Mean value=-0.023), while for indirect effects the opposite was true: they were lower for CDMM-O (Mean value=-0.026) than for CDMM-L (Mean value=-0.029). Concerning standard errors, the mean SE values for both direct and indirect effects were lower for CDMM-L (0.005 and 0.003, respectively) than for CDMM-O (0.007 and 0.004, respectively).

More details of the effects decomposition can be found in C.2 and C.3.




**Correlation among observed and latent Coefficients and Standard Errors by type of effect**

| Type of effect | Coefficients | SEs |
|---|---|---|
| Direct | 0.96 | 0.87 |
| Indirect | 0.94 | 0.76 |
| Total | 0.98 | 0.89 |

Figure 5.5: Total effects of conditions for CDMM-O (O) and CDMM-L (L)
The bars are broken down into direct (yellow) and indirect (green) effects. The error bars represent the standard errors of the direct and indirect effects. Conditions are sorted by the highest to the lowest total effect for CDMM-L in absolute value.

#### 5.3.3.2 Overall Indirect Contributions (OICs) and Specific Indirect effects

Figure 5.6 depicts the specific indirect effects for CDMM-O and CDMM-L presented in terms of Specific Contributions (SCs, percentage of specific indirect over overall indirect effect).

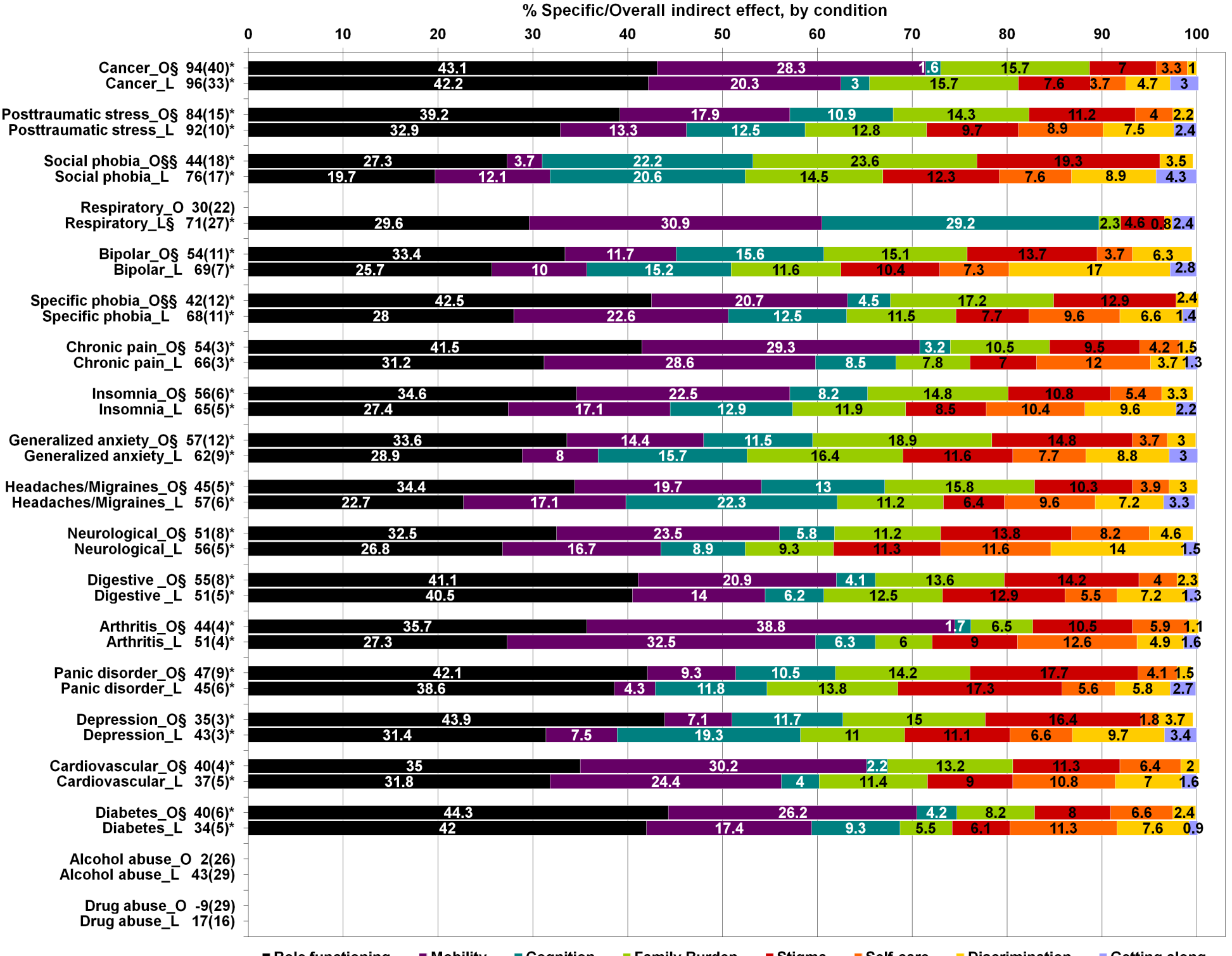
The OIC contribution for Cancer was almost 100% (96%), thus its direct effect was near 0. In contrast, Diabetes had the lowest OIC for CDMM-L (34%), and therefore most of its total effect (two thirds) was unmediated. Note that Drug abuse presented a negative percentage for CDMM-O because of the opposite sign of its direct and indirect effects, as it shows Figure 5.5.

On average, indirect effects represented a lower proportion of the total effect for CDMM-O (mean=45.4, median=44.8, IQR=39.8-54.5), while for CDMM-L the opposite was true: OICs had a mean value of 57.8, a median of 57.2, and an IQR of 44.0-68.9. In Figure 5.6 conditions were sorted from higher to lower OIC according to CDMM-L, and the order would have been quite different according to CDMM-O; in particular, Respiratory would have been moved from the fourth to the seventeenth place. Besides, the OIC of Respiratory was not significant for CDMM-O.

This disagreement between observed and latent OICs was in line with the correlation value of 0.79 among observed and latent OICs; although strong, it was below the corresponding value of Total, Direct, and Indirect effects. In contrast, the correlation among standard errors of OICs was 0.93, above the corresponding values of Total, Direct, and Indirect.

The mean value of standard errors for OICs was 12.78 for CDMM-O versus 10.86 for CDMM-L.

Role functioning was the dimension with the highest contribution to overall indirect effects in all the conditions, with the exception of Arthritis in CDMM-O and CDMM-L, and Social phobia and Respiratory for CDMM-L, where Mobility or Cognition were the main contributors. In CDMM-O the specific indirect effects for Getting along were positive for all the conditions; consequently its percentages were negative, and they were not represented in the diagram. In CDMM-O, seven instead of eight dimensions made up the 100% percentage. The dimensions were sorted from most to least contributing for CDMM-L. For CDMM-O, Cognition moved from the third to the fifth position. In fact, the percentages for four of the five most contributing dimensions (Role functioning, Mobility, Family burden and Stigma) were higher for CDMM-O, and for Cognition and the three least contributing dimensions (Self-care, Discrimination and Getting along) they were greater for CDMM-L.




§The Specific Contributions (SCs) have been rescaled to obtain a percentage of 100%.

* p.value<0.05.

Figure 5.6: Specific Contributions (SCs) for CDMM-O(O) and CDMM-L(L) by condition

Conditions are sorted from highest to lowest Overall Indirect Contributions (OIC, percentage of overall indirect over total effect) for CDMM-L, values indicated with their SE. Specific indirect effects are presented in terms of Specific Contributions (SCs, percentage of specific indirect over overall indirect effect), values indicated within the bars. Dimensions are sorted from highest to lowest SC for CDMM-L. SCs are only represented for conditions with significant OIC. Some conditions had one (§) or two (§§) dimensions with negative SC, which had been set up to 0 and the remaining dimensions had been rescaled to 100%.

The contribution of Role functioning ranged from 27.3% to 44.3% in CDMM-O, and from 19.7% to 42.0% in CDMM-L; and for Discrimination from 1.0% to 6.3% in CDMM-O and from 0.8% to 17.0% in CDMM-L, so Discrimination played a more important role in CDMM-L, which was consistent with having a significantly higher Direct effect (see Table 5.5). There were dimensions with a greater influence in mental than in physical conditions, and vice versa. Mobility had a clearer mediation contribution in physical conditions, with a median of 18.9 in physical conditions and 10.0 in mental for CDMM-L; and Cognition in mental disorders, with a median of 15.2 for mental and 8.7 for physical conditions for CDMM-L. Role functioning and Self-care also presented higher values for physical conditions, and Family burden and Stigma were more important mediators for mental disorders (see Table 5.6).

| Dimension | Model | All conditions | | Mental disorders | | Physical conditions | |
|---|---|---|---|---|---|---|---|
| | | Mean | Median (IQR) | Mean | Median (IQR) | Mean | Median (IQR) |
| Role functioning | CDMM-O | 38.0 | 37.6 (34.4 - 42.5) | 37.8 | 39.5 (33.9 - 42.5) | 38.1 | 35.6 (34.8 - 41.6) |
| | CDMM-L | 31.0 | 30.3 (27.3 - 32.9) | 29.3 | 28.9 (26.9 - 32.2) | 32.2 | 30.8 (27.3 - 38.3) |
| Mobility | CDMM-O | 20.5 | 20.9 (14.0 - 26.8) | 12.4 | 12.1 (8.6 - 16.4) | 26.7 | 26.3 (22.7 - 29.4) |
| | CDMM-L | 17.5 | 17.1 (12.1 - 22.6) | 11.1 | 10.0 (7.8 - 12.7) | 22.0 | 18.9 (17.1 - 27.6) |
| Cognition | CDMM-O | 8.4 | 7.2 (4.0 - 11.9) | 12.8 | 11.8 (11.1 - 14.0) | 5.0 | 4.2 (2.2 - 6.0) |
| | CDMM-L | 12.9 | 12.5 (8.5 - 15.7) | 15.4 | 15.2 (12.5 - 17.5) | 11.1 | 8.7 (6.2 - 12.0) |
| Family burden | CDMM-O | 14.5 | 14.8 (12.8 - 15.9) | 17.3 | 15.5 (15.0 - 18.2) | 12.3 | 13.2 (10.6 - 15.0) |
| | CDMM-L | 10.9 | 11.5 (9.3 - 12.8) | 13.1 | 12.8 (11.6 - 14.2) | 9.4 | 10.3 (6.5 - 11.8) |
| Stigma | CDMM-O | 12.8 | 12.2 (10.5 - 14.5) | 15.5 | 15.1 (13.5 - 17.4) | 10.7 | 10.5 (9.6 - 11.3) |
| | CDMM-L | 9.6 | 9.0 (7.6 - 11.3) | 11.4 | 11.1 (10.1 - 12.0) | 8.3 | 8.1 (6.6 - 9.0) |
| Self-care | CDMM-O | 4.2 | 4.3 (3.9 - 5.7) | 2.6 | 4.0 (1.0 - 4.2) | 5.4 | 5.6 (4.2 - 6.4) |
| | CDMM-L | 8.0 | 8.9 (6.6 - 10.8) | 7.6 | 7.6 (7.0 - 8.3) | 8.3 | 10.6 (6.5 - 11.5) |
| Discrimination | CDMM-O | 3.0 | 2.5 (2.0 - 3.6) | 3.6 | 3.3 (2.5 - 4.1) | 2.5 | 2.4 (1.6 - 3.2) |
| | CDMM-L | 7.7 | 7.2 (5.8 - 8.9) | 9.2 | 8.8 (7.1 - 9.3) | 6.7 | 7.1 (4.8 - 7.5) |
| Getting along | CDMM-O | -1.2 | -1.1 (-1.8 - -0.5) | -1.9 | -2.0 (-2.4 - -1.7) | -0.7 | -0.7 (-1 - -0.4) |
| | CDMM-L | 2.3 | 2.4 (1.5 - 3.0) | 2.9 | 2.8 (2.6 - 3.2) | 2.0 | 1.6 (1.4 - 2.8) |

Table 5.6: Mean and median values of Specific Contributions (SCs) for CDMM-O and CDMM-L
All the conditions, Mental disorders, and Physical conditions are considered.

As opposite to the fact that in CDMM-O each condition had one or two negative proportions, in CDMM-L it only happened once, for Respiratory, where Self-care had a positive specific indirect effect. Hence, for specific indirect effects, CDMM-L presented more acceptable estimates.

Table 5.7 shows the correlations among the overall indirect effect and each dimension specific indirect effect. Role funcioning and Stigma presented the highest correlations, and Getting along and Cognition the lowest. Mobility and Getting along were the dimensions with a wider difference across observed and latent models. For CDMM-O, Getting along

| | Overall indirect effect | |
|---|---|---|
| Specific indirect effect | CDMM-O | CDMM-L |
| Cognition | 0.65 | 0.71 |
| Getting along | -0.64 | 0.66 |
| Mobility | 0.85 | 0.71 |
| Self-care | 0.89 | 0.82 |
| Role functioning | 0.97 | 0.93 |
| Discrimination | 0.79 | 0.84 |
| Family burden | 0.92 | 0.90 |
| Stigma | 0.94 | 0.91 |

Table 5.7: Correlations among overall indirect effect and specific indirect effects for CDMM-O and CDMM-L

had a negative value, suggesting that the effect of this particular dimension went in the opposite sense that the overall effect of the set of dimensions.

Table 5.8 shows the correlations among specific observed and latent effects.

| | Cognition | Getting along | Mobility | Self-care | Role functioning | Discri-mination | Family burden | Stigma |
|---|---|---|---|---|---|---|---|---|
| Cognition | **0.90** | -0.95 | 0.21 | 0.43 | 0.59 | 0.78 | 0.79 | 0.73 |
| Getting along | 0.85 | -**0.87** | -0.19 | -0.46 | -0.61 | -0.78 | -0.8 | -0.73 |
| Mobility | 0.24 | 0.08 | **0.88** | 0.86 | 0.81 | 0.50 | 0.63 | 0.68 |
| Self-care | 0.45 | 0.48 | 0.70 | **0.79** | 0.83 | 0.76 | 0.73 | 0.80 |
| Role functioning | 0.53 | 0.50 | 0.66 | 0.69 | **0.98** | 0.71 | 0.89 | 0.91 |
| Discrimination | 0.71 | 0.74 | 0.38 | 0.72 | 0.69 | **0.99** | 0.79 | 0.80 |
| Family burden | 0.72 | 0.79 | 0.42 | 0.63 | 0.85 | 0.78 | **0.97** | 0.92 |
| Stigma | 0.66 | 0.64 | 0.50 | 0.61 | 0.90 | 0.79 | 0.89 | **0.98** |

Table 5.8: Correlations among observed and latent specific indirect effects
The diagonal contains the correlation between the corresponding observed and latent dimensions. The upper-diagonal contains the correlations among observed dimensions, and the lower diagonal the correlations among latent dimensions.

Again, values concerning Getting along (observed) were negative, contrary to what is expected in the disability model and also inconsistent with the positive correlation values for Getting along in Table 5.3. The observed dimension with the corresponding latent dimension correlated from strongly to perfectly: five of the eight scales had values above 0.9, and the minimum was 0.79 in Self-care. Despite presenting much bigger SCs for CDMM-L, Discrimination was the dimension with the highest correlation among observed and latent models. The median correlation of observed-latent dimensions was 0.93.

Cognition and Getting along, Family burden and Stigma, Role functioning and Family burden, and Role functioning and Stigma were the dimensions with the highest correlation values in both CDMM-O and CDMM-L (above 0.85). Mobility and Self-care also correlated strongly in CDMM-O (0.86) and in CDMM-L (0.70), altough the latter was on the verge of moderately.

In general, the correlation among observed dimensions, on the one hand, and among latent dimensions, on the other hand, was around the moderate-strong threshold (0.7). In CDMM-O the median value was 0.72; and in CDMM-L it was 0.68.

More details of the OICs and Specific Indirect effects can be found in C.2 and C.3.

# Chapter 6

# Discussion and Conclusions

In this master thesis the mediating role of disability in the impact of common conditions on perceived health has been assessed through a structural equation modeling approach. Two models have been considered: a model with all the variables observed, CDMM-O, and another with four of the eight mediators treated as latent variables, CDMM-L. The structural equations with latent variables models are more realistic in their allowance for measurement error in the observed variables, and therefore the relation between latent variables can be analized unobscured by measurement error.

Certainly, evidence of instability has been observed in CDMM-O: a negative proportion of indirect effect over total effect for Drug abuse, as well as inconsistent positive direct and specific indirect effects for Getting along have been obtained. Positive specific effects also exist for other dimensions, and in fact the positive specific indirect effect of Mobility on Drug abuse (0.31) is the main responsible for the overall positive indirect effect. Indeed, the negative value of the regression coefficient of Mobility on Drug abuse (not shown but available under request) suggests that the presence of Drug abuse diminishes (improves) the Mobility score. In spite of being an exception in the set of associations among conditions and disability, this is in line with the result from the descriptive analysis that the mean of Mobility is lower for individuals suffering from Drug abuse (0.85) than the overall Mobility mean (0.89). CDMM-L also provides a positive specific indirect effect of Mobility for Drug abuse (0.20), but the negative specific indirect effects through the other dimensions make up for it, and the overall indirect effect is negative, a more realistic fact.

This particular situation suggests that the Mobility model might be inadequate in the Drug abuse population. Certainly, Mobility is not usually a problem for individuals with substance disorders. Conversely, other symptoms relating with the condition (and other comorbid behavioral disorders) might be explaining their response to mobility items, such as restlessness, psychomotor agitation or anxiety-related symptoms (APA, 2000).

With regard to the inconsistencies found for Getting along dimension, in the descriptive analysis we saw that all the dimensions were negatively correlated to Perceived health, which is consistent with the fact that the higher the disability sum score, the lower the perceived health. However, a positive (but not significant) direct effect for Getting along in CDMM-O was obtained, meaning that the higher the disability score, the higher the VAS score, fact that makes no sense, and it is inconsistent with the negative correlation coefficient among Getting along and Perceived health (-0.26). Moreover, this positive direct effect is the responsible for the positive specific indirect effects (negative Specific Contributions, SC) for Getting along. This is not the case of CDMM-L. Moreover, the negative correlation values involving Getting along for CDMM-O (with the overall indirect effect, with its latent counterpart, and with the other observed specific effects) are also inconsistent.

In contrast, CDMM-L does not present the inconsistencies found for CDMM-O. In addition, it shows a better fit. The goodness of fit indices CFI and TLI are not acceptable for CDMM-O (Hu and Bentler, 1999) but they are satisfactory for CDMM-L. RMSEA is valid for both models, and the proportion of variance explained is significantly higher for CDMM-L (0.43 versus 0.39), as expected (Dhrymes, 1978). Explaining a 4% more of variability is a substantial amount if we take into account that we do not compare nested models. In social sciences where self-reported data are frequent, $R^2$ values above 0.14 are considered large, and they are rarely found as the product of a single predictor (Cohen, 1988). There are a number of factors intervening in what we want to study, and it can be difficult to measure or to take into account all of them. Hence, the proportion of variance explained for both models can be considered large. However, a future line of research would be to determine additional factors to include in the model in order to increase the proportion of unexplained variance.

Another advantage of using latent variables is that factors do not have floor/ceiling effects, while in this study observed sum scores were clearly affected by floor effect. As factor scores have an unrestricted range, they can get any value from $-\infty$ to $\infty$, including those values not considered in the observed sum score range. In the case of factor models, there is a score value for each of the possible $Ncategories^{Nitems}$ response patterns, which widens the range of the function score beyond that of the observed sum score.

Standard errors for direct effects, indirect effects, and Overall Indirect Contributions (OICs) for chronic conditions were always lower for CDMM-L than for CDMM-O. In the case of direct effects of mediators, the standard errors for the latent mediators were higher for CDMM-L, while for the observed mediators they kept consistently lower for CDMM-L. In the field of PRO, where small effects are obtained, the maximum precision in parameters is desired, and the pure relations among latent factors contribute to enlarge this precision.

In turn, the parameters were more distinct among latent and observed models than expected; the mediating role of disability was more important in CDMM-L because more than half of the total effect was mediated by dimensions. In contrast, less than half of the total effect was mediated according to CDMM-O. Nevertheless, the total, direct, and indirect effects were more correlated among observed and latent models than their respective standard errors, showing a greater correspondence among effects than among standard errors.

An important strength of SEM methodology is that collinearity among predictors can be taken into account. Multicollinearity is the extent to which a linear dependence exists between an explanatory variable and the other explanatory variables in an equation. Collinearity generally increases the standard errors of the coefficients of the collinear variables, and so it exists a greater uncertainty in the inferences that we make about the parameters (Bollen, 1989). Once associations among factors are defined in the model, collinearity is not an issue. However, when those associations are not defined and they actually exist, serious consequences can arise. In particular, CDMM-L was previously run without considering factor associations: convergence problems arose and inconsistent estimates were obtained, such as positive direct effects on perceived health for almost all the conditions. Therefore, the possibility of accounting for collinearity was crucial for CDMM-L to provide meaningful estimates. Apart from that, CDMM-L was more sensitive to collinearity than CDMM-O, because the latent factors were more correlated among them than the corresponding observed sum scores, as expected: factors do not contain measurement error, and all the relation existing among factors is considered in the model; on the other hand, the random errors present in observed scores do not correlate, and therefore the degree of association in the model is lower.

As we have shown here, SEM can handle complex mediated models, considering multiple predictors and mediators (and even more than one outcome if necessary) simultaneously. Total, direct, indirect, specific indirect effects, and OICs, with their standard errors, can be explicitly obtained, which would be considerably ackward if one tried to compute those estimates with single paths. Moreover, with single paths, effects could not be simultaneously taken into account, and asymptotic standard errors could not be obtained.

Our findings highlight that disability mediates the impact of common conditions on perceived health. However, variation across individual disorders exists in the extent to which their impact on perceived health is mediated by disability dimensions, ranging from a non significant 17% for Drug abuse to 96% for Cancer. However, a higher decrement in perceived health (total effect) does not necessarily imply a higher contribution of disability on this total effect. For instance, Neurological, the condition with the highest total effect, has 56% of its total effect mediated by disability dimensions, much lower than the 96% of

Cancer which, in turn, is the condition with the second to last total effect. Bipolar is the condition with both an important total effect (-7.4) and OIC (69%).

On average, the disability mediated effect on perceived health is substantial (mean OIC=58%, median OIC =57%), and similar for the 9 mental disorders and the 10 physical conditions. Nevertheless, the specific disability dimension which mediates such effect tends to be different for physical and for mental conditions.

Role functioning is the predominant mediator of conditions on perceived health. Mobility, the second most important mediator, is a frequent mediator of physical conditions on perceived health (median value of 18.9%), while it is much less important for mental conditions (median value of 10%). Many of the physical conditions considered in the study imply either pain (arthritis, chronic pain) or impairment on the extremities and their functional performance (neurological conditions, cardiovascular, respiratory) or general weakness (cancer and others). All of them have an impact on the mobility function and modify the perception of health of the individual (Alonso et al., 2004, 2011; Garin et al., 2010). On the other hand, this dimension is not a very relevant mediator of the impact of mental conditions on perceived health, while Cognition (third most important mediator), Family burden, and Stigma are. In particular, the mediating role of Cognition on Depression has been assessed (Buist-Bouwman et al., 2008; Knouse et al., 2013). In line with that, we found that Cognition contributes the 19% on the overall indirect effect of Depression on perceived health, just after Role functioning (31.4%). Therefore, our results also indicate that addressing Cognition should help to ameliorate the perceived health of individuals suffering from Depression.

It is important to highlight that the latent model gave much more importance to the specific contribution of Cognition on the mediation role: it is considered the third most important mediator while the observed model placed it in the fifth position, after Family burden and Stigma. As stated in the previous paragraph, in the literature there is evidence of Cognition as a mediator of mental conditions and health outcomes; therefore, the classification made for the latent model seems more consistent with what is theoretically known.

Nevertheless, the results must be interpreted considering the following limitations. First, only two estimators, ML for CDMM-O and WLSMV for CDMM-L, have been used, while there are a variety of them: these estimators aim at reproducing the variance covariance matrix of the variables. A line of further research would be to use other estimators (Item Response Theory [IRT] via Full Information Maximum Likelihood [FIML]) aimed at reproducing the observed data and obtaining of individual scores. However, the objective of this project was to assess the associations among variables, not to make assessment and predictions on individuals. Second, only four variables could be treated as latent, and

therefore a reduced number of variables could be analyzed without measurement error. Third, chronic physical conditions and mental disorders were differently assessed: physical conditions were self-reported, and therefore underreporting could be present. Finally, only 12-month physical and mental conditions were considered, to increase the accuracy of recalls, while perceived health (VAS) and WHODAS questions referred to the 30 days preceding the interview. Due to different time frames it is not possible to definitively relate either the health statuts nor the disability reported by the respondents to their underlying mental or physical health condition for the preceding 12 months. Nevertheless, as VAS and WHODAS use the same recall period, any such bias should not influence the analyses of the intermediating role of disability in the impact of conditions on perceived health. Similarly, the duration of the disability was not possible to be assessed.

**Implications**

Regarding the comparison of both approaches, a model with latent variables is preferred: benefits of assessing pure relations, without measurement error, were observed even treating a few number (4) of variables as latent: CDMM-L corrected the inconsistencies present in CDMM-O, and more precise estimates —of utmost importance in PRO area—, were obtained. Nevertheless, one has to bear in mind that the WHODAS questionnaire is highly reliable (Garin et al., 2010; Noonan et al., 2010), fact that attenuates the differences among latent and observed approaches; with a less reliable questionnaire, even broader differences would have been obtained.

In general, the results from both the CDMM-O and CDMM-L call attention on the need to assess and consider disability to better understand how perceived health is influenced by common mental and physical conditions. According to CDMM-L, more than a half of the decrements in perceived health are mediated by disability dimensions and would not be a direct effect to these conditions. Both CDMM-O and CDMM-L highlight the importance of addressing disability to increase health status among individuals with common conditions: there is a need to learn more about the strength and ways of indirect association between chronic conditions and perceived health. In particular, evaluating whether interventions addressed to improve specific disability areas may improve perceived health of individuals with common chronic conditions beyond benefits that would be obtained with the usual treatment for these conditions.

# Appendices

# Appendix A

# Questionnaire for Disability and Perceived health

| Variable | Item questions | Response scale |
|---|---|---|
| Cognition | How much difficulty did you have in<br>· Concentrating on doing something for ten minutes?<br>· Understanding what was going on around you?<br>· Remembering to do important things?<br>· Learning a new task<br>(for example, learning how to get to a new place)? | 5-point Likert type<br>with a score<br>from 0: no disability<br>to 16: complete disability |
| Getting along | How much difficulty did you have in<br>· Starting and maintaining a conversation?<br>· Dealing with people you did not know well?<br>· Maintaining friendships?<br>· Making new friends?<br>· Controlling your emotions when you were around people? | 5-point Likert type<br>with a score<br>from 0: no disability<br>to 20: complete disability |
| Mobility | How much difficulty did you have in<br>· Standing for long periods, such as 30 minutes?<br>· Moving around inside your home?<br>· Walking a long distance such as (a kilometer/half a mile)? | 5-point Likert type<br>with a score<br>from 0: no disability<br>to 12: complete disability |
| Self-care | How much difficulty did you have in<br>· Washing your whole body?<br>· Getting dressed?<br>· Staying by yourself for a few days? | 5-point Likert type<br>with a score<br>from 0: no disability<br>to 12: complete disability |
| Role functioning | How many days out of the past 30<br>· were you totally unable to work or carry out<br>your normal activities?<br>· were you able to work and carry out your<br>normal activities, but had to cut down on what you did<br>or not get as much done as usual?<br>· did you cut back on the quality of your work or<br>how carefully you worked because of problems with<br>either your physical health, your mental health,<br>or your use of alcohol or drugs?<br>· did it take an extreme effort to perform up to your<br>usual level at work or at your other normal daily activities<br>because of problems with either your physical health,<br>your mental health, or your use of alcohol or drugs? | Weighted number of days<br>from 0: no disability<br>to 30: complete disability |
| Stigma | During the past 30 days,<br>How much embarrassment did you experience because of<br>your health problems? | 5-point Likert type<br>with a score<br>from 0 to 4 |
| Discrimination | During the past 30 days,<br>How much discrimination or unfair treatment did you<br>experience because of your health problems? | 5-point Likert type<br>with a score<br>from 0 to 4 |
| Family burden | During the past 30 days,<br>How much did your health-related difficulties interfere<br>with the life and activities of your close friends and<br>family members? | 5-point Likert type<br>with a score<br>from 0 to 4 |
| Perceived health | During the past 30 days,<br>What number would you use to describe your own<br>overall physical and mental health? | Score from 0: worst<br>to 100: perfect<br>health status |

Table A.1: Item questions corresponding to disability dimensions and perceived health
The response scale used is for continuous observed variables.

# Appendix B

# General model with intercept terms

In the whole project, all random variables were assumed to have zero means. This assumption will now be relaxed and the model will be extended to include four new parameters in addition to the previous eight. These new parameter matrices contain an intercept term in the relationships and mean values of the latent variables.

The model is now defined by the following three equations

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{B.1}$$

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \epsilon \\ \boldsymbol{x} &= \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \end{aligned} \tag{B.2}$$

,

where $\boldsymbol{\alpha}, \boldsymbol{\nu}_y$, and $\boldsymbol{\nu}_x$ are vectors of constant intercept terms. As before, we assume that $\boldsymbol{\zeta}$ is uncorrelated with $\boldsymbol{\xi}$, $\boldsymbol{\epsilon}$ is uncorrelated with $\boldsymbol{\eta}$ and that $\boldsymbol{\delta}$ is uncorrelated with $\boldsymbol{\xi}$. We also assume, as before, that $E(\boldsymbol{\zeta}) = \boldsymbol{0}$, $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$, and $E(\boldsymbol{\delta}) = \boldsymbol{0}$. However, we will denote $E(\boldsymbol{\xi}) = \boldsymbol{\kappa}$, and deduce $E(\boldsymbol{\eta})$, $E(\boldsymbol{y})$, and $E(\boldsymbol{x})$ from equations (B.1) and (B.2):

$$\begin{aligned} E(\boldsymbol{\eta}) &= E[(\boldsymbol{I} - \boldsymbol{B})^{-1}(\alpha + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})] = (\boldsymbol{I} - \boldsymbol{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\kappa}) \\ E(\boldsymbol{y}) &= \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y(\boldsymbol{I} - \boldsymbol{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\kappa}) \\ E(\boldsymbol{x}) &= \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x\boldsymbol{\kappa} \end{aligned} \tag{B.3}$$

As the first equation in (B.3) shows, the mean of $\boldsymbol{\eta}$ is not only a function of $\boldsymbol{\kappa}$, the mean of the exogenous variables, but also a function of the structural parameters in $\boldsymbol{B}, \boldsymbol{\Gamma}$, and

$\boldsymbol{\alpha}$. Similarly, the mean of $\boldsymbol{y}$ is determined by these matrices as well as by $\boldsymbol{\nu}_y$ and $\boldsymbol{\Lambda}_y$. The expected value of $\boldsymbol{x}$ is influenced by $\boldsymbol{\nu}_x$, $\boldsymbol{\Lambda}_x$, and $\boldsymbol{\kappa}$.

In general, all the mean parameters $\boldsymbol{\nu}_y, \boldsymbol{\nu}_x, \boldsymbol{\alpha}$, and $\boldsymbol{\kappa}$ will not be identified without further conditions imposed.

## Appendix C

# Coefficients and SE for Total, Direct, Indirect and Specific Indirect effects, and OICs for CDMM-O and CDMM-L

| **Sociodemographic variable** | **Categories** | **Direct effects for CDMM-O** | **Direct effects for CDMM-L** |
|---|---|---|---|
| | | Coef (SE) | Coef (SE) |
| Age | | -0.15 (0.01)* | -0.17 (0.01)* |
| Country | Belgium | -3.20 (0.75)* | -4.46 (0.74)* |
| | France | -1.34 (0.63)* | -2.87 (0.69)* |
| | Germany | -4.88 (0.54)* | -5.38 (0.70)* |
| | Israel | -3.17 (0.34)* | -6.22 (0.42)* |
| | Italy | -4.79 (0.55)* | -5.10 (0.61)* |
| | Japan | -6.01 (0.53)* | -5.96 (0.67)* |
| | Netherlands | -0.76 (0.69) | -3.36 (0.94)* |
| | Northern Ireland | -0.09 (0.48) | -1.21 (0.55)* |
| | Portugal | -4.20 (0.66)* | -5.45 (0.61)* |
| | Spain | -5.06 (0.67)* | -5.90 (0.59)* |
| Employment status | Student | 0.54 (0.64) | 0.27 (1.21) |
| | Homemaker | -1.34 (0.53)* | -1.96 (0.53)* |
| | Retired | -3.22 (0.45)* | -4.33 (0.44)* |
| | Other | -5.33 (0.49)* | -9.99 (0.36)* |
| Marital status | Married/Cohabiting | 1.80 (0.33)* | 1.63 (0.39)* |
| | Separated/Widowed/ Divorced | 0.45 (0.53) | -0.49 (0.51) |
| Sex | Male | 0.10 (0.27) | 0.30 (0.30) |

Table C.1: Direct effects of sociodemographic variables

| Condition | Total effect | Direct effect | Indirect effect | OIC | Cognition | Getting along | Mobility | Self-care | Role functioning | Discri-mination | Family burden | Stigma |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coeff (SE) | Coeff (SE) | Coeff (SE) | %(SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) |
| Alcohol abuse | -1.5 (0.8) | -1.4 (0.7)* | 0.0 (0.4) | 2.0 (26.0) | -0.11 (0.06) | 0.01 (0.01) | 0.15 (0.10) | 0.05 (0.03) | -0.14 (0.19) | 0.01 (0.03) | 0.04 (0.08) | -0.03 (0.08) |
| Bipolar | -7.6 (1.5)* | -3.5 (1.2)* | -4.1 (1.0)* | 54.3 (10.8)* | -0.66 (0.18)* | 0.10 (0.14) | -0.50 (0.22)* | -0.17 (0.10) | -1.39 (0.35)* | -0.28 (0.16) | -0.64 (0.23)* | -0.58 (0.21)* |
| Depression | -8.2 (0.5)* | -5.3 (0.5)* | -2.9 (0.3)* | 34.8 (3.4)* | -0.34 (0.08)* | 0.05 (0.06) | -0.21 (0.07)* | -0.06 (0.03)* | -1.26 (0.13)* | -0.12 (0.06)* | -0.44 (0.09)* | -0.48 (0.11)* |
| Drug abuse | -2.8 (1.5) | -3.0 (1.2)* | 0.3 (0.7) | -9.2 (28.6) | -0.01 (0.08) | 0.02 (0.04) | 0.31 (0.12)* | 0.00 (0.05) | -0.11 (0.31) | -0.05 (0.07) | -0.03 (0.15) | 0.12 (0.14) |
| Generalized anxiety | -5.1 (1.1)* | -2.2 (0.9)* | -2.9 (0.7)* | 56.8 (12.1)* | -0.35 (0.11)* | 0.06 (0.08) | -0.43 (0.18)* | -0.12 (0.07) | -0.99 (0.29)* | -0.10 (0.06) | -0.56 (0.16)* | -0.44 (0.15)* |
| Panic disorder | -6.0 (1.0)* | -3.2 (0.9)* | -2.8 (0.5)* | 46.5 (8.7)* | -0.31 (0.09)* | 0.06 (0.09) | -0.27 (0.13)* | -0.13 (0.07) | -1.19 (0.23)* | -0.05 (0.05) | -0.41 (0.14)* | -0.51 (0.15)* |
| Posttraumatic stress | -4.7 (1.0)* | -0.8 (0.8) | -4.0 (0.8)* | 84.0 (15.2)* | -0.44 (0.13)* | 0.07 (0.10) | -0.72 (0.18)* | -0.17 (0.10) | -1.56 (0.30)* | -0.10 (0.06) | -0.58 (0.19)* | -0.46 (0.16)* |
| Social phobia | -2.4 (0.8)* | -1.3 (0.8) | -1.1 (0.4)* | 44.1 (17.7)* | -0.24 (0.07)* | 0.04 (0.05) | -0.05 (0.1) | 0.01 (0.03) | -0.30 (0.13)* | -0.04 (0.04) | -0.26 (0.11)* | -0.21 (0.09)* |
| Specific phobia | -2.7 (0.6)* | -1.6 (0.6)* | -1.1 (0.4)* | 41.8 (12.4)* | -0.05 (0.03) | 0.00 (0.01) | -0.24 (0.10)* | 0.00 (0.03) | -0.49 (0.17)* | -0.03 (0.03) | -0.20 (0.08)* | -0.15 (0.07)* |
| Arthritis | -4.8 (0.5)* | -2.7 (0.4)* | -2.1 (0.2)* | 43.6 (4.5)* | -0.04 (0.02)* | 0.00 (0.01) | -0.81 (0.1)* | -0.12 (0.04)* | -0.75 (0.11)* | -0.02 (0.02) | -0.14 (0.04)* | -0.22 (0.05)* |
| Cancer | -1.9 (0.9)* | -0.1 (0.8) | -1.8 (0.4)* | 93.6 (39.9)* | -0.03 (0.04) | 0.01 (0.02) | -0.51 (0.15)* | -0.06 (0.04) | -0.78 (0.18)* | -0.02 (0.03) | -0.29 (0.09)* | -0.13 (0.07) |
| Cardiovascular | -5.3 (0.4)* | -3.2 (0.4)* | -2.1 (0.2)* | 39.8 (4.0)* | -0.05 (0.02)* | 0.01 (0.01) | -0.64 (0.09)* | -0.14 (0.06)* | -0.74 (0.11)* | -0.04 (0.02) | -0.28 (0.07)* | -0.24 (0.05)* |
| Chronic pain | -6.7 (0.4)* | -3.1 (0.3)* | -3.6 (0.2)* | 54.0 (3.1)* | -0.12 (0.03)* | 0.02 (0.02) | -1.06 (0.11)* | -0.16 (0.05)* | -1.50 (0.12)* | -0.06 (0.03) | -0.38 (0.07)* | -0.35 (0.07)* |
| Diabetes | -5.6 (0.8)* | -3.4 (0.7)* | -2.2 (0.4)* | 39.7 (6.3)* | -0.10 (0.04)* | 0.01 (0.02) | -0.58 (0.13)* | -0.15 (0.05)* | -0.98 (0.19)* | -0.06 (0.04) | -0.18 (0.07)* | -0.18 (0.07)* |
| Digestive | -7.0 (1.1)* | -3.2 (1.0)* | -3.8 (0.5)* | 54.7 (8.1)* | -0.16 (0.05)* | 0.03 (0.04) | -0.80 (0.16)* | -0.16 (0.06)* | -1.57 (0.24)* | -0.09 (0.06) | -0.52 (0.14)* | -0.54 (0.14)* |
| Headaches/ Migraines | -4.2 (0.4)* | -2.3 (0.4)* | -1.9 (0.2)* | 44.8 (5.4)* | -0.25 (0.06)* | 0.03 (0.04) | -0.37 (0.07)* | -0.08 (0.03)* | -0.65 (0.11)* | -0.06 (0.04) | -0.30 (0.07)* | -0.20 (0.05)* |
| Insomnia | -6.1 (0.6)* | -2.7 (0.6)* | -3.5 (0.3)* | 56.3 (5.8)* | -0.29 (0.07)* | 0.04 (0.05) | -0.78 (0.10)* | -0.20 (0.06)* | -1.20 (0.16)* | -0.12 (0.06)* | -0.52 (0.11)* | -0.38 (0.10)* |
| Neurological | -11.1 (1.4)* | -5.4 (1.2)* | -5.7 (1.1)* | 51.4 (8.5)* | -0.34 (0.12)* | 0.06 (0.08) | -1.36 (0.34)* | -0.48 (0.19)* | -1.87 (0.41)* | -0.27 (0.15) | -0.65 (0.24)* | -0.8 (0.25)* |
| Respiratory | -0.6 (0.3) | -0.4 (0.3) | -0.2 (0.1) | 30.3 (22.5) | -0.05 (0.02)* | 0.00 (0.00) | -0.12 (0.05)* | 0.01 (0.02) | -0.08 (0.07) | 0.01 (0.01) | 0.03 (0.03) | 0.02 (0.02) |

*p.value $< 0.05$.

Table C.2: Total, direct and indirect effects, OICs and specific indirect effects of conditions for CDMM-O.
OIC=Overall Indirect Contribution, percentage of indirect over total effect.
The model was adjusted for Age, Country, Employment status, Marital status, and Sex.

| Condition | Total effect | Direct effect | Indirect effect | OIC | Cognition | Getting along | Mobility | Self-care | Role functioning | Discri-mination | Family burden | Stigma |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coeff (SE) | Coeff (SE) | Coeff (SE) | %(SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) | Coeff (SE) |
| Alcohol abuse | -1.8 (0.8)* | -1.0 (0.8) | -0.8 (0.5) | 42.7 (29.1) | -0.31 (0.11)* | -0.02 (0.05) | -0.12 (0.18) | 0.00 (0.15) | -0.25 (0.14) | 0.01 (0.05) | -0.01 (0.06) | -0.07 (0.05) |
| Bipolar | -7.4 (0.8)* | -2.3 (0.7)* | -5.1 (0.5)* | 69.4 (7.1)* | -0.78 (0.18)* | -0.14 (0.22) | -0.51 (0.20)* | -0.38 (0.17)* | -1.32 (0.14)* | -0.87 (0.14)* | -0.60 (0.07)* | -0.53 (0.06)* |
| Depression | -7.9 (0.4)* | -4.5 (0.4)* | -3.4 (0.2)* | 43.2 (3.0)* | -0.66 (0.15)* | -0.12 (0.17) | -0.26 (0.09)* | -0.23 (0.10)* | -1.07 (0.07)* | -0.33 (0.05)* | -0.38 (0.04)* | -0.38 (0.04)* |
| Drug abuse | -4.4 (1.2)* | -3.6 (1.0)* | -0.7 (0.8) | 16.8 (15.8) | -0.12 (0.15) | -0.07 (0.11) | 0.20 (0.23) | -0.33 (0.26) | -0.15 (0.23) | -0.20 (0.07)* | -0.11 (0.07) | 0.04 (0.08) |
| Generalized anxiety | -4.7 (0.6)* | -1.8 (0.6)* | -2.9 (0.4)* | 61.5 (9.0)* | -0.45 (0.12)* | -0.09 (0.14) | -0.23 (0.16) | -0.22 (0.12) | -0.83 (0.12)* | -0.25 (0.05)* | -0.47 (0.05)* | -0.33 (0.04)* |
| Panic disorder | -5.8 (0.6)* | -3.2 (0.6)* | -2.6 (0.4)* | 44.8 (5.9)* | -0.31 (0.09)* | -0.07 (0.11) | -0.11 (0.12) | -0.15 (0.11) | -1.01 (0.11)* | -0.15 (0.04)* | -0.36 (0.05)* | -0.45 (0.05)* |
| Posttraumatic stress | -4.5 (0.6)* | -0.3 (0.5) | -4.2 (0.4)* | 92.4 (10.5)* | -0.52 (0.13)* | -0.10 (0.15) | -0.55 (0.15)* | -0.37 (0.16)* | -1.37 (0.11)* | -0.31 (0.05)* | -0.53 (0.06)* | -0.41 (0.05)* |
| Social phobia | -3.0 (0.7)* | -0.7 (0.6) | -2.3 (0.3)* | 76.3 (17.0)* | -0.47 (0.11)* | -0.10 (0.15) | -0.27 (0.13)* | -0.17 (0.11) | -0.45 (0.08)* | -0.20 (0.04)* | -0.33 (0.04)* | -0.28 (0.04)* |
| Specific phobia | -3.3 (0.5)* | -1.0 (0.5)* | -2.2 (0.3)* | 68.5 (10.7)* | -0.28 (0.08)* | -0.03 (0.05) | -0.51 (0.13)* | -0.22 (0.10)* | -0.63 (0.10)* | -0.15 (0.04)* | -0.26 (0.04)* | -0.17 (0.03)* |
| Arthritis | -4.7 (0.4)* | -2.3 (0.3)* | -2.4 (0.2)* | 50.8 (4.2)* | -0.15 (0.05)* | -0.04 (0.06) | -0.77 (0.14)* | -0.30 (0.11)* | -0.64 (0.06)* | -0.12 (0.03)* | -0.14 (0.03)* | -0.21 (0.03)* |
| Cancer | -1.7 (0.6)* | -0.1 (0.6) | -1.6 (0.4)* | 95.7 (33.5)* | -0.05 (0.08) | -0.05 (0.07) | -0.32 (0.15)* | -0.06 (0.10) | -0.67 (0.09)* | -0.08 (0.04) | -0.25 (0.04)* | -0.12 (0.03)* |
| Cardiovascular | -4.7 (0.3)* | -3.0 (0.3)* | -1.7 (0.2)* | 36.6 (4.5)* | -0.07 (0.05) | -0.03 (0.04) | -0.42 (0.10)* | -0.19 (0.09)* | -0.55 (0.07)* | -0.12 (0.03)* | -0.20 (0.03)* | -0.16 (0.02)* |
| Chronic pain | -6.4 (0.3)* | -2.2 (0.3)* | -4.2 (0.2)* | 65.7 (3.2)* | -0.36 (0.08)* | -0.06 (0.08) | -1.21 (0.20)* | -0.51 (0.16)* | -1.32 (0.07)* | -0.16 (0.03)* | -0.33 (0.03)* | -0.29 (0.03)* |
| Diabetes | -4.9 (0.5)* | -3.2 (0.4)* | -1.7 (0.3)* | 34.1 (5.3)* | -0.15 (0.08)* | -0.02 (0.03) | -0.29 (0.11)* | -0.19 (0.10) | -0.70 (0.09)* | -0.13 (0.03)* | -0.09 (0.03)* | -0.10 (0.03)* |
| Digestive | -6.4 (0.7)* | -3.1 (0.6)* | -3.3 (0.3)* | 51.1 (5.3)* | -0.2 (0.08)* | -0.04 (0.07) | -0.46 (0.13)* | -0.18 (0.11) | -1.33 (0.10)* | -0.24 (0.04)* | -0.41 (0.05)* | -0.42 (0.05)* |
| Headaches/ Migraines | -4.0 (0.3)* | -1.7 (0.3)* | -2.3 (0.2)* | 57.2 (5.6)* | -0.51 (0.11)* | -0.07 (0.11) | -0.39 (0.09)* | -0.22 (0.09)* | -0.52 (0.06)* | -0.16 (0.03)* | -0.26 (0.03)* | -0.15 (0.02)* |
| Insomnia | -5.8 (0.4)* | -2.0 (0.4)* | -3.8 (0.2)* | 65.3 (4.7)* | -0.49 (0.11)* | -0.08 (0.13) | -0.65 (0.12)* | -0.40 (0.14)* | -1.04 (0.07)* | -0.36 (0.06)* | -0.45 (0.04)* | -0.32 (0.04)* |
| Neurological | -10.5 (0.8)* | -4.6 (0.7)* | -5.9 (0.5)* | 56.1 (4.7)* | -0.53 (0.15)* | -0.09 (0.13) | -0.98 (0.25)* | -0.68 (0.23)* | -1.58 (0.15)* | -0.83 (0.13)* | -0.55 (0.07)* | -0.66 (0.07)* |
| Respiratory | -0.9 (0.3)* | -0.3 (0.3) | -0.6 (0.2)* | 70.9 (27.2)* | -0.18 (0.05)* | -0.02 (0.03) | -0.19 (0.07)* | 0.03 (0.06) | -0.18 (0.05)* | -0.01 (0.02) | -0.02 (0.02) | -0.03 (0.02)* |

*p.value $< 0.05$.

Table C.3: Total, direct and indirect effects, OICs and specific indirect effects of conditions for CDMM-L.
OIC=Overall Indirect Contribution, percentage of indirect over total effect.
The model was adjusted for Age, Country, Employment status, Marital status, and Sex.

# Appendix D

# SUDAAN, MPLUS, and R code

## D.1 SUDAAN

### Preprocessing

```
data dtset_final;
set dtset_tfm;

/*** Observed sum score for all indicators ***/
sindis0=scog0+smove0+scare0+ssoci0;

/*** Create dichotomous variables for affected individuals ***/
if scog0>0 then cogn0=1; else cogn0=0;
if smove0>0 then moven0=1; else moven0=0;
if scare0>0 then caren0=1; else caren0=0;
if ssoci0>0 then socin0=1; else socin0=0;
if outrol30>0 then outroln0=1; else outroln0=0;
if sstig0>0 then stign0=1; else stign0=0;
if sdiscr0>0 then discrin0=1; else discrin0=0;
if sfambu0>0 then fambun0=1; else fambun0=0;

if sindis0>0 then indisn0=1; else indisn0=0;

run;

proc sort data=dtset_final ; by str secu; run;

proc format;

 /*** Format of countries ***/
 value ctrf 1='Belgium' 2='France' 3='Germany' 4='Israel' 5='Italy' 6='Japan' 7='Netherlands'
 8='N.Ireland' 9='Portugal' 10='Spain' 11='United States' 12='All countries'
 13='Comparison among countries';
```

```
 /*** Format of dimensions and perceived health ***/
 value dimf 1='Cognition' 2='Getting along' 3='Mobility' 4='Self-care' 5='Role functioning'
 6='Discrimination' 7='Family burden' 8='Stigma' 9='Perceived health';

 /*** Format of dimensions and all indicators ***/
 value dimif 1='Cognition' 2='Getting along' 3='Mobility' 4='Self-care' 5='Role functioning'
 6='Discrimination' 7='Family burden' 8='Stigma' 9='Indis' 10='VAS';

 /*** Format of chronic conditions ***/
 value ccf 1='Alcohol abuse' 2='Bipolar' 3='Depression' 4='Drug abuse' 5='Generalized anxiety'
 6='Panic disorder' 7='Posttraumatic stress' 8='Social Phobia' 9='Specific phobia'
 10='Arthritis' 11='Cancer' 12='Cardiovascular' 13='Chronic pain' 14='Diabetes'
 15='Digestive' 16='Headaches/Migraines' 17='Insomnia' 18='Neurological' 19='Respiratory';

run;
```

## Sociodemographics table

```
/*** macro to create the demographic tables by country ***/
/*** catm: category for marital status ***/
/*** cate: category for employment status ***/

\%macro tabledem (catm,cate);
 \%let vars=sex marcat emp age;
 \%let levs=2 3 5 0;

 \%do i=1 \%to 4;

   \%let var=\%scan(&vars,&i);
   \%let lev=\%scan(&levs,&i);

   /*** For categorical variables: Prevalence and SE by means of crosstab ***/
   \%if &i ne 4 \%then \%do;

    /*** Obtain Wald test for difference in the variable across countries ***/
      proc crosstab data=dtset_final design=wr;
       nest str secu/missunit;
       weight finalp2wt;
       subgroup &var Countryhi;
       levels   &lev    11;
       table &var*countryhi;
       subpopn chronic_notasked=0;
       test chisq;
       output colper secol nsum/filename=out&i filetype=sas replace;
       output stestval sdf spval/filename=p&i filetype=sas replace;
      run;

     \%if &i=1 \%then \%do;
      data pp&i._1 (keep= var&i countryhi);
```

```
   set p&i;
   length var&i \$ 25;
   if spval<.0001 then var&i=put(stestval,7.1)||" (<0.0001)";
    else var&i=put(stestval,7.1)||' ('||put(spval,5.1)||')'; countryhi=13;
  run;
 \%end;

\%if &i=2 \%then \%do;
  data pp&i._1 (keep= var&i._&catm countryhi);
   set p&i;
   length var&i._&catm \$ 25;
   if spval<.0001 then var&i._&catm =put(stestval,7.1)||" (<0.0001)";
    else var&i._&catm =put(stestval,7.1)||' ('||put(spval,5.1)||')'; countryhi=13;
   run;
 \%end;

\%if &i=3 \%then \%do;
  data pp&i._1 (keep= var&i._&cate countryhi);
   set p&i;
   length var&i._&cate \$ 25;
   if spval<.0001 then var&i._&cate=put(stestval,7.1)||" (<0.0001)";
    else var&i._&cate =put(stestval,7.1)||' ('||put(spval,5.1)||')'; countryhi=13;
   run;
\%end;

data out&i._1;
 set out&i;
  \%if &i=1 \%then \%do;
    length var&i \$ 25;
    if &var=2;
    var&i=put(colper,5.1)||' ('||put(secol,4.1)||')';
  \%end;

  \%if &i=2 \%then \%do;
    length var&i._&catm \$ 25;
    if &var=&catm;
    var&i._&catm=put(colper,5.1)||' ('||put(secol,4.1)||')';
  \%end;

  \%if &i=3 \%then \%do;
    length var&i._&cate \$ 25;
    if &var=&cate;
    var&i._&cate=put(colper,5.1)||' ('||put(secol,4.1)||')';
  \%end;
run;

data col&i;
 set out&i._1 pp&i._1;
run;

proc sort data=col&i; by countryhi; run;
```

```
/* end if i ne 4 */
\%end;

  \%else \%do;

  /*** For the continuous variable: Mean and SE by means of descript ***/
    proc descript data=dtset_final design=wr;
     nest str15 secu/missunit;
     weight finalp2wt;
     class countryhi;
     var &var;
     subpopn chronic_notasked=0 ;
     output mean semean nsum/filename=out&i filetype=sas replace;
    run;

   /*** Obtain Wald test for difference in age across countries ***/
    proc regress data=dtset_final design=wr;
     nest str15 secu/missunit;
     weight finalp2wt;
     subgroup countryhi;
     levels   11;
     model &var = countryhi;
     subpopn chronic_notasked=0 ;
     output waldf df waldchp/filename=p_mean&i filetype=sas replace;
    run;

    data p_mean&i._2 (keep= p_0);
     set p_mean&i nobs=last;
     length p_0 \$ 25;
     if waldchp<.0001 then p_0=put(waldf,7.1)||' (<0.0001)';
     else p_0=put(waldf,7.1)||' ('||put(waldchp,5.1)||')';
     if _n_=last;
    run;

   data out&i._1;
    set out&i;
    length var&i \$ 25;
    \%if &i=4 \%then \%do; num=put(nsum,5.0); \%end;
     var&i=put(mean,5.1)||' ('||put(semean,4.1)||')';
   run;

   data p&i._1;
    set p_mean&i._2;
    var&i=p_0;
    countryhi=13;
   run;

   data col&i;
    set out&i._1 p&i._1 ;
   run;

   proc sort data=col&i; by countryhi; run;
```

```
  /* end else do */
   \%end;

/* end do for i=1 to 4 */
 \%end;

   data table_1_&cate;
    merge col1 col2 col3 col4;
    by countryhi;
   run;

   data table_1_&cate;
    length label \$35;
    set table_1_&cate;
    if countryhi=0 then countryhi=12;
    label=put(countryhi,ctrf.);
   run;

   proc sort data=table_1_&cate; by countryhi;run;
\%mend tabledem;

/*** Execute macro for each of the categories of marital status and employment status ***/
\%tabledem (1,1);
\%tabledem (2,2);
\%tabledem (3,3);
\%tabledem (1,4);
\%tabledem (1,5);

data table_1;
merge table_1_1 table_1_2 table_1_3 table_1_4 table_1_5; by countryhi; run;

/*** Export to RTF ***/
ods RTF file="sociodem_&sysdate..rtf";
title "Sociodemographics";
proc report nowd data=table_1 headline headskip split="*";
        column label  num var4 var3_1 var3_2 var3_3 var3_4 var3_5;
        define  label /width=25 left "Country";
        define  countryhi /width=25 left "Ctr code";
        define  num /width=25  " N ";
        define  var4 /width=25  "Age*Mean (SE)";
        define  var3_1 /width=25  "Working*\% (SE)";
        define  var3_2 /width=25  "Student*\% (SE)";
        define  var3_3 /width=25  "Homemaker*\% (SE)";
        define  var3_4 /width=25  "Retired*\% (SE)";
        define  var3_5 /width=25  "Other*\% (SE)";
run;

proc report nowd data=table_1 headline headskip split="*";
        column label countryhi num var2_1 var2_2 var2_3 var1;
        define  label /width=25 left "Country";
        define  countryhi /width=25 left "Ctr code";
```

```
        define  num /width=25  " N ";
        define  var2_1 /width=25  "Married/Cohabiting*\% (SE)";
        define  var2_2 /width=25  "Separated/Widowed/Divorced*\% (SE)";
        define  var2_3 /width=25  "Never married*\% (SE)";
        define  var1 /width=25  "Females*\% (SE)";
        run;

ODS RTF Close;
title "";
```

---

## Percentage of indicators on the latent variable

---

```
/*** Obtain percentages for each set of indicators ***/
/*** Cognition ***/
proc crosstab data=dtset_final design=wr;
 nest str secu /missunit ;
 weight finalp2wt;
 class fd11a0 fd11b0 fd11c0 fd11d0;
 tables fd11a0 fd11b0 fd11c0 fd11d0;
 output nsum rowper serow / filename = cog replace;
run;

/*** Getting along ***/
proc crosstab data=dtset_final design=wr;
 nest str secu /missunit ;
 weight finalp2wt;
 class fd17a0 fd17b0 fd17c0 fd17d0 fd17e0;
 tables fd17a0 fd17b0 fd17c0 fd17d0 fd17e0;
 output nsum rowper serow / filename = soci replace;
run;

/*** Mobility ***/
proc crosstab data=dtset_final design=wr;
 nest str secu /missunit ;
 weight finalp2wt;
 class fd13a0 fd13b0 fd13c0;
 tables fd13a0 fd13b0 fd13c0;
 output nsum rowper serow / filename = move replace;
run;

/*** Self-care ***/
proc crosstab data=dtset_final design=wr;
 nest str secu /missunit ;
 weight finalp2wt;
 class fd15a0 fd15b0 fd15c0;
 tables fd15a0 fd15b0 fd15c0;
 output nsum rowper serow / filename = care replace;
run;
```

```
data cog;
 set cog;
 name='cog';
run;

data soci;
 set soci;
 name='soci';
run;

data move;
 set move;
 name='move';
run;

data care;
 set care;
 name='care';
run;

/*** Join the previous tables ****/
data indis;
 set cog soci move care;
run;

data indis (keep=rowper name);
 set indis;
run;

/*** Export to RTF ***/
\%macro indis;
options orientation=landscape;
ods RTF file="Proportion_indicators_&sysdate..rtf";
  title "Proportion of indicators";
  proc report nowd data=indis headline headskip split="*";
run;
ODS RTF Close;
title "";
options orientation=portrait;
\%mend indis;
\%indis;
```

## Descriptives of WHODAS and VAS

```
\%macro table1_meanwho(weight, indata);

/*** dimensions for affected individuals ***/
\%let dimn0=  cogn0 socin0 moven0 caren0 outroln0 discrin0 fambun0 stign0 indisn0;
/*** dimensions for all individuals ***/
\%let dim=  scog0 ssoci0 smove0 scare0 outrol30 sdiscr0 sfambu0 sstig0 sindis0;
```

```
/*** macro variable for VAS ***/
\%let va=newfd23d_final;

\%do i=1 \%to 9;
/*** v dimension variable for affected individuals ***/
    \%let v=\%scan(&dimn0,&i);
/*** v1 dimension variable for all individuals ***/
    \%let v1=\%scan(&dim,&i);

/*** Prevalence of affected individuals ***/
    proc crosstab data = &indata design = wr;
     nest str secu /missunit ;
     weight &weight;
     class &v;
     tables &v;
     output nsum rowper serow / filename = preval_&v replace;
    run;

/*** Mean of dimension among affected individuals ***/
    proc descript data = &indata design = wr;
     nest str secu /missunit ;
     weight &weight;
     var  &v1;
     subpopn &v=1;
     output nsum mean semean/ filename=mean_&v filetype=SAS replace;
    run;

    data _preval_&v.;
     set preval_&v.;
     dim=put(&i,dimif.);
     prev_se=put(rowper,5.2)||" ("||put(serow,5.2)||")";
     if _n_=3;
    run;

    data _mean_&v.;
     set mean_&v.;
     dim=put(&i,dimif.);
     meann0_se=put(mean,5.2)||" ("||put(semean,5.2)||")";
     if _n_=2;
    run;

/*** Mean of dimension among all individuals ***/
    proc descript data = &indata design = wr;
     nest str secu /missunit ;
     weight &weight;
     var  &v1;
     output nsum mean semean/ filename=mean_&v1 filetype=SAS replace;
    run;

    data _mean_&v1.;
     set mean_&v1.;
     dim=put(&i,dimif.);
```

```
     mean_se=put(mean,5.2)||" ("||put(semean,5.2)||")";
     if _n_=2;
    run;

/*** Mean of VAS among affected individuals ***/
    proc descript data = &indata design = wr;
     nest str secu /missunit ;
     weight &weight;
     var  &va;
     subpopn &v=1;
     output nsum mean semean/ filename=mean_&v._&va filetype=SAS replace;
    run;

    data _mean_&v._&va.;
     set mean_&v._&va.;
     dim=put(&i,dimif.);
     mean_sevas=put(mean,5.2)||" ("||put(semean,5.2)||")";
     if _n_=2;
    run;

\%end;

\%mend table1_meanwho;

\%table1_meanwho(finalp2wt,dtset_final);

/*** Join the previous 9 tables ***/
\%macro join;

data _descri_whono0 (keep= dim prev_se meann0_se mean_se mean_sevas);
 \%let dimn0=  cogn0 socin0 moven0 caren0 outroln0 discrin0 fambun0 stign0 indisn0;
 \%let dim=  scog0 ssoci0 smove0 scare0 outrol30 sdiscr0 sfambu0 sstig0 sindis0 ;
 \%let va=newfd23d_final;
 merge
  \%do i=1 \%to 9;
   \%let v=\%scan(&dimn0,&i);
   \%let v1=\%scan(&dim,&i);
   _preval_&v. _mean_&v. _mean_&v1. _mean_&v._&va.
 \%end;
 ;
 by dim;
run;

\%mend join;
\%join;

/*** Obtain mean and SE for VAS for all individuals ***/
proc descript data = dtset_final design = wr;
 nest str secu /missunit ;
 weight finalp2wt ;
 var  newfd23d_final;
 output nsum mean semean/ filename=mean_vas filetype=SAS replace;
```

```
run;

/*** Export to RTF ***/
\%macro rtfdes;
options orientation=landscape;
ods RTF file="&path\Descriptive whodasno0indis_sc0_&sysdate..rtf";
  title "Table 1. Distribution of WHODAS";
  proc report nowd data=_descri_whono0 headline headskip split="*";
run;
ODS RTF Close;
title "";
options orientation=portrait;
\%mend rtfdes;
\%rtfdes;
```

## Correlations among observed endogenous variables

```
\%macro corrs(weight,indata);

\%let vars= newfd23d_final scog0 ssoci0 smove0 scare0 outrol30 sdiscr0 sfambu0 sstig0;

\%do i=1 \%to 9;
\%let var1=\%scan(&vars,&i);
   \%do j=&i+1 \%to 9;
   \%let var2=\%scan(&vars,&j);
   proc corr data=&indata;
    weight &weight;
    var &var1 &var2;
   run;
  \%end;
\%end;

\%mend corrs;

\%corrs(finalp2wt,dtset_final);
```

## Prevalences of conditions and mean of WHODAS and VAS for individuals with the condition

```
\%macro table1_prevs(weight, indata);

/*** Dimensions and VAS ***/
\%let dim=  scog0 ssoci0 smove0 scare0 outrol30 sdiscr0 sfambu0 sstig0 newfd23d_final ;

/*** Conditions ***/
\%let dis= alcohol_abuse bipolar_dxn depression_mdd drug_abuse d_gadh12 panic_dx
```

```
d_pts12 d_so12 d_sp12
arthritis cancer cardiovascular musculoskeletor diabetes  digestive head_migrane
d_ins12_new neurological respiratory;

/*** Names of dimensions and disorders ***/
\%let ndim=scog ssoci smove scare role discri fambu stig vas;
\%let ndis= alc bip dep drug gad pan pts so sp arth cancer cv chr diab dig head ins neuro respi;

\%do k=1 \%to 19;

/*** cc: Condition ***/
 \%let cc=\%scan(&dis,&k);
/*** ccn: Name of the condition ***/
 \%let ccn=\%scan(&ndis,&k);

/*** Prevalence ***/
    proc crosstab data = &indata design = wr;
     nest str secu /missunit ;
     weight &weight;
     class &cc;
     tables &cc;
     output nsum rowper serow / filename = prev_&ccn replace;
    run;

    data prev_&ccn.;
     set prev_&ccn.;
     cc=put(&k,ccf.);
     perc_se=put(rowper,5.2)||" ("||put(serow,5.1)||")";
     if _n_=3;
    run;

    \%do i=1 \%to 9;
    \%let v1=\%scan(&dim,&i);
    \%let v1n=\%scan(&ndim,&i);

/*****************************************
**** Mean of WHODAS/VAS when condition=1 ****
*****************************************/
    proc descript data = &indata design = wr;
     nest str secu /missunit ;
     weight &weight;
     var  &v1;
     subpopn &cc=1;
     output nsum mean semean/ filename=mean_&v1n._&ccn filetype=SAS replace;
    run;

    data _mean_&v1n._&ccn.;
     set mean_&v1n._&ccn.;
     dim=put(&i,dimf.);
     cc=put(&k,ccf.);
     mean_se&v1n=put(mean,5.2)||" ("||put(semean,5.1)||")";
     if _n_=2;
```

```
    run;

    /*end i*/
    \%end;
/*end k*/
\%end;

\%mend table1_prevs;

\%table1_prevs(finalp2wt,dtset_final);

/*** Join the previous 19 tables ***/
\%macro joincc;

data _prevs_dis (keep= cc perc_se mean_sescog  mean_sessoci mean_sesmove mean_sescare
mean_serole mean_sediscri mean_sefambu mean_sestig mean_sevas);

 \%let ndim=scog ssoci smove scare role discri fambu stig vas;
 \%let ndis= alc bip dep drug gad pan pts so sp arth cancer cv chr diab dig head ins neuro respi;

 merge
  \%do k=1 \%to 19;
   \%let ccn=\%scan(&ndis,&k);
   prev_&ccn.
   \%do i=1 \%to 9;
    \%let v1n=\%scan(&ndim,&i);
    _mean_&v1n._&ccn.
   \%end;
  \%end;
;
by cc;
run;

\%mend joincc;
\%joincc;


/*** Export to RTF ***/
\%macro rtfprevs;
options orientation=landscape;
ods RTF file="Prevalences_&sysdate..rtf";
  title "Prevalences of CC and mean Disab/Perceived health among individuals with the condition";
  proc report nowd data=_prevs_dis headline headskip split="*";
run;
ODS RTF Close;
title "";
options orientation=portrait;
\%mend rtfprevs;
\%rtfprevs;
```

## D.2 MPLUS

Only the latent model is presented

```
!Read the dataset
    DATA: FILE IS "dtset_TFM_high_sc0.dat";

    VARIABLE:

!Specify the names of the variables from the dataset
     NAMES ARE

     weightp2 str secu vas sexm sexf age agesqr agesexm agesexf empw emps emph empr empo
     inclo  incloa inchia inchi marco mardi marnot
     cbe cfr cde cil cit cjp cnl cni cpt ces cus
     ncom dep bip pan sp so gad alab drab pts ins hemi arth chrpain cv respi diab dige
     neuro cancer
     whodas cog move care getalo role stigma discri famby scog smove scare sgetalo
     zvas zcog zmove zcare zgetalo zoutr zstig zdiscr zfambu
     concent under remem newtas stan movin wal was dres stayb
     conver dealunk maifrie makfrie conemo sstig sdiscr sfambu o1 o2 o3 o4
     scog0 smove0 scare0 ssoci0 rol30 sstig0 sdiscr0 sfambu0
     concentr unders rememb newtask stand movins walk wash dress stayby
     convers dealunkn maifrien makfrien conemot;

!Specify only the variables to be used
     USEVARIABLES ARE vas sexm age
     emps emph empr empo marco mardi
     cbe cfr cde cil cit cjp cnl cni cpt ces
     dep bip pan sp so gad alab drab pts ins hemi arth chrpain
     cv respi diab dige neuro cancer
     sdiscr0 sfambu0 sstig0 rol30
     concentr unders rememb newtask stand movins walk wash dress stayby
     convers dealunkn maifrien makfrien conemot;

!Indicators declared as ordinal
     CATEGORICAL ARE concentr unders rememb newtask stand movins walk
     wash dress stayby convers dealunkn maifrien makfrien conemot;

!Complex design variables
    WEIGHT IS weightp2;
    STRATIFICATION=str;
    CLUSTER=secu;

!Missing indicator
    MISSING ARE .;

!Define type of analysis
    ANALYSIS: TYPE=COMPLEX;
              PROCESSORS=8;
    MODEL:
```

```
!Declare latent factors by means of indicators. Parameters are fixed from CFA
!Cognition
    fcog by concentr@1 unders@0.987 rememb@1.006 newtask@0.988;
!Self-care
    fcare by wash@1 dress@0.994 stayby@0.971;
!Mobility
    fmove by stand@1 movins@0.991 walk@1.001;
!Getting along
    fgetalo by convers@1 dealunkn@1.003 maifrien@0.988 makfrien@0.993 conemot@0.964;

!Set threshold values of indicators from CFA
  [CONCENTR\$1@1.513];
  [CONCENTR\$2@1.764];
  [CONCENTR\$3@2.268];
  [CONCENTR\$4@3.151];
  [UNDERS\$1@1.75];
  [UNDERS\$2@2.083];
  [UNDERS\$3@2.608];
  [UNDERS\$4@3.464];
  [REMEMB\$1@1.569];
  [REMEMB\$2@1.856];
  [REMEMB\$3@2.294];
  [REMEMB\$4@3.256];
  [NEWTASK\$1@1.707];
  [NEWTASK\$2@1.94];
  [NEWTASK\$3@2.247];
  [NEWTASK\$4@2.863];
  [STAND\$1@1.106];
  [STAND\$2@1.239];
  [STAND\$3@1.563];
  [STAND\$4@2.189];
  [MOVINS\$1@1.209];
  [MOVINS\$2@1.464];
  [MOVINS\$3@1.957];
  [MOVINS\$4@2.703];
  [WALK\$1@1.118];
  [WALK\$2@1.201];
  [WALK\$3@1.363];
  [WALK\$4@1.872];
  [WASH\$1@1.78];
  [WASH\$2@1.871];
  [WASH\$3@2.102];
  [WASH\$4@2.592];
  [DRESS\$1@1.789];
  [DRESS\$2@1.924];
  [DRESS\$3@2.205];
  [DRESS\$4@2.698];
  [STAYBY\$1@1.964];
  [STAYBY\$2@2.069];
  [STAYBY\$3@2.216];
  [STAYBY\$4@2.578];
```

```
  [CONVERS\$101.836];
  [CONVERS\$202.078];
  [CONVERS\$302.448];
  [CONVERS\$403.307];
  [DEALUNKN\$101.858];
  [DEALUNKN\$202.041];
  [DEALUNKN\$302.359];
  [DEALUNKN\$402.954];
  [MAIFRIEN\$101.933];
  [MAIFRIEN\$202.168];
  [MAIFRIEN\$302.501];
  [MAIFRIEN\$403.171];
  [MAKFRIEN\$101.876];
  [MAKFRIEN\$201.98];
  [MAKFRIEN\$302.159];
  [MAKFRIEN\$402.688];
  [CONEMOT\$101.868];
  [CONEMOT\$202.109];
  [CONEMOT\$302.472];
  [CONEMOT\$403.326];


!Set correlation among mediators
!Cognition with Self-care, Mobility, Getting along, Stigma, Discrimination, Family burden,
!Role funcitoning
  fcog with fcare fmove fgetalo sstig0 sdiscr0 sfambu0 rol30;
  fcare with fmove fgetalo sstig0 sdiscr0 sfambu0 rol30;
  fmove with fgetalo sstig0 sdiscr0 sfambu0 rol30;
  fgetalo with sstig0 sdiscr0  sfambu0 rol30;

  sstig0 with sdiscr0 sfambu0 rol30;
  sdiscr0 with sfambu0 rol30;
  sfambu0 with rol30;

!SEM
!Outcome (vas) on sociodemographics (sex, age, employment status, marital status, country)
    vas on
    age cbe cfr cde cil cit cjp cnl cni cpt emps emph empr empo marco mardi ces sexm;

!Outcome on chronic conditions (Depression, Bipolar, Panic, Specific fobia, Social fobia,
!Generalized anxiety, Alcohol abuse)
    vas on
    dep bip pan sp so gad alab (dde dbi dpa dsp dso dga dal);
    vas on
!(Drug abuse, Posttraumatic stress, Insomnia, Headaches/Migraines, Arthritis, Chronic pain)
    drab pts ins hemi arth chrpain (ddr dpt din dhe dar dmu);
    vas on
!(Cardiovascular, Respiratory, Diabetes, Digestive, Neurological, Cancer)
    cv respi diab dige neuro cancer (dcv dre ddia ddig dne dca);

!Mediators on chronic conditions
    fcare on
```

```
dep bip pan sp so (icade icabi icapa icasp icaso);
fcare on
gad alab drab pts ins (icaga icaal icadr icapt icain);
fcare on
hemi arth chrpain cv respi (icahe icaar icamu icacv icare);
fcare on
diab dige neuro cancer (icadia icadig icane icaca);

fcog on
dep bip pan sp so(icode icobi icopa icosp icoso);
fcog on
gad alab drab pts ins (icoga icoal icodr icopt icoin);
fcog on
hemi arth chrpain cv respi (icohe icoar icomu icocv icore);
fcog on
diab dige neuro cancer (icodia icodig icone icoca);

sdiscr0 on
dep bip pan sp so (idde idbi idpa idsp idso);
sdiscr0 on
gad alab drab pts ins (idga idal iddr idpt idin);
sdiscr0 on
hemi arth chrpain cv respi (idhe idar idmu idcv idre);
sdiscr0 on
diab dige neuro cancer (iddia iddig idne idca);

sfambu0 on
dep bip pan sp so (ifde ifbi ifpa ifsp ifso);
sfambu0 on
gad alab drab pts ins (ifga ifal ifdr ifpt ifin);
sfambu0 on
hemi arth chrpain cv respi (ifhe ifar ifmu ifcv ifre);
sfambu0 on
diab dige neuro cancer (ifdia ifdig ifne ifca);

fmove on
dep bip pan sp so (imde imbi impa imsp imso);
fmove on
gad alab drab pts ins (imga imal imdr impt imin);
fmove on
hemi arth chrpain cv respi (imhe imar immu imcv imre);
fmove on
diab dige neuro cancer (imdia imdig imne imca);

rol30 on
dep bip pan sp so (iode iobi iopa iosp ioso);
rol30 on
gad alab drab pts ins (ioga ioal iodr iopt ioin);
rol30 on
hemi arth chrpain cv respi (iohe ioar iomu iocv iore);
rol30 on
diab dige neuro cancer (iodia iodig ione ioca);
```

```
    fgetalo on
    dep bip pan sp so (isode isobi isopa isosp isoso);
    fgetalo on
    gad alab drab pts ins (isoga isoal isodr isopt isoin);
    fgetalo on
    hemi arth chrpain cv respi (isohe isoar isomu isocv isore);
    fgetalo on
    diab dige neuro cancer (isodia isodig isone isoca);

    sstig0 on
    dep bip pan sp so (istde istbi istpa istsp istso);
    sstig0 on
    gad alab drab pts ins (istga istal istdr istpt istin);
    sstig0 on
    hemi arth chrpain cv respi (isthe istar istmu istcv istre);
    sstig0 on
    diab dige neuro cancer (istdia istdig istne istca);

!Outcome on mediators
    vas on fcare (ivca);
    vas on fcog (ivco);
    vas on sdiscr0 (ivd);
    vas on sfambu0 (ivf);
    vas on fmove (ivm);
    vas on rol30 (ivo);
    vas on fgetalo (ivso);
    vas on sstig0 (ivst);

!Declare the 152 indirect effects: Outcome on mediators * mediators on chronic conditions
   MODEL INDIRECT:

!Self-care
  vas IND fcare dep;
  vas IND fcare bip;
  vas IND fcare pan;
  vas IND fcare sp;
  vas IND fcare so;
  vas IND fcare gad;
  vas IND fcare alab;
  vas IND fcare drab;
  vas IND fcare pts;
  vas IND fcare ins;
  vas IND fcare hemi;
  vas IND fcare arth;
  vas IND fcare chrpain;
  vas IND fcare cv;
  vas IND fcare respi;
  vas IND fcare diab;
  vas IND fcare dige;
  vas IND fcare neuro;
  vas IND fcare cancer;
```

```
!Cognition
  vas IND fcog dep;
  vas IND fcog bip;
  vas IND fcog pan;
  vas IND fcog sp;
  vas IND fcog so;
  vas IND fcog gad;
  vas IND fcog alab;
  vas IND fcog drab;
  vas IND fcog pts;
  vas IND fcog ins;
  vas IND fcog hemi;
  vas IND fcog arth;
  vas IND fcog chrpain;
  vas IND fcog cv;
  vas IND fcog respi;
  vas IND fcog diab;
  vas IND fcog dige;
  vas IND fcog neuro;
  vas IND fcog cancer;

!Discrimination
  vas IND sdiscr0 dep;
  vas IND sdiscr0 bip;
  vas IND sdiscr0 pan;
  vas IND sdiscr0 sp;
  vas IND sdiscr0 so;
  vas IND sdiscr0 gad;
  vas IND sdiscr0 alab;
  vas IND sdiscr0 drab;
  vas IND sdiscr0 pts;
  vas IND sdiscr0 ins;
  vas IND sdiscr0 hemi;
  vas IND sdiscr0 arth;
  vas IND sdiscr0 chrpain;
  vas IND sdiscr0 cv;
  vas IND sdiscr0 respi;
  vas IND sdiscr0 diab;
  vas IND sdiscr0 dige;
  vas IND sdiscr0 neuro;
  vas IND sdiscr0 cancer;

!Family burden
  vas IND sfambu0 dep;
  vas IND sfambu0 bip;
  vas IND sfambu0 pan;
  vas IND sfambu0 sp;
  vas IND sfambu0 so;
  vas IND sfambu0 gad;
  vas IND sfambu0 alab;
  vas IND sfambu0 drab;
```

```
  vas IND sfambu0 pts;
  vas IND sfambu0 ins;
  vas IND sfambu0 hemi;
  vas IND sfambu0 arth;
  vas IND sfambu0 chrpain;
  vas IND sfambu0 cv;
  vas IND sfambu0 respi;
  vas IND sfambu0 diab;
  vas IND sfambu0 dige;
  vas IND sfambu0 neuro;
  vas IND sfambu0 cancer;

!Mobility
  vas IND fmove dep;
  vas IND fmove bip;
  vas IND fmove pan;
  vas IND fmove sp;
  vas IND fmove so;
  vas IND fmove gad;
  vas IND fmove alab;
  vas IND fmove drab;
  vas IND fmove pts;
  vas IND fmove ins;
  vas IND fmove hemi;
  vas IND fmove arth;
  vas IND fmove chrpain;
  vas IND fmove cv;
  vas IND fmove respi;
  vas IND fmove diab;
  vas IND fmove dige;
  vas IND fmove neuro;
  vas IND fmove cancer;

!Role functioning
  vas IND rol30 dep;
  vas IND rol30 bip;
  vas IND rol30 pan;
  vas IND rol30 sp;
  vas IND rol30 so;
  vas IND rol30 gad;
  vas IND rol30 alab;
  vas IND rol30 drab;
  vas IND rol30 pts;
  vas IND rol30 ins;
  vas IND rol30 hemi;
  vas IND rol30 arth;
  vas IND rol30 chrpain;
  vas IND rol30 cv;
  vas IND rol30 respi;
  vas IND rol30 diab;
  vas IND rol30 dige;
  vas IND rol30 neuro;
```

```
  vas IND rol30 cancer;

!Getting along
  vas IND fgetalo dep;
  vas IND fgetalo bip;
  vas IND fgetalo pan;
  vas IND fgetalo sp;
  vas IND fgetalo so;
  vas IND fgetalo gad;
  vas IND fgetalo alab;
  vas IND fgetalo drab;
  vas IND fgetalo pts;
  vas IND fgetalo ins;
  vas IND fgetalo hemi;
  vas IND fgetalo arth;
  vas IND fgetalo chrpain;
  vas IND fgetalo cv;
  vas IND fgetalo respi;
  vas IND fgetalo diab;
  vas IND fgetalo dige;
  vas IND fgetalo neuro;
  vas IND fgetalo cancer;

!Stigma
  vas IND sstig0 dep;
  vas IND sstig0 bip;
  vas IND sstig0 pan;
  vas IND sstig0 sp;
  vas IND sstig0 so;
  vas IND sstig0 gad;
  vas IND sstig0 alab;
  vas IND sstig0 drab;
  vas IND sstig0 pts;
  vas IND sstig0 ins;
  vas IND sstig0 hemi;
  vas IND sstig0 arth;
  vas IND sstig0 chrpain;
  vas IND sstig0 cv;
  vas IND sstig0 respi;
  vas IND sstig0 diab;
  vas IND sstig0 dige;
  vas IND sstig0 neuro;
  vas IND sstig0 cancer;


!Declare new variables
  MODEL CONSTRAINT:

!Indirect effects
  NEW (ivcade ivcabi ivcapa ivcasp ivcaso ivcaga ivcaal ivcadr ivcapt ivcain);
  NEW (ivcahe ivcaar ivcamu ivcacv ivcare ivcadia ivcadig ivcane ivcaca);
```

```
  NEW (ivcode ivcobi ivcopa ivcosp ivcoso ivcoga ivcoal ivcodr ivcopt ivcoin);
  NEW (ivcohe ivcoar ivcomu ivcocv ivcore ivcodia ivcodig ivcone ivcoca);

  NEW (ivdde ivdbi ivdpa ivdsp ivdso ivdga ivdal ivddr ivdpt ivdin);
  NEW (ivdhe ivdar ivdmu ivdcv ivdre ivddia ivddig ivdne ivdca);

  NEW (ivfde ivfbi ivfpa ivfsp ivfso ivfga ivfal ivfdr ivfpt ivfin);
  NEW (ivfhe ivfar ivfmu ivfcv ivfre ivfdia ivfdig ivfne ivfca);

  NEW (ivmde ivmbi ivmpa ivmsp ivmso ivmga ivmal ivmdr ivmpt ivmin);
  NEW (ivmhe ivmar ivmmu ivmcv ivmre ivmdia ivmdig ivmne ivmca);

  NEW (ivode ivobi ivopa ivosp ivoso ivoga ivoal ivodr ivopt ivoin);
  NEW (ivohe ivoar ivomu ivocv ivore ivodia ivodig ivone ivoca);

  NEW (ivsode ivsobi ivsopa ivsosp ivsoso ivsoga ivsoal ivsodr ivsopt ivsoin);
  NEW (ivsohe ivsoar ivsomu ivsocv ivsore ivsodia ivsodig ivsone ivsoca);

  NEW (ivstde ivstbi ivstpa ivstsp ivstso ivstga ivstal ivstdr ivstpt ivstin);
  NEW (ivsthe ivstar ivstmu ivstcv ivstre ivstdia ivstdig ivstne ivstca);

!Overall Indirect Contributions (OIC) of conditions
  NEW (perde perbi perpa persp perso perga peral perdr perpt perin perhe);
  NEW (perar permu percv perre perdia perdig perne perca);

!Overall Indirect effects of conditions
  NEW (sumide sumibi sumipa sumisp sumiso sumiga sumial sumidr sumipt sumiin);
  NEW (sumihe sumiar sumimu sumicv sumire sumidia sumidig sumine sumica);

!Total effects of conditions
  NEW (totde totbi totpa totsp totso totga total totdr totpt totin tothe);
  NEW (totar totmu totcv totre totdia totdig totne totca);

!Compute the 152 indirect effects of conditions (single paths)
!Outcome on mediators * mediators on conditions
!Self care
  ivcade=ivca*icade;
  ivcabi=ivca*icabi;
  ivcapa=ivca*icapa;
  ivcasp=ivca*icasp;
  ivcaso=ivca*icaso;
  ivcaga=ivca*icaga;
  ivcaal=ivca*icaal;
  ivcadr=ivca*icadr;
  ivcapt=ivca*icapt;
  ivcain=ivca*icain;
  ivcahe=ivca*icahe;
  ivcaar=ivca*icaar;
  ivcamu=ivca*icamu;
  ivcacv=ivca*icacv;
  ivcare=ivca*icare;
  ivcadia=ivca*icadia;
```

```
  ivcadig=ivca*icadig;
  ivcane=ivca*icane;
  ivcaca=ivca*icaca;

!Cognition
  ivcode=ivco*icode;
  ivcobi=ivco*icobi;
  ivcopa=ivco*icopa;
  ivcosp=ivco*icosp;
  ivcoso=ivco*icoso;
  ivcoga=ivco*icoga;
  ivcoal=ivco*icoal;
  ivcodr=ivco*icodr;
  ivcopt=ivco*icopt;
  ivcoin=ivco*icoin;
  ivcohe=ivco*icohe;
  ivcoar=ivco*icoar;
  ivcomu=ivco*icomu;
  ivcocv=ivco*icocv;
  ivcore=ivco*icore;
  ivcodia=ivco*icodia;
  ivcodig=ivco*icodig;
  ivcone=ivco*icone;
  ivcoca=ivco*icoca;

!Discrimination
  ivdde=ivd*idde;
  ivdbi=ivd*idbi;
  ivdpa=ivd*idpa;
  ivdsp=ivd*idsp;
  ivdso=ivd*idso;
  ivdga=ivd*idga;
  ivdal=ivd*idal;
  ivddr=ivd*iddr;
  ivdpt=ivd*idpt;
  ivdin=ivd*idin;
  ivdhe=ivd*idhe;
  ivdar=ivd*idar;
  ivdmu=ivd*idmu;
  ivdcv=ivd*idcv;
  ivdre=ivd*idre;
  ivddia=ivd*iddia;
  ivddig=ivd*iddig;
  ivdne=ivd*idne;
  ivdca=ivd*idca;

!Family burden
  ivfde=ivf*ifde;
  ivfbi=ivf*ifbi;
  ivfpa=ivf*ifpa;
  ivfsp=ivf*ifsp;
  ivfso=ivf*ifso;
```

```
  ivfga=ivf*ifga;
  ivfal=ivf*ifal;
  ivfdr=ivf*ifdr;
  ivfpt=ivf*ifpt;
  ivfin=ivf*ifin;
  ivfhe=ivf*ifhe;
  ivfar=ivf*ifar;
  ivfmu=ivf*ifmu;
  ivfcv=ivf*ifcv;
  ivfre=ivf*ifre;
  ivfdia=ivf*ifdia;
  ivfdig=ivf*ifdig;
  ivfne=ivf*ifne;
  ivfca=ivf*ifca;

!Mobility
  ivmde=ivm*imde;
  ivmbi=ivm*imbi;
  ivmpa=ivm*impa;
  ivmsp=ivm*imsp;
  ivmso=ivm*imso;
  ivmga=ivm*imga;
  ivmal=ivm*imal;
  ivmdr=ivm*imdr;
  ivmpt=ivm*impt;
  ivmin=ivm*imin;
  ivmhe=ivm*imhe;
  ivmar=ivm*imar;
  ivmmu=ivm*immu;
  ivmcv=ivm*imcv;
  ivmre=ivm*imre;
  ivmdia=ivm*imdia;
  ivmdig=ivm*imdig;
  ivmne=ivm*imne;
  ivmca=ivm*imca;

!Role functioning
  ivode=ivo*iode;
  ivobi=ivo*iobi;
  ivopa=ivo*iopa;
  ivosp=ivo*iosp;
  ivoso=ivo*ioso;
  ivoga=ivo*ioga;
  ivoal=ivo*ioal;
  ivodr=ivo*iodr;
  ivopt=ivo*iopt;
  ivoin=ivo*ioin;
  ivohe=ivo*iohe;
  ivoar=ivo*ioar;
  ivomu=ivo*iomu;
  ivocv=ivo*iocv;
  ivore=ivo*iore;
```

```
  ivodia=ivo*iodia;
  ivodig=ivo*iodig;
  ivone=ivo*ione;
  ivoca=ivo*ioca;

!Getting along
  ivsode=ivso*isode;
  ivsobi=ivso*isobi;
  ivsopa=ivso*isopa;
  ivsosp=ivso*isosp;
  ivsoso=ivso*isoso;
  ivsoga=ivso*isoga;
  ivsoal=ivso*isoal;
  ivsodr=ivso*isodr;
  ivsopt=ivso*isopt;
  ivsoin=ivso*isoin;
  ivsohe=ivso*isohe;
  ivsoar=ivso*isoar;
  ivsomu=ivso*isomu;
  ivsocv=ivso*isocv;
  ivsore=ivso*isore;
  ivsodia=ivso*isodia;
  ivsodig=ivso*isodig;
  ivsone=ivso*isone;
  ivsoca=ivso*isoca;

!Stigma
  ivstde=ivst*istde;
  ivstbi=ivst*istbi;
  ivstpa=ivst*istpa;
  ivstsp=ivst*istsp;
  ivstso=ivst*istso;
  ivstga=ivst*istga;
  ivstal=ivst*istal;
  ivstdr=ivst*istdr;
  ivstpt=ivst*istpt;
  ivstin=ivst*istin;
  ivsthe=ivst*isthe;
  ivstar=ivst*istar;
  ivstmu=ivst*istmu;
  ivstcv=ivst*istcv;
  ivstre=ivst*istre;
  ivstdia=ivst*istdia;
  ivstdig=ivst*istdig;
  ivstne=ivst*istne;
  ivstca=ivst*istca;

!Compute overall indirect effects by condition
  sumide=ivcade+ivcode+ivdde+ivfde+ivmde+ivode+ivsode+ivstde;
  sumibi=ivcabi+ivcobi+ivdbi+ivfbi+ivmbi+ivobi+ivsobi+ivstbi;
  sumipa=ivcapa+ivcopa+ivdpa+ivfpa+ivmpa+ivopa+ivsopa+ivstpa;
  sumisp=ivcasp+ivcosp+ivdsp+ivfsp+ivmsp+ivosp+ivsosp+ivstsp;
```

```
  sumiso=ivcaso+ivcoso+ivdso+ivfso+ivmso+ivoso+ivsoso+ivstso;
  sumiga=ivcaga+ivcoga+ivdga+ivfga+ivmga+ivoga+ivsoga+ivstga;
  sumial=ivcaal+ivcoal+ivdal+ivfal+ivmal+ivoal+ivsoal+ivstal;
  sumidr=ivcadr+ivcodr+ivddr+ivfdr+ivmdr+ivodr+ivsodr+ivstdr;
  sumipt=ivcapt+ivcopt+ivdpt+ivfpt+ivmpt+ivopt+ivsopt+ivstpt;
  sumiin=ivcain+ivcoin+ivdin+ivfin+ivmin+ivoin+ivsoin+ivstin;
  sumihe=ivcahe+ivcohe+ivdhe+ivfhe+ivmhe+ivohe+ivsohe+ivsthe;
  sumiar=ivcaar+ivcoar+ivdar+ivfar+ivmar+ivoar+ivsoar+ivstar;
  sumimu=ivcamu+ivcomu+ivdmu+ivfmu+ivmmu+ivomu+ivsomu+ivstmu;
  sumicv=ivcacv+ivcocv+ivdcv+ivfcv+ivmcv+ivocv+ivsocv+ivstcv;
  sumire=ivcare+ivcore+ivdre+ivfre+ivmre+ivore+ivsore+ivstre;
  sumidia=ivcadia+ivcodia+ivddia+ivfdia+ivmdia+ivodia+ivsodia+ivstdia;
  sumidig=ivcadig+ivcodig+ivddig+ivfdig+ivmdig+ivodig+ivsodig+ivstdig;
  sumine=ivcane+ivcone+ivdne+ivfne+ivmne+ivone+ivsone+ivstne;
  sumica=ivcaca+ivcoca+ivdca+ivfca+ivmca+ivoca+ivsoca+ivstca;

!Compute OICs
  perde=100*sumide/(dde+sumide+0.00000000000000001);
  perbi=100*sumibi/(dbi+sumibi+0.00000000000000001);
  perpa=100*sumipa/(dpa+sumipa+0.00000000000000001);
  persp=100*sumisp/(dsp+sumisp+0.00000000000000001);
  perso=100*sumiso/(dso+sumiso+0.00000000000000001);
  perga=100*sumiga/(dga+sumiga+0.00000000000000001);
  peral=100*sumial/(dal+sumial+0.00000000000000001);
  perdr=100*sumidr/(ddr+sumidr+0.00000000000000001);
  perpt=100*sumipt/(dpt+sumipt+0.00000000000000001);
  perin=100*sumiin/(din+sumiin+0.00000000000000001);
  perhe=100*sumihe/(dhe+sumihe+0.00000000000000001);
  perar=100*sumiar/(dar+sumiar+0.00000000000000001);
  permu=100*sumimu/(dmu+sumimu+0.00000000000000001);
  percv=100*sumicv/(dcv+sumicv+0.00000000000000001);
  perre=100*sumire/(dre+sumire+0.00000000000000001);
  perdia=100*sumidia/(ddia+sumidia+0.00000000000000001);
  perdig=100*sumidig/(ddig+sumidig+0.00000000000000001);
  perne=100*sumine/(dne+sumine+0.00000000000000001);
  perca=100*sumica/(dca+sumica+0.00000000000000001);

!Compute total effects
  totde=dde+sumide;
  totbi=dbi+sumibi;
  totpa=dpa+sumipa;
  totsp=dsp+sumisp;
  totso=dso+sumiso;
  totga=dga+sumiga;
  total=dal+sumial;
  totdr=ddr+sumidr;
  totpt=dpt+sumipt;
  totin=din+sumiin;
  tothe=dhe+sumihe;
  totar=dar+sumiar;
  totmu=dmu+sumimu;
  totcv=dcv+sumicv;
```

```
  totre=dre+sumire;
  totdia=ddia+sumidia;
  totdig=ddig+sumidig;
  totne=dne+sumine;
  totca=dca+sumica;

!Output options
  OUTPUT:  SAMPSTAT;
           STANDARDIZED (STDYX);
           STANDARDIZED (STDY);
           STANDARDIZED (STD);
           tech4;
```

## D.3 R

```
#Libraries to produce the barplot of Decompositon of effects
library(ggplot2)
library(reshape)
library(scales)

#Read data file from Excel defined with appropriate format for ggplot
db<-read.csv2("cdmm_both_sc0.csv",dec=".")

#Disorders are ordered by Total effect in CDMM-L
#s1, ... , s19 are auxiliary variables to separate each couple of bars
orderlist =c("Neurological_O","Neurological_L","s1","Depression_O","Depression_L","s2",
"Bipolar_O","Bipolar_L","s3","Chronic pain_O","Chronic pain_L","s4","Digestive_O","Digestive_L",
"s5","Panic disorder_O","Panic disorder_L","s6","Insomnia_O","Insomnia_L","s7",
"Diabetes_O","Diabetes_L","s8","Cardiovascular_O","Cardiovascular_L","s9",
"Generalized anxiety_O","Generalized anxiety_L","s10","Arthritis_O","Arthritis_L","s11",
"Posttraumatic stress_O","Posttraumatic stress_L","s12","Drug abuse_O","Drug abuse_L","s13",
"Headaches/Migraines_O","Headaches/Migraines_L","s14",
"Specific phobia_O","Specific phobia_L","s15","Social phobia_O","Social phobia_L","s16",
"Alcohol abuse_O","Alcohol abuse_L","s17","Cancer_O","Cancer_L","s18",
"Respiratory_O","Respiratory_L","s19")

#Order conditions by Total effect in CDMM-L
db\$Condition<-factor(db\$Condition, levels=orderlist,ordered=TRUE)

#Declare two datasets, one for positive and one for negative values
dat1 <- subset(db,Value >= 0)
dat2 <- subset(db,Value < 0)

#Set Direct effect of Drug abuse (positive) to 0 represent the correct bar
db[67,2]<-0

#Create a dataset with ddply to distinguish Direct and Indirect effects in the graph
zz<-ddply(db,.(Type),summarise,Condition=Condition,Value=Value,SE=SE,Total=Total)

#ggplot function
ggplot(data = db, aes(x = Condition, y = Value, fill = Type),stat = "identity") +
geom_bar(position="stack")+
geom_bar(data = dat1, aes(x = Condition, y = Value, fill = Type),
stat = "identity",position="stack") +
geom_errorbar(data=zz,aes(ymin=Total-SE,ymax=Total+SE,colour=Type))+
scale_fill_brewer(palette="YlGn")+
theme(axis.text.x = element_text(angle = 90, hjust = 1))+
ylab("Total effect")+
xlab("")+
opts(legend.position="top")
```

# Bibliography

D Borsboom, GJ Mellenbergh, and J van Heerden. The theoretical status of latent variables. *Psychological Review*, 110:203–219, 2003.

KA Bollen. *Structural equations with latent variables*. John Wiley & Sons, Chapel Hill, North Carolina, 1989.

JE Rohrer, R Young, V Sicola, and M Houston. Overall self-rated health: a new quality indicator for primary care. *Journal of Evaluation in Clinical Practice*, 13:150–153, 2007.

AV Perruccio, JD Power, and EM Badley. The relative impact of 13 chronic conditions across three different outcomes. *Journal of Epidemiology and Community Health*, 61: 1056–1061, 2007.

J Alonso. La medida de la calidad de vida relacionada con la salud en la investigación y la práctica. *Gaceta Sanitaria*, 14:163–167, 2000.

SI Saarni, T Harkanen, H Sintonen, J Suvisaari, and S Koskinen et al. The impact of 29 chronic conditions on health-related quality of life: A general population survey in Finland using 15D and EQ-5D. *Quality of Life Research*, 15:1403–1414, 2006.

SE Schultz and JA Kopec. Impact of chronic conditions. *Health reports*, 14(4):41–53, 2003.

J Alonso, M Ferrer, B Gandek, JE Ware, and NK Aaronson et al. Health-related quality of life associated with chronic conditions in eight countries: results from the International Quality of Life Assessment (IQOLA) Project. *Quality of Life Research*, 13(2):283–298, 2004.

A van Dijk, PA McGrath, W Pickett, and EG Van Den Kerkhof. Pain and self-reported health in Canadian children. *Pain Research & Management*, 13(5):407–411, 2008.

J Alonso, G Vilagut, S Chatterji, S Heeringa, and M Schoenbaum et al. Including information about co-morbidity in estimates of disease burden: results from the World Health Organization World Mental Health Surveys. *Psychological Medicine*, 41(4):873–886, 2011.

J Ormel, GI Kempen, DJ Deeg, EI Brilman, E van Sonderen, and J Relyveld. Functioning, well-being, and health perception in late middle-aged and older people: comparing the effects of depressive symptoms and chronic medical conditions. *Journal of the American Geriatrics Society*, 46(1):39–48, 1998.

IB Wilson and PD Cleary. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *The Journal of the American Medical Association*, 273):59–65, 1995.

RM Baron and DA Kenny. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51:1173–1182, 1986.

HY Lee, SN Jang, S Lee, SI Cho, and EO Park. The relationship between social participation and self-rated health by sex and age: a cross-sectional survey. *International Journal of Nursing Studies*, 45(7):1042–1054, 2008.

R Leinonen, E Heikkinen, and M Jylhä. Predictors of decline in selfassessments of health among older people–a 5-year longitudinal study. *Social Science & Medicine*, 52(9): 1329–1341, 2001.

J Ormel, M Petukhova M, S Chatterji, S Aguilar-Gaxiola S, and J Alonso et al. Disability and treatment of specific mental and physical disorders across the world. *The British Journal of Psychiatry*, 192:368–375, 2008.

World Mental Health Surveys initiative. http://www.hcp.med.harvard.edu/wmh. Accessed: 2013-12-19.

J Alonso, G Vilagut, ND Adroher, S Chatterji, and Y He et al. Disability Mediates the Impact of Common Conditions on Perceived Health. *Plos One*, 8(6):e65858, 2013.

APA. *Diagnostic and Statistical Manual of Mental Disorders. 4th edition, text revised*. American Psychiatric Association, Washington, DC, 2000.

H Baumeister, L Kriston, J Bengel J, and M Harter. High agreement of self-report and physician-diagnosed somatic conditions yields limited bias in examining mental-physical comorbidity. *Journal of Clinical Epidemiology*, 63(5):558–565, 2010.

World Health Organization Disability Assessment Schedule (WHODAS) 2.0. http://www.who.int/classifications/icf/whodasii/en/. Accessed: 2013-12-13.

A Paul-Dauphin, F Guillemin, JM Virion, and S Briançon. Bias and precision in visual analogue scales: a randomized controlled trial. *American Journal of Epidemiology*, 150(10):1117–1127, 1999.

DA Binder. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51:279–292, 1983.

J Fox and S Weisberg. *An R Companion to Applied Regression, Second Edition. Appendix: Structural Equation Modeling*. Sage Publications, Inc, Thousand Oaks, CA, 2011. URL http://socserv.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-SEMs.pdf.

JJ McArdle and RP McDonald. Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, 37: 234–251, 1984.

KA Bollen. Total, direct and indirect effects in structural equation models. *Sociological Methodology*, 17:37–69, 1987.

PM Bentler and EH Freeman. Tests for stability in linear structural equation systems. *Psychometrika*, 48:143–145, 1983.

DF Alwin and RM Hauser. The decomposition of effects in path analysis. *American Sociological Review*, 40:37–47, 1975.

U Olsson. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44:443–460, 1979.

MW Browne. Asymptotic distribution free methods in analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37:62–83, 1984.

B Muthén. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49:115–132, 1984.

DB Flora and PJ Curran. An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods*, 9(4): 466–491, 2004.

F Yang-Wallentin, KG Jöreskog, and H Luo. Confirmatory Factor Analysis of Ordinal Variables With Misspecified Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 17:392–423, 2010.

KG Jöreskog. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34:183–202, 1969.

LT Hu and PM Bentler. Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6:1–55, 1999.

CP Dancey and J Reidy. *Statistics without maths for psychology: using SPSS for windows)*. Prentice Hall, London, 2004.

P Dhrymes. *Introductory Econometrics*. Springer-Verlag, New York, 1978.

J Cohen. *Statistical Power Analysis for the Behavioral Sciences (2nd edition)*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

O Garin, JL Ayuso-Mateos, J Almansa, M Nieto, and S Chatterji et al. Validation of the World Health Organiation Disability Assessment Schedule (WHODAS-2) in patients with chronic diseases. *Health and Quality of Life Outcomes*, 8:51, 2010.

MA Buist-Bouwman, J Ormel, R de Graaf, P de Jonge, and E van Sonderen et al. Mediators of the association between depression and role functioning. *Acta Psychiatrica Scandinavica*, 118:451–458, 2008.

LE Knouse, I Zvorsky, and SA Safren. Depression in Adults with Attention-Deficit/Hyperactivity Disorder (ADHD): The Mediating Role of Cognitive-Behavioral Factors. *Cognitive Therapy and Research*, 37(6):1220–1232, 2013.

VK Noonan, JA Kopec, L Noreau, J Singer, LC Mâsse, and MF Dvorak. Comparing the reliability of five participation instruments in persons with spinal conditions. *Journal of Rehabilitation Medicine*, 42:735–743, 2010.