Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences

School for Information Technology

Master of Statistics and Data Science

Master's thesis

Health literacy as the missing link between socioeconomic vulnerability and poor health outcomes.

Ayleen Mufudza

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :
Prof. dr. Annelies AGTEN

SUPERVISOR :
Prof. dr. Annelies BOONEN

2024
2025

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Health literacy as the missing link between socioeconomic vulnerability and poor health outcomes.

Ayleen Mufudza

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :
Prof. dr. Annelies AGTEN
SUPERVISOR :
Prof. dr. Annelies BOONEN

Abstract

Introduction: Inflammatory rheumatic and musculoskeletal diseases (iRMDs), including rheumatoid arthritis, spondyloarthritis, and gout, are chronic conditions characterized by immune-driven joint inflammation requiring ongoing management. While advances in treatment have improved outcomes, disparities persist in health status and disease control. Socioeconomic status (SES) influences these disparities, but the pathways through which SES affects health outcomes remain unclear. Factors such as health literacy and mastery, which reflect patients' cognitive and psychological resources, may help explain these relationships.

Aims and Objectives : The primary aim of this project was to examine whether health literacy mediates the relationship between SES and three health outcomes number of comorbidities, patient global assessment of health, and professional severity score among patients with iRMDs. Secondary objectives explored whether mastery, either alone or together with health literacy, also mediates the SES and health outcomes relationship.

Methodology: Confirmatory factor analysis (CFA) was first conducted as a preliminary step to validate the measurement models of the latent variables, health literacy and SES. This step ensured that the observed indicators reliably represented the underlying constructs. With valid measurement models established, mediation analysis was then carried out within the structural equation modeling (SEM) framework to examine how health literacy and mastery mediate the relationship between SES and health outcomes.

Results: In all models, higher socioeconomic status (SES) was associated with higher health literacy and better patient-reported health scores, indicating improved health outcomes. Higher SES was also associated with higher health literacy and lower professional disease severity scores . For the number of comorbidities, the effect of SES was mainly direct, with little or no mediation. Additionally, the model including mastery alone showed a stronger mediation effect than models with health literacy alone or both mediators combined.

Discussion and Conclusion: Mastery demonstrated a stronger mediating effect than health literacy in the relationship between socioeconomic status (SES) and health outcomes. However, including both mediators simultaneously did not significantly enhance the model beyond using health literacy alone. For self-reported comorbidities, SES had a direct effect without evidence of mediation, unlike the other health outcomes, where mediation was observed.

Keywords: socioeconomic status (SES), health literacy, mastery, mediation analysis, confirmatory factor analysis(CFA), structural equation modeling (SEM), inflammatory rheumatic and musculoskeletal diseases (iRMDs)

1 Introduction

According to a consensus document endorsed by the European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR), rheumatic and musculoskeletal diseases (RMDs) are a diverse group of conditions that commonly affect the joints but can also involve muscles, bones, and other organs. There are more than 200 different RMDs, affecting both children and adults. Immune system dysfunction, chronic inflammation, infections, or gradual deterioration of musculoskeletal tissues usually cause these diseases. Many RMDs are long-term and progressive, often leading to pain, impaired function, and in severe cases, significant disability that impacts quality of life and life expectancy [1].

Within the 200 different rheumatic and musculoskeletal diseases (RMDs), there is a subset called inflammatory rheumatic and musculoskeletal diseases (iRMDs), characterized by immune system-driven inflammation that primarily affects the joints and surrounding tissues. Key examples of iRMDs include rheumatoid arthritis (RA), spondyloarthritis (SpA), and gout. RA is an autoimmune condition that causes progressive joint damage and inflammation. Gout is marked by sudden episodes of intense joint pain due to uric acid crystal buildup. Spondyloarthritis is an auto-infflammatory disease that primarily targets the spine and larger joints through inflammatory processes.[2] [3, 4]

Fortunately, advances in understanding these disease mechanisms have paved the way for targeted therapies that effectively control inflammation, prevent joint damage, and improve long-term outcomes. However, achieving optimal results depends not only on medical treatments but also on effective disease management. This requires regular monitoring of disease activity and timely adjustments to therapy. Central to this approach is the active involvement of patients in their care, as patient engagement is essential for improving treatment adherence and maintaining quality of life.

However, people with inflammatory arthritis from lower socioeconomic backgrounds consistently experience worse health outcomes[5, 6, 7], with individuals from lower SES backgrounds experiencing approximately 10 years lower life expectancy and higher rates of comorbidities, yet the mechanisms underlying this association are not fully understood. One possibility is that socioeconomic disadvantage restricts people's ability to navigate the healthcare system effectively, engaging with healthcare providers, accessing and utilizing health information. These challenges fall under the broader concept of health literacy and may contribute to poorer disease management and adverse health outcomes [8].

While theoretical models have outlined the relationships between SES, health literacy, and health outcomes [9], their application in iRMD populations remains largely untested. Furthermore, it remains unclear whether the role of health literacy is consistent across different types of outcomes, such as patient global assessment of health, health professional-disease severity assessment, or self-reported comorbidities. It is also unknown whether the relationship between health literacy and outcomes varies across different types of iRMDs.

That said, health literacy is a complex and multifaceted concept that can be difficult to measure or apply consistently. To illustrate this complexity, Paasche-Orlow and Wolf [10] developed a conceptual model with causal pathways linking health literacy to health outcomes. Their model highlights three key pathways influenced by health literacy: access to and utilization of healthcare, provider–patient interactions, and self-care (see Figure 1).

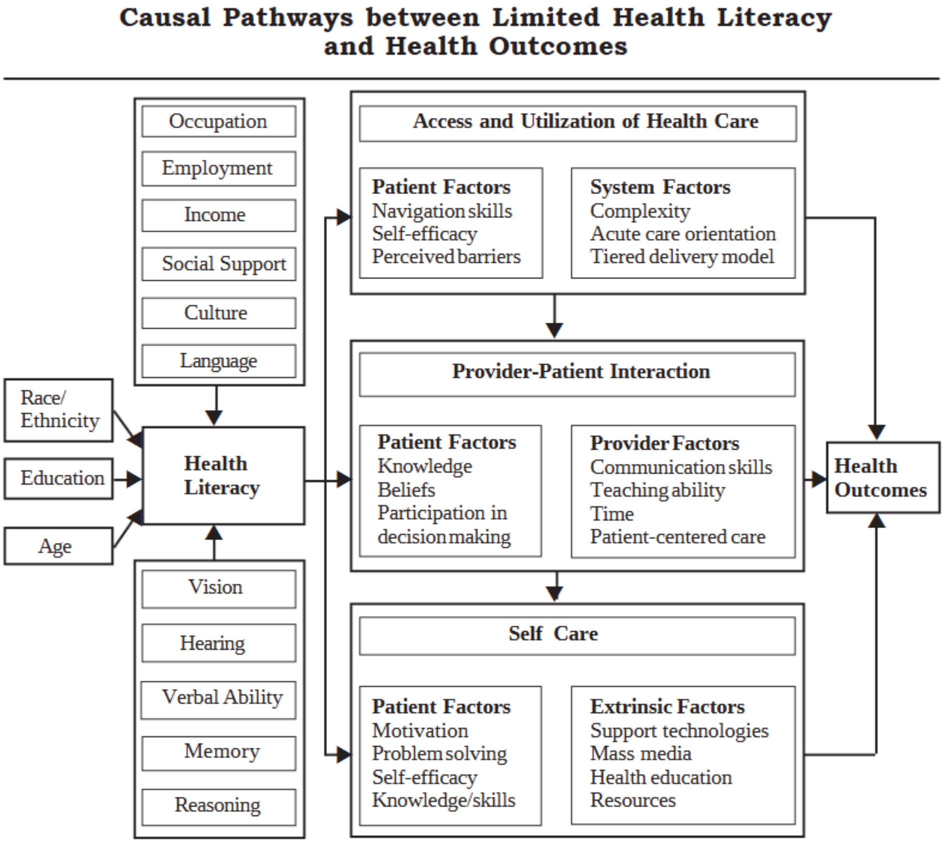


Figure 1: Paasche-Orlow and Wolf model illustrating the key pathways through which health literacy influences health outcomes.

The model emphasizes that successful self-care requires both cognitive comprehension of health information and the psychological confidence to apply it effectively. This focus on self-care resonates with Pearlin's concept of mastery [11], an individual's perceived control over their life circumstances, including health. While the Paasche-Orlow model centers on health literacy primarily within healthcare settings, mastery represents a psychological factor that influences health outcomes beyond healthcare environments. Notably, mastery has long been associated with health outcomes in iRMDs, predating the Paasche-Orlow model. However, it remains unclear whether health literacy serves as a stronger mediator in the relationship between socioeconomic status and health outcomes or if its influence is partly indirect through mastery. Understanding these pathways is crucial for designing interventions that address health disparities more effectively.

Therefore, to test these proposed pathways, this secondary analysis examines data from 895 individuals with iRMDs from a multicenter study conducted in the Netherlands. The analysis focuses on whether health literacy, mastery or both mediate the relationship between SES and health outcomes:

self-reported comorbidities, patient global assessment of health, and health professional-reported disease severity scores.

1.1 Research Questions

Primary research question:

1. To examine whether health literacy mediates the relationship between socioeconomic status and each of the three health outcomes (number of comorbidities, patient global assessment of health, and professional severity score) among inflammatory rheumatic and musculoskeletal diseases (iRMDs).

Secondary research questions:

1. To examine whether mastery alone mediates the relationship between socioeconomic status and the health outcomes among iRMDs patients.
2. To examine whether both health literacy and mastery jointly mediate the relationship between socioeconomic status and the health outcomes among iRMDs patients.

2 Methodology

2.1 Study design, setting and study population

This project utilized secondary data from a multicenter, cross-sectional study [7] conducted in the Netherlands involving 895 patients with inflammatory rheumatic and musculoskeletal diseases (iRMDs). The study was conducted across three hospitals in the Netherlands: Maastricht University Medical Centre, Maasstad Hospital in Rotterdam, and Medisch Spectrum Twente in Enschede. These sites were selected to capture a range of socioeconomic contexts, including urban and rural settings, economic variability, and migration patterns.

2.2 Data Collection and Procedures

The original study collected data between May 2018 and May 2019 to examine patterns of health literacy among patients with iRMDs. Data were gathered through patient questionnaires, healthcare provider questionnaires, and a structured health literacy assessment derived from patient responses. These questionnaires covered a broad range of information, including patient characteristics, self-reported health, and clinical assessments. The patient questionnaire yielded N = 895 responses. For the healthcare professionals' questionnaire, data collection varied across participating study centers due to resource constraints [12]. Two centers used a predefined sampling protocol where health professionals completed the questionnaire for every second patient on their center list, while other centers collected data on all or nearly all patients, depending on staffing capacity. This approach resulted in a total of N = 778 responses for the healthcare professionals' questionnaire.

2.3 Data Variables

The dataset used for this analysis was provided in `.dta` format and included data on 895 patients with 237 variables. For this project, a focused subset of variables relevant to the research questions was selected. These variables are summarized in Table 1. In the following subsections, the selected variables are described, detailing how they were measured and categorized, as well as their role in the measurement or structural models.

Table 1: Summary of the Variables Analysed

Variable	Type/Options
Socioeconomic Status	
Education	Categorical: Low, Medium, High
Age	Continuous (Years)
Gender	Categorical: Male, Female
Migration background	Categorical: Non Western, Western, Native
Employment	Categorical:Not employed; Work disabled; Retired; Employed
Living situation	Categorical: Living alone; Living with others
HLQ Scales and Items [14] (Mean Score for Each Domain, Continuous)	
Feeling understood (HPS)	Mean score for 4 items (4-point Likert scale)
Having sufficient information (HSI)	Mean score for 4 items (4-point Likert scale)
Actively managing health (AMH)	Mean score for 5 items (4-point Likert scale)
Social support for health (SS)	Mean score for 5 items (4-point Likert scale)
Appraisal of health information (CA)	Mean score for 5 items (4-point Likert scale)
Ability to engage with providers (AE)	Mean score for 5 items (5-point Likert scale)
Navigating healthcare system (NHS)	Mean score for 6 items (5-point Likert scale)
Ability to find good health information (FHI)	Mean score for 5 items (5-point Likert scale)
Understanding health information (UHI)	Mean score for 5 items (5-point Likert scale)
Mastery of Health (Higher scores indicate greater mastery)	
Level of mastery over the disease	Total score of 7 items (Ordinal: 7-28 score)
Health Outcomes	
Comorbidities	Count of 10 self-reported chronic conditions (see outcome variables section for full list 2.3.3)
Patient Global Assessment of Health	Ordinal (10-point Likert scale)
Physician score of impact on health	Ordinal (10-point Likert scale)

2.3.1 Explanatory variables

Socioeconomic status (SES) was conceptualized as a latent variable represented by six observed indicators: age, gender, education level, migration background, employment status, and living situation. Age was treated as a continuous variable in years. Gender was coded categorically as female (1), male (2). Migration background was classified into three groups: non-Western migrants (1), Western migrants (2), and natives (3). Education was categorized into three levels: low (1), medium (2), and high (3), based on the Dutch educational system [13]. The specific education categories are listed in Table 11 in the Supplementary Material.

Employment status originally included multiple categories (student, paid work, unemployed, unable to work, housewife or husband, and retired) but was recoded into four mutually exclusive groups: not employed (combining unemployed, students, and housewives or husbands) (1), work disabled (unable to work)(2), retired (3) and employed (paid work)(4). Living situation was originally categorized into five household composition groups but was regrouped into two categories: living alone (1) or living with others (2). Recoding for these two variables ensured that each participant belonged to one and only one category per variable, maintaining clarity and consistency for subsequent analyses.

2.3.2 Mediators

Health Literacy

Health literacy (HL) refers to an individual's ability to access, understand, evaluate, and use health information to make informed decisions. In this project, HL was conceptualized as a latent construct and measured using the Health Literacy Questionnaire (HLQ), developed by Osborne et al. [14]. The HLQ items were extracted from the patient-reported questionnaire and comprised 44 items grouped into nine conceptually distinct domains, each reflecting a unique aspect of health literacy following the Osborne framework. For each domain, a mean score was calculated across its corresponding items. These nine domain-level mean scores, treated as continuous variables, served as observed indicators of the latent construct of health literacy in the analysis. The specific domains are presented in Table 1 and the domain specific items are detailed in the HLQ questionnaire [14].

Mastery

Mastery of health was treated as an ordinal observed variable, measured using a 7-item scale adapted from the Pearlin Mastery Scale [11]. This construct captures an individual's sense of personal control over health outcomes, conceptually distinct from both self-management skills and health literacy [11]. Participants responded to each item on a 4-point Likert scale: "strongly disagree," "disagree," "agree," and "strongly agree." A total mastery score was computed by summing responses across all seven items, yielding a possible range from 7 to 28, with higher scores indicating a greater sense of mastery over one's health.

2.3.3 Outcome variables

Comorbidities

Participants were presented with a list of common health conditions and asked to indicate whether they currently had, previously experienced, or never had the condition. The conditions included: lung diseases (e.g., asthma or chronic obstructive pulmonary disease), cardiovascular issues (heart attack, cardiac arrest, or other heart-related problems), stroke or intracranial bleeding, hypertension, any type of cancer, fractures (leg, hip, or spine), ulcers or other stomach-related problems, diabetes, and depression.

A total comorbidity score was calculated by summing the number of conditions reported as present, resulting in a count variable ranging from 0 to 10. Higher scores reflected a greater burden of additional health conditions alongside the primary iRMDs. This composite measure was used as one of the outcome variables.

Patient Global Assessment of Health

Patient global assessment of health was measured using a Visual Analogue Scale (VAS) where participants rated their overall health from 0 (worst possible health) to 10 (best possible health). This single-item measure captured participants' subjective assessment of their overall health status and was treated as an ordinal variable in the analysis.

Professional Severity Score

The professional severity score was an outcome variable reflecting the healthcare provider's assessment of the impact of the iRMD on the patient's functioning and overall health. It was recorded on a Likert-type scale ranging from 0 (no influence) to 10 (very severe influence). We treated this variable as an ordinal variable in the analysis.

2.4 Rationale: Structural Equation Modelling with Mediation Analysis

Structural Equation Modeling (SEM) is a multivariate statistical technique that enables researchers to examine complex relationships among variables simultaneously [15]. Unlike traditional statistical methods like simple or multiple regression analysis, SEM analyzes both observed and unobserved (latent) variables within a single unified framework. This makes it particularly valuable in social and behavioral sciences where abstract constructs are common [16]. The primary advantage of SEM over traditional regression approaches lies in its ability to account for measurement error, which often produces more accurate parameter estimates and reduces bias in research findings [17][18]. Furthermore, SEM allows for the assessment of both direct and indirect effects among variables, providing a more comprehensive understanding of complex theoretical relationships.

SEM integrates several analytical techniques, with Confirmatory Factor Analysis (CFA) serving as a cornerstone of the measurement model component. CFA tests predetermined relationships between latent constructs and their observed indicators, therefore requiring the specification of the factor structure in advance based on theoretical foundations or prior research [19]. This confirmatory approach validates whether measured variables accurately represent the underlying theoretical constructs. In contrast, one can also do Exploratory Factor Analysis (EFA) which typically precedes CFA when the underlying structure is unknown. EFA helps to identify potential latent constructs without assuming prior theoretical models. Once factor structures are established through EFA, CFA validates these structures before proceeding to structural modeling.

In addition to the measurement models, there is path analysis which forms the structural component of SEM, extending multiple regression by examining both direct and indirect relationships among variables [16][18]. When combined with measurement models from CFA, SEM can model the causal relationships between latent or theoretical concepts that cannot be measured directly but are inferred from observable indicators. These latent variables might include abstract ideas such as SES or health literacy, while observable variables act as proxies. For example, SES is a multi-dimensional construct which can be represented by indicators such as education level, employment status, or income [20], while health literacy can be measured using a validated instrument like the Health Literacy Questionnaire [14].

One important application of path analysis within SEM is mediation analysis, which explores whether the relationship between two variables is explained by a third, intermediate variable which is called the mediator [21][22]. In the context of this study, health literacy may mediate the relationship between SES and health outcomes. That is, individuals with higher SES may have better health literacy, which in turn contributes to improved health outcomes. SEM enables the simultaneous testing of these direct effects for example, SES directly affecting health outcomes and indirect effects for example, SES influencing health outcomes through health literacy.

Once the model is specified and estimated, regardless of whether it includes mediation pathways, latent constructs from CFA, or structures derived from EFA, it is evaluated for how well it fits the observed data. This is typically done using a set of model fit indices, each providing a different perspective on the adequacy of the model. Further details on fit indices are outline in Subsection 3.5

3 Statisical Data Analysis

3.1 Statistical Software

Data analysis was performed using `R` version 4.4.0. The following R packages were used for data management, analysis, and visualization: `lavaan`, `DiagrammeR`, `ggplot2`, `tidyr`, `psych`, `GGally`, and `haven`.

3.2 Descriptive Statistics

Descriptive analyses were conducted overall and stratified by inflammatory rheumatic and musculoskeletal disease (iRMD) type for the socioeconomic status (SES) indicator variables. Frequencies and relative percentages were computed for categorical data, while means and standard deviations were computed for continuous variables. The categories for the variables are outlined in the Data Variables section (see Subsection 2.3).

3.3 Structural Equation Modeling Procedure

To answer the research questions, we employed structural equation modeling in stages. This process begins with confirmatory factor analysis to establish the measurement models for our latent constructs. Subsequently, we proceed to the structural component, focusing specifically on mediation analysis, which integrates the measurement models with path analysis to examine hypothesized relationships. The analytical stages are outlined below:

3.3.1 Confirmatory Factor Analysis: Measurement Models

Confirmatory Factor Analysis (CFA) was used to specify the relationship between observed indicators and the two underlying latent constructs: socioeconomic status and health literacy. The general measurement model is defined as:

$$\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta} \tag{1}$$

Where:

- $\mathbf{x}$ is a vector of observed variables (indicators),
- $\boldsymbol{\Lambda}$ is the matrix of factor loadings,
- $\boldsymbol{\xi}$ is the vector of latent variables (e.g., SES or HL),
- $\boldsymbol{\delta}$ is the vector of measurement errors.

For each latent construct, observed indicators were selected based on theoretical relevance and prior literature. Each measurement model was estimated separately prior to inclusion in the structural equation models.

3.3.2 Measurement model for Socio Economic Status

To identify the strongest contributors to the latent construct of socioeconomic status (SES), a confirmatory factor analysis (CFA) was conducted using six observed indicators selected a priori (see Data Variables Subsubsection 2.3.1 for variable definitions). The hypothesis was that these variables would significantly load onto a single latent SES variable. The specific measurement equations are given by:

$$\text{Employment Status} = 1.0 \times \xi_{SES} + \delta_1 \quad \text{(reference indicator)} \tag{2}$$
$$\text{Education Level} = \lambda_2 \xi_{SES} + \delta_2 \tag{3}$$
$$\text{Living Situation} = \lambda_3 \xi_{SES} + \delta_3 \tag{4}$$
$$\text{Gender} = \lambda_4 \xi_{SES} + \delta_4 \tag{5}$$
$$\text{Age} = \lambda_5 \xi_{SES} + \delta_5 \tag{6}$$
$$\text{Migration Background} = \lambda_6 \xi_{SES} + \delta_6 \tag{7}$$

Given that the observed indicators include both categorical and continuous variables, the CFA was estimated using the Weighted Least Squares Mean and Variance-adjusted (WLSMV) estimator. WLSMV is particularly well-suited for this analysis as it does not assume multivariate normality and provides robust parameter estimates and standard errors when working with ordinal or categorical data [25], making it ideal for CFA models with mixed variable types.

Standardized factor loadings were examined to assess the strength of association between each observed indicator and the SES construct. Standardized factor loading enables direct comparison of contributions across indicators measured on different scales. Variables that did not load significantly onto the latent factor would be excluded, and the model re-estimated using only significant indicators. The final interpretation would focus on those variables with the highest factor loadings, representing the strongest contributors to the latent SES construct. Model fit was evaluated using fit indices as outlined in Subsection 3.5

3.3.3 Measurement Model for Health Literacy

Health literacy (HL) was modeled as a latent construct measured by nine domain from the validated Health Literacy Questionnaire (HLQ), used in the original study (see Table 1) for the observed domain names. Each domain represented a distinct dimension of HL and was treated as an observed indicator in the confirmatory factor analysis (CFA).

One domain was fixed to 1.0 to serve as the reference indicator for model identification, while all other factor loadings were freely estimated. Standardized factor loadings were examined to determine the strength of association between each domain and the latent HL construct. Domains with higher loadings were interpreted as the strongest contributors to health literacy in this population.

The CFA model for HL was specified analogously to that of SES, and the full set of measurement equations is presented in the Supplementary material (see Subsection 8.2). Model fit was evaluated using the indices described in Subsection 3.5.

3.3.4 Measurement model assumptions

Confirmatory Factor Analysis (CFA) relies on several important assumptions to ensure valid and reliable results. First, it assumes a linear relationship between latent variables and their observed indicators, meaning that changes in the latent construct are reflected proportionally in the observed measures. Second, independence of observations assumes each data point represents a separate, unrelated case. This assumption was supported by evidence from the primary study showing no substantial differences between hospitals in demographic and health variables [7], indicating that patients' characteristics were not systematically influenced by which center they attended. Third, the selection of observed indicators must be theoretically justified to ensure meaningful representation of the underlying latent construct, which was achieved by selecting variables based on prior literature and conceptual relevance. Fourth, adequate sample size is critical for stable parameter estimates. The project included 895 participants, exceeding commonly recommended thresholds for CFA such as a minimum of 200 cases or 5–10 observations per estimated parameter [18]. Finally, while multivariate normality is typically assumed for maximum likelihood estimation, the Weighted Least Squares Mean and Variance-adjusted (WLSMV) estimator was employed to address potential normality violations and accommodate the mix of categorical and continuous indicators used in this analysis.

3.4 Mediation Analysis

The second part of our SEM involved mediation analysis. This was used to investigate the mechanisms through which socioeconomic status influences health outcomes. Specifically, we examined three models: (1) the pathway from SES through health literacy to outcomes, (2) the pathway from SES through mastery to outcomes, (3) a sequential mediation model, the pathway from SES through health literacy and mastery to outcomes. Each model tested these pathways across the three health outcomes.

Given that all three models follow the same analytical structure, we present the first model (primary objective) in full detail as a worked example. The remaining two models, along with their corresponding path diagrams and statistical specifications, are provided in the Supplementary materials (see Subsection 8.3).

3.4.1 Mediation analysis: SES-Health Literacy-Outcomes Mediation Model

To answer the primary objective which sought to examine whether health literacy mediates the relationship between socioeconomic status and each of the three health outcomes among iRMDs patients, a conceptual framework illustrated in Figure 2 was followed. Here, socioeconomic status (X) serves as the independent variable, health literacy (M) is the mediator, and the outcomes (Y) represent the dependent variables. In the model, path a represents the effect of socioeconomic status on health literacy, path b captures the effect of health literacy on the outcome while adjusting for socioeconomic status, and path c denotes the direct effect of socioeconomic status on the outcome after accounting for the mediator.

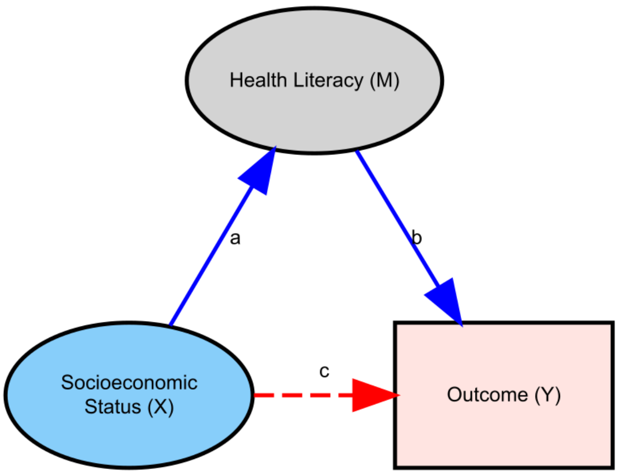


Figure 2: Path diagram illustrating the mediation model with health literacy as mediator between SES and health outcomes. Oval shapes represent latent variables (theoretical constructs measured through multiple indicators), while rectangles represent observed variables (directly measured indicators). Solid arrows represent hypothesized causal relationships, with path coefficients a, b, and c indicating the strength and direction of these relationships.

The decision on whether or not to conduct mediation analysis was based on steps outlined by Baron and Kenny (1986) [21]. Using their approach on our conceptual model, firstly, socioeconomic status (X) must significantly predict the outcome (Y). Second, socioeconomic status (X) must significantly predict the proposed mediator, health literacy (M). Third, health literacy (M) must significantly predict the outcome (Y). If these conditions are met, a mediation analysis is justified, and a fourth step is performed to test whether health literacy (M) significantly predicts the outcome (Y) while controlling for socioeconomic status (X). These relationships can be expressed through the following system of regression equations:

$$Y = b_{01} + cX + e_1 \tag{8}$$

$$M = b_{02} + aX + e_2 \tag{9}$$

$$Y = b_{03} + bM + e_3 \tag{10}$$

$$Y = b_{04} + c'X + bM + e_4 \tag{11}$$

The terms b_{01}, b_{02}, b_{03}, and b_{04} represent the intercepts for each respective equation, while e_1, e_2, e_3, and e_4 denote the residuals or unexplained variance. Equation (8) specifies the total effect (c) of the independent variable X on the outcome variable Y. Equation (9) estimates the effect (a) of X on the mediator M. Equation (10) reflects the effect (b) of the mediator M on the outcome Y, without adjusting for X. Finally, Equation (11) represents the full mediation model, where both the mediator M and independent variable X are included as predictors of Y. In this model, c' denotes the direct effect of X on Y after accounting for the mediation pathway through M, and b represents the effect of the mediator M on the outcome Y. The indirect effect of X on Y via M is quantified by the product ab, and the total effect decomposes as $c = c' + ab$. The proportion mediated is calculated as *indirect effect / total effect*.

However, if one or more of these hypotheses are not supported, the evidence for mediation is absent. For example, if socioeconomic status X does not significantly predict the outcome Y in Equation 8, there is no total effect to be mediated. If X does not significantly predict the mediator M in Equation 9, the pathway through the mediator is not supported. Similarly, if M does not significantly predict Y in Equation 10 when controlling for X, its role as a mediator is not established. Even if the individual paths are significant, a non-significant indirect effect (ab) provides no statistical support for mediation. In such instances, the results suggest that health literacy may not mediate the relationship between socioeconomic status and the outcomes.[26][27][28]

It is also important to note that the decomposition $c = c' + ab$ (where ab is the indirect effect) holds exactly under certain conditions, including the use of simple multiple regression or SEM without latent variables [29]. However, in our study, we employed SEM with latent variables which can lead to discrepancies between the directly estimated total effect (c) and the sum of the direct and indirect effects ($c' + ab$). Therefore, we interpreted the total effect as approximately equal to $c' + ab$, rather than identical.

3.5 Model Fit Assessment

The adequacy of both the measurement and structural models was evaluated using multiple fit indices to provide a comprehensive assessment of model fit. The Chi-square test examines the null hypothesis that the model-implied covariance matrix equals the observed covariance matrix; however, it is sensitive to sample size, meaning that with large samples, even trivial model misspecifications can result in significant chi-square values and model rejection [18][23]. Therefore, we supplemented this with incremental fit indices including the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), which compare the proposed model's fit relative to a baseline null model that assumes no relationships among variables. Additionally, we employed approximate fit measures: the Root Mean Square Error of Approximation (RMSEA), which evaluates model fit per degree of freedom and includes a penalty for model complexity, and the Standardized Root Mean Square Residual (SRMR), which quantifies the standardized difference between the observed and model-implied correlation matrices.

Thus, we assessed model fit using the following established criteria: CFI and TLI values above 0.90 were considered acceptable fit and above 0.95 as good fit; SRMR values below 0.08 were considered indicative of good fit; RMSEA values below 0.08 were considered acceptable fit and below 0.05 as good fit. The RMSEA was also evaluated along with its 90% confidence interval. [23][24].

3.6 Analytical Considerations for Selected Variables

Our analysis encountered some methodological challenges that required careful consideration and transparent reporting.

Missing Professional Severity Scores. The first challenge stemmed from the study's sampling design: professional severity scores were only available for 778 of the 895 patients (see Section 2.2). This missingness was by design, rather than due to patient non-response. As such, no imputation was applied, and analyses were conducted using available cases for this outcome variable. Bakker et al. [12] reported no systematic differences in baseline patient characteristics between patients with and without professional severity data.

Count Variable Constraints. The second challenge involved the comorbidity outcome variable. The number of comorbidities per patient represents count data, which is ideally modeled using Poisson or negative binomial regression. However, the `lavaan` package used for structural equation modeling does not currently support generalized SEM for count outcomes. Consequently, comorbidities were treated as ordinal variables and analyzed using the WLSMV estimator. While this approach is methodologically defensible, it introduces approximation into model estimates and should be interpreted accordingly.

Missing data Mastery. The final issue concerned the mastery mediator. Mastery data were available for 894 patients out the 895 enrolled. One respondent was excluded because the mastery scale was not available in Arabic, as reported by Bakker et al. [12]. Given the minimal extent and known reason for missingness, we considered it negligible and conducted analyses using complete cases.

These analytical considerations informed both our modeling strategy and the interpretation of findings, and are further addressed in the discussion on methodological limitations.

4 Results

4.1 Descriptive statistics

We began our analysis with descriptive statistics to characterise the study population and explore the distributions of key variables related to socioeconomic status (SES), and the three outcomes: number of comorbidities, patient global assessment of health and the professional severity score. These summaries provide the necessary context for understanding and interpreting the relationships explored later in the models.

Table 2: Descriptive statistics stratified by disease type. Continuous variables are presented as Mean±SD and categorical variables are presented as number (%)

Variable	Description	Total (n=895)	RA (n=369)	SpA (n=319)	Gout (n=207)
Gender	Female	436(48.72)	259(70.18)	147(46.08)	30(14.49)
	Male	459(51.28)	110(29.81)	172(53.92)	177(85.51)
Age	Mean±SD	61.07±13.92	64.14±12.93	55.08±13.46	64.83±13.28
Education	Low	454(50.73)	203 (55.01)	138(43.26)	113(54.59)
	Medium	222(24.80)	82(22.22)	88(27.59)	52(25.12)
	High	219(24.47)	84(22.76)	93(29.15)	42(20.29)
Migration Background	Non Western	74(8.27)	23(6.23)	30(9.40)	21(10.14)
	Western	83(9.27)	33(8.94)	32(10.03)	18(8.70)
	Native	738(82.46)	313(84.82)	257(80.57)	168(81.16)
Employment	Not employed	30(3.35)	8(2.17)	14(4.39)	8(3.87)
	Work disabled	257 (28.72)	130(35.23)	97(30.41)	30(14.49)
	Retired	324(36.20)	145(39.30)	69(21.63)	110(53.14)
	Employed	284(31.73)	86(23.30)	139(43.57)	59(28.50)
Living Situation	Living alone	220(24.58)	102(27.64)	68(21.32)	50(24.15)
	Living with others	675(75.42)	267(72.36)	251(78.68)	157(75.85)
Comobidities	0	188(21.01)	76(20.60)	88(27.59)	24(11.59)
	1-2	444(49.61)	198(53.66)	159(49.84)	87(42.03)
	3-5	238(26.59)	89(24.12)	65(20.38)	84(40.58)
	6-10	25(2.80)	6(1.63)	7(2.19)	12(5.80)
Patient score	Mean±SD	6.41(1.75)	6.46(1.63)	6.14(1.97)	6.74(1.53)
Physician score	Mean±SD	4.23(2.39)	4.39(2.32)	4.78(2.30)	3.39(2.42)

Descriptive characteristics of the full study sample (n = 895) are presented in Table 2. Patients had a mean age of 61.1 years (SD = 13.9), and 48.7% were female. Most patients had low educational attainment (50.7%), and the majority were of native background (82%), and living with others (75.4%). Regarding employment, 36% were retired and 29% were work disabled. Comorbidities were common, with nearly 80% of participants reporting at least one, and 29.4% reporting three or more.

The descriptive statistics are also shown for the different iRMDs (rheumatoid arthritis (RA), spondyloarthritis (SpA), and gout to describe patient distribution. RA patients tended to be older and predominantly female, with a higher rate of incapacity and comorbidity. SpA patients were younger and more likely to be employed, with a more balanced gender distribution. Gout patients were mostly male, older, and had the highest comorbidity burden.

4.2 Structural Equation Modeling Procedure

Next, we outline the structural equation modeling procedure, starting with confirmatory factor analysis (CFA) to determine the contribution of our observed indicators to our latent variables. These measurement models serve as an important part of the structural models both for the primary and secondary objectives. The findings and detailed descriptions are found in the following subsections: The measurement models for SES and HL, are found on Subsubsections 4.2.1 and 4.2.2 respectively. The results for the structural models (mediation analysis) are found on Subsections 4.3 through 4.5.

4.2.1 Confirmatory Factor Analysis: Measurement Model for Socio Economic Status

Table 3: Confirmatory factor analysis for the latent variable socioeconomic status

Latent Variable: SES	**Baseline model**				**Without Migration Background**			
	Est.	**SE**	**P-val**	**Std. All**	**Est.**	**SE**	**P-val**	**Std. All**
Employment Status	1.000	–	–	0.499	1.000	–	–	0.491
Education Level	0.900	0.159	< 0.001	0.449	0.916	0.161	< 0.001	0.450
Living Situation	0.918	0.183	< 0.001	0.458	0.936	0.186	< 0.001	0.459
Gender	0.537	0.138	< 0.001	0.268	0.543	0.140	< 0.001	0.266
Age	-11.186	2.024	< 0.001	-0.401	-11.599	2.096	< 0.001	-0.409
Migration Background	0.045	0.129	0.724	0.023	–	–	–	–

Table 3 shows a confirmatory factor analysis for the latent variable socioeconomic status. The CFA was conducted under the assumption that the proposed six observed indicators would load onto the single latent variable SES. However, upon fitting the first measurement model, migration background had a negligible and non-significant association with the latent construct (standardized loading = 0.023, p = 0.724), indicating that it does not meaningfully contribute to SES latent variable and was excluded from the model.

As a result, a second model was fitted without migration background and the remaining five indicators retained significant standardised factor loadings. From this model, Employment status (standardised loading = 0.499), Living situation (standardised loading = 0.458) and education level (standardised loading = 0.449) were the strongest contributors of SES, indicating that those who are employed, live with others and having a higher education were more likely to have a higher SES. Age showed a negative association (standardised loading = -0.40), suggesting that older patients tended to have a lower SES. Gender contributed more weakly (standardised loading = 0.27), with men more likely having slightly higher SES than women.

While the model demonstrated reasonable approximate fit RMSEA = 0.085[0.061; 0.111, SRMR = 0.070], incremental fit indices (CFI = 0.857, TLI = 0.714) were below the conventional thresholds ($\geq 0.90-0.95$) $\chi^2_{WLSMV}(5) = 37.127$, p< 0.001. However, given the complexity of measuring socioeconomic status and the theoretical coherence of the latent variable structure, the five retained indicators were considered adequate for constructing the measurement model for the SES latent variable in the subsequent analyses.

4.2.2 Confirmatory Factor Analysis: Measurement Model for Health literacy

Table 4: Confirmatory factor analysis for the latent variable health literacy

Indicator	Estimate	SE	P-value	Std. All
Healthcare Providers support(HPS)	1.000	–	–	0.646
Having sufficient information (HSI)	1.144	0.063	<0.001	0.739
Actively managing health (AMH)	0.646	0.065	<0.001	0.417
Social support (SS)	0.841	0.056	<0.001	0.544
Appraisal of health(CA)	0.746	0.064	<0.001	0.482
Ability to engage (AE)	1.221	0.079	<0.001	0.789
Navigating healthcare system (NHS)	1.236	0.081	<0.001	0.799
Finding health information (FHI)	1.146	0.081	<0.001	0.740
Understanding health information (UHI)	1.133	0.082	<0.001	0.732

The measurement model for the latent variable Health literacy retained all nine proposed domains (p< 0.001) as indicated by Table 4. Standardized factor loadings ranged from 0.417 to 0.799. The strongest contributors to health literacy were "Navigating the Healthcare System" (0.799), and "Ability to Engage with Providers" (0.789) suggesting that the more functional and interactive aspects of health literacy were central to the construct among the patients. In addition, the need for information that is "Having sufficient information", "Finding health information" and "Understanding health information" all contributed significantly to the latent construct, with standardised loading approximately (0.74) for the three. Other domains, such as "Social Support" (0.544), "Appraisal of Health Information" (0.482), and "Actively Managing Health" (0.417), contributed less but still meaningfully. Model fit was acceptable: $\chi^2_{WLSMV}(27) = 213.822$, p< 0.001; CFI = 0.953; TLI = 0.938; RMSEA = 0.088 [0.077, 0.099]; and SRMR = 0.096. While the RMSEA and SRMR slightly exceeded ideal cutoffs, the high CFI and TLI values support a reasonably acceptable overall model fit.

Mediation analysis

4.3 Primary Objective: Examining whether health literacy mediates the relation between socioeconomic factors and the three outcomes among iRMD patients

Building on the results of the measurement models, we proceeded to address the study's primary objective using structural equation modelling (SEM). Prior to conducting the mediation analysis, we first performed an individual path analysis between SES, HL, and each of the health outcomes. This preliminary step was critical to determine whether the conditions to perform mediation analysis were met.

Following this preliminary assessment, we estimated the full SEM to test whether health literacy mediates the relationship between SES and each of the three health outcomes. Results are presented in both tabular form and path diagrams to facilitate clear interpretation of the relationships among the latent constructs and observed variables. Mediation was assessed by decomposing the total effect of SES on each outcome into its direct and indirect components, with the latter operating through HL. The proportion of the effect mediated was also calculated to quantify the extent to which HL explained the SES-health outcomes path. Standardized coefficients were used throughout, enabling comparison across constructs that differ in measurement scale or are latent.

We also report the relative contributions of individual SES and HL indicators based on their standardized factor loadings. The results of the SEM are presented separately for each of the three outcomes, detailing the direct, indirect, and total effects. These findings for the outcomes: patient global assessment of health, professional severity score and comorbidities are detailed in Subsection 4.3.1; 4.3.2 and 4.3.3 respectively.

4.3.1 Health literacy as mediator: Patient health score outcome

Determining the significance of the mediation paths

To determine whether mediation analysis could be conducted for the above mentioned objective, we assessed the association between SES and health literacy (path a), health literacy and patient global assessment of health (path b), and SES and patient global assessment of health directly (path c) as shown in Table 5. All three paths were statistically significant (p $<$ 0.001), suggesting the presence of both direct and indirect effects. The analysis was based on data from 895 patients. Given these results, we proceeded with formal mediation analysis.

Table 5: Individual path analysis for the outcome patient global assessment of health

Model	Estimate	Std. Error	p-value	Std. All
SES$\rightarrow$Health literacy(a)	0.673	0.136	< 0.001	0.358
Health Literacy $\rightarrow$ Patient Health(b)	0.672	0.115	< 0.001	0.245
SES $\rightarrow$ Patient Health (c)	0.628	0.148	< 0.001	0.231

Mediation analysis results

Results show that upon inclusion of health literacy and patient health score outcome in the model, the contribution of gender to the SES construct became very low and statistically non-significant as shown in Table 6 (standardised factor=0.006, p value= 0.931). This attenuation implies that gender may not have meaningful influence on the SES-patient global assessment of health outcome relationship once variation in health literacy is accounted for. Meanwhile, education stood out and demonstrated a notably stronger association (0.679, $p < 0.001$). For the health literacy construct, Navigating health care system and Ability to engage with healthcare providers stood out with significant factor loadings

Table 6: Structural equation model path coefficients. Paths b and c' are estimated simultaneously in the same equation predicting Patient Health Score.

Latent Variable: SES	**Estimate**	**Std Error**	**P-value**	**Std. All**
Employment Status	1.000	–	–	0.379
Education Level	1.792	0.353	< 0.001	0.679
Living Situation	1.112	0.265	< 0.001	0.422
Age	-12.946	2.887	< 0.001	-0.353
Gender	0.016	0.180	0.931	0.006
Latent Variable: Health Literacy	**Estimate**	**Std Error**	**P-value**	**Std. All**
Healthcare providers support(HPS)	1.000	–	–	0.645
Having sufficient information (HSI)	1.106	0.045	< 0.001	0.713
Actively managing health (AMH)	0.649	0.048	< 0.001	0.418
Social support (SS)	0.846	0.041	< 0.001	0.545
Appraisal of health(CA)	0.753	0.049	< 0.001	0.485
Ability to engage (AE)	1.234	0.058	< 0.001	0.796
Navigating healthcare system (NHS)	1.244	0.061	< 0.001	0.802
Finding health information (FHI)	1.204	0.063	< 0.001	0.776
Understanding health information (UHI)	1.193	0.063	< 0.001	0.769
Regressions	**Estimate**	**Std Error**	**P-value**	**Std. All**
Health Literacy $\sim$ SES (a)	0.602	0.119	< 0.001	0.354
Patient Health Score $\sim$ Health Literacy (b)	0.572	0.105	< 0.001	0.210
Patient Health Score$\sim$ SES (c')	0.448	0.242	0.064	0.097
Defined Parameters	**Estimate**	**Std Error**	**P-value**	**Std. All**
Indirect Effect (a $\times$ b)	0.344	0.083	< 0.001	0.074
Total Effect (c'+ a $\times$ b)	0.792	0.238	0.001	0.171
Proportion Mediated	0.434	–	–	0.434

of 0.802 and 0.796 respectively.

The model fit was acceptable: $\chi^2_{WLSMV}(88) = 731.305$, p< 0.001; CFI = 0.926; TLI = 0.9312 RMSEA = 0.090 [90% CI: 0.084, 0.0979]; and SRMR = 0.087. While the RMSEA and SRMR slightly exceeded ideal cutoffs, the high CFI and TLI values support a reasonably acceptable overall fit between the model and the observed data.

Presentation of results as a Path diagram

In addition to results from Table 6, a path diagram is presented in Figure 3. The mediation analysis revealed significant indirect effects through health literacy. The standardized coefficient from SES to health literacy (path a) was 0.354 ($p < 0.001$), indicating that for every one standard deviation increase in SES, health literacy increased by approximately 0.354 standard deviations. The path from health literacy to patient health (path b) yielded a coefficient of 0.210 ($p < 0.001$), meaning that every one standard deviation increase in health literacy is associated with a 0.21 standard deviation

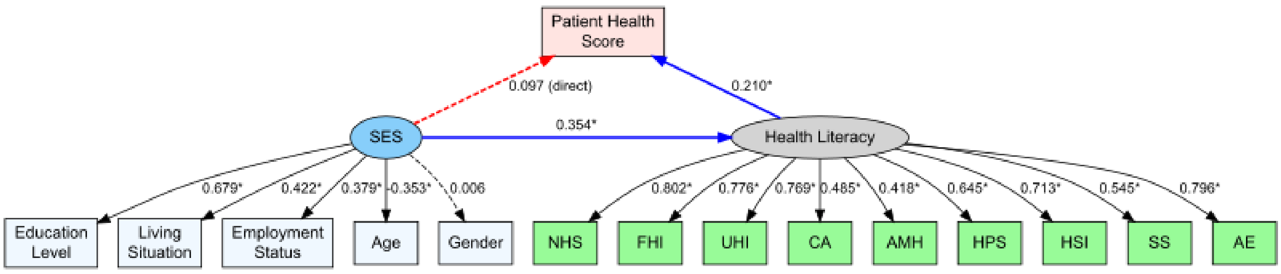


Figure 3: Path diagram for the self-reported health score outcome. The blue arrows represent the indirect paths, while the red dashed arrow represents the direct effect of SES on the outcome. Significant standardised path coefficients are indicated with an asterisk (*)($p < 0.05$). All constructs are scored such that higher values reflect more favourable outcomes: higher SES indicates greater socioeconomic advantage, higher health literacy reflects better health knowledge and skills, and higher self-reported health scores indicate better perceived health.

improvement in patient health, while controlling for SES.

The indirect effect, computed as $a \times b = 0.354 \times 0.210 = 0.074$, was significant and the direct effect of SES on patient health (path c') was 0.097 and was not statistically significant ($p > 0.05$), suggesting that SES does not have a meaningful direct effect on patient health after accounting for the mediating role of health literacy. The total effect, calculated as $c' + (a \times b) = 0.097 + 0.074 = 0.171$, captures the overall association between SES and patient health.

The proportion mediated was $\frac{a \times b}{\text{total effect}} = \frac{0.074}{0.171} \approx 0.434$, indicating that approximately 43.4% of the total effect of SES on patient health was explained by the mediating pathway through health literacy. While the direct effect was non-significant, suggesting health literacy is an important mediator, the remaining portion of the total effect likely operates through other unmeasured mediators not included in this model.

4.3.2 Health literacy as mediator: Professional severity score outcome

Determining the significance of the mediation paths

Table 7: Individual path analysis for the outcome professional severity score

Model	Estimate	Std. Error	p-value	Std. All
SES → Health Literacy (a)	0.673	0.136	< 0.001	0.358
Health Literacy → Professional Severity Score (b)	-0.437	0.143	0.002	-0.122
SES → Professional Severity Score (c)	-0.828	0.241	0.001	-0.192

Individual path results for the professional severity scores in Table 7, shows that all three paths were statistically significant, with p values below 0.001. The analyses were based on responses from 778 patients. Given these findings, a mediation analysis was subsequently performed to formally assess whether health literacy mediates the relationship between SES and professional severity score.

Mediation analysis results

Table 8: Structural Equation Modelling with mediation analysis results

Latent Variable: SES	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Employment Status	1.000	–	–	0.370
Education Level	2.035	0.413	< 0.001	0.754
Living Situation	1.075	0.270	< 0.001	0.398
Age	-13.774	3.074	< 0.001	-0.367
Gender	0.071	0.192	0.714	0.026
Latent Variable: Health Literacy	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Healthcare providers support(HPS)	1.000	–	–	0.652
Having sufficient information (HSI)	1.108	0.048	< 0.001	0.716
Actively managing health (AMH)	0.636	0.049	< 0.001	0.419
Social support (SS)	0.837	0.043	< 0.001	0.558
Appraisal of health(CA)	0.727	0.051	< 0.001	0.484
Ability to engage (AE)	1.201	0.060	< 0.001	0.795
Navigating healthcare system (NHS)	1.227	0.065	< 0.001	0.803
Finding health information (FHI)	1.198	0.067	< 0.001	0.779
Understanding health information (UHI)	1.172	0.065	< 0.001	0.775
Regressions	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Health Literacy $\sim$ SES (a)	0.643	0.130	< 0.001	0.357
Professional severity score $\sim$ Health Literacy (b)	-0.333	0.149	0.025	-0.093
Professional severity score $\sim$ SES (c')	-0.563	0.350	0.108	-0.087
Defined Parameters	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Indirect Effect (a $\times$ b)	-0.214	0.100	0.031	-0.033
Total Effect (c'+ a $\times$ b)	-0.777	0.327	0.017	-0.120
Proportion Mediated	0.275	–	–	0.275

For the professional severity scores outcome, the overall model demonstrated good fit to the data ($\chi^2_{WLSMV}(88) = 608.093$, df = 88, $p < 0.001$; CFI = 0.930, TLI = 0.917, RMSEA = 0.087 [90% CI: 0.081, 0.094], SRMR = 0.087), with incremental fit indices slightly exceeding conventional thresholds and approximate fit measures within acceptable ranges.

The SES showed similar trends to the patient global assessment of health outcome; gender showed a lower and non significant contribution (0.026, $p = 0.714$), while education level demonstrated a notably stronger association (0.754, $p < 0.001$). For health literacy, navigating healthcare system and ability to engage with healthcare providers also had the most contribution with significant factor loadings of 0.803 and 0.795 respectively. The structural relationships and their corresponding effects are explored in detail through the path diagram presented in Figure 4, including the results for direct, indirect, and the total effects .

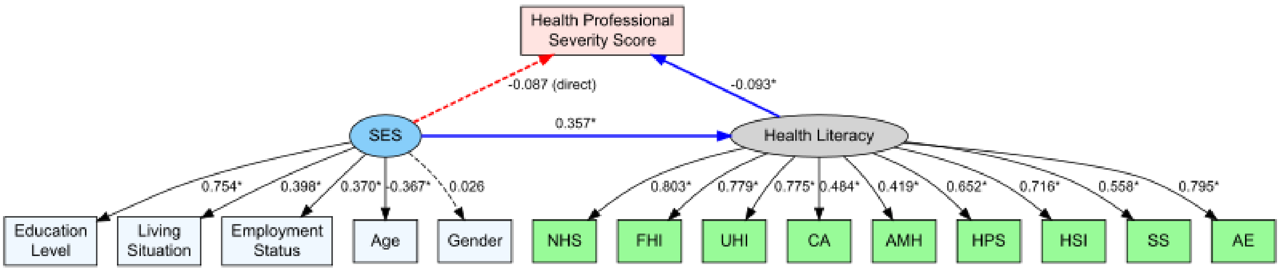


Figure 4: Path diagram for the professional severity score outcome. Significant standardised path coefficients are indicated with an asterisk (*) for significant associations ($p < 0.05$). For this outcome, higher SES and HL constructs indicate better health hence lower severity scores.

Presentation of results as a Path diagram

The mediation analysis showed that a one standard deviation increase in SES was associated with a 0.357 standard deviation increase in health literacy ($p < 0.001$), indicating that individuals with higher socioeconomic status tend to possess better health-related knowledge and skills. In turn, a one standard deviation increase in health literacy was associated with a 0.093 standard deviation decrease in professional health severity scores ($-0.093, p = 0.025$), suggesting that better health understanding translates to reports of less severe health outcomes.

These individual pathways combined to produce a significant indirect effect ($a \times b = -0.033$, $p = 0.031$), confirming health literacy as a meaningful mediator in this relationship. However, when health literacy was included in the model, the direct association between SES and professional severity score outcome became non-significant ($-0.087, p = 0.108$) pointing to health literacy as a key mechanism rather than SES having independent effects on the professional severity score.

Overall, the total effect of SES on professional severity was -0.120, with the proportion mediated through the health literacy pathway accounting for approximately 27.5% of this association. While this represents a meaningful contribution, the majority of the SES effect appears to operate through additional pathways not captured in the current model.

4.3.3 Health literacy as mediator: Number of comorbidities outcome

Determining the significance of the mediation paths

Table 9: Individual path analysis for the outcome self- reported comorbidities

Model	Estimate	Std. Error	p-value	Std. All
SES → Health Literacy (a)	0.673	0.136	< 0.001	0.358
Health Literacy → Comorbidities (b)	-0.414	0.091	< 0.001	-0.169
SES → Comorbidities (c)	-1.630	0.244	< 0.001	-0.419

For the comorbidities outcome, all individual paths, SES to health literacy (a), health literacy to comorbidities (b), and SES directly to comorbidities (c) were statistically significant ($p < 0.001$), as shown in Table 9. Based on data analysis from 895 patients, these results supported the conditions for mediation, hence a formal mediation analysis was conducted.

Mediation analysis results

Table 10: Structural Equation Modelling with mediation analysis results

Latent Variable: SES	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Employment Status	1.000	–	–	0.353
Education Level	1.708	0.312	< 0.001	0.603
Living Situation	1.171	0.265	< 0.001	0.414
Age	-16.666	3.371	< 0.001	-0.423
Gender	-0.154	0.188	0.414	-0.054
Latent Variable: Health Literacy	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Healthcare providers support(HPS)	1.000	–	–	0.645
Having sufficient information (HSI)	1.097	0.044	< 0.001	0.708
Actively managing health (AMH)	0.653	0.048	< 0.001	0.422
Social support (SS)	0.837	0.041	< 0.001	0.540
Appraisal of health(CA)	0.755	0.050	< 0.001	0.487
Ability to engage (AE)	1.232	0.058	< 0.001	0.795
Navigating healthcare system (NHS)	1.239	0.061	< 0.001	0.799
Finding health information (FHI)	1.208	0.064	< 0.001	0.780
Understanding health information (UHI)	1.199	0.063	< 0.001	0.774
Regressions	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Health Literacy $\sim$ SES (a)	0.664	0.129	< 0.001	0.364
Comorbidities $\sim$ Health Literacy (b)	-0.099	0.093	0.288	-0.041
Comorbidities $\sim$ SES (c')	-1.713	0.331	< 0.001	-0.387
Defined Parameters	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Indirect Effect (a $\times$ b)	-0.066	0.059	0.268	-0.015
Total Effect	-1.779	0.315	< 0.001	-0.402
Proportion Mediated	–	–	–	–

For the comorbidities outcome, we estimated a structural equation model following the same theoretical framework. The model demonstrated acceptable fit to the data ($\chi^2_{WLSMV}(88) = 724.904$, df = 88, $p < 0.001$; CFI = 0.926, TLI = 0.912, RMSEA = 0.090 [90% CI: 0.084, 0.096], SRMR = 0.087), with incremental fit indices slightly above the conventional thresholds and approximate fit measures within acceptable ranges.

Consistent with the previous models, education level remained the strongest contributor (standardized loading = 0.603, $p < 0.001$), Gender again showed minimal and non significant contribution (standardized loading = -0.054, $p = 0.414$). Similary, the two domains, navigating healthcare system and ability to engage with healthcare providers also remained as the strongest contributors of health literacy with significant standardized factor loadings of 0.799 and 0.795 respectively.

The structural pathways examining how SES influences comorbidities directly and through health literacy are detailed in the corresponding path diagram below, which illustrates the direct, indirect, and total effects of this mediation model.

Presentation of results as a Path diagram

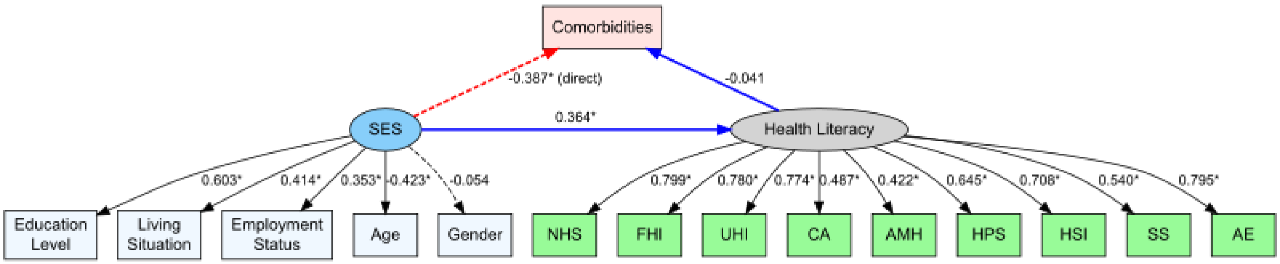


Figure 5: Path diagram for the number of self reported comorbidities outcome. Significant standardised path coefficients are indicated with an asterisk (*) for significant associations ($p < 0.05$). The SES and HL constructs are scored such that higher values reflect more favorable outcomes that is a lower number of comorbidities

The mediation analysis for comorbidities revealed a different pattern of relationships compared to the professional health severity model and the patient global assessment of health model. Results as presented in Figure 5 show that one standard deviation increase in SES is associated with a 0.364 standard deviation increase in health literacy ($p < 0.001$), demonstrating a consistent positive relationship between socioeconomic advantage and health knowledge across the outcome models. However, the pathway from health literacy to comorbidities showed a non-significant standardized coefficient (-0.041, $p = 0.288$), suggesting that health literacy does not meaningfully predict the number of comorbidities among the patients.

Given the non-significant health literacy-comorbidities pathway, the indirect effect through health literacy was also non-significant ($a \times b = -0.015$, $p = 0.268$), indicating that health literacy does not serve as a mediator in the SES and comorbidities relationship. In contrast, the direct effect of SES on comorbidities remained strong and significant (-0.387, $p < 0.001$), suggesting that a one standard deviation increase in SES is associated with a 0.387 standard deviation decrease in comorbidities, independent of health literacy.

The total effect of SES on comorbidities was (-0.402 $p < 0.001$), with this effect operating primarily through the direct pathway rather than through health literacy mediation. These findings suggest that health literacy might not be important in the mediating mechanism of the SES and comorbidities relationship. Instead, SES appears to influence comorbidities directly.

4.4 Secondary objective (1): Investigating whether mastery mediates the relation between socioeconomic factors and the three outcomes among the iRMD patients.

Following the primary objective mediation analysis involving health literacy (HL) as the mediator, we moved on to addressing the second objectives, starting with a simpler model, where mastery was used as a mediator between socioeconomic status (SES) and health outcomes. This analysis was motivated by two key considerations. Firstly, while HL has been widely recognized as a mediator linking SES to health, it does not fully capture the internal psychological resources individuals draw upon in managing iRMDs [9]. Second, mastery is conceptually more straightforward than health literacy, as it reflects an individual's perceived control over health and life circumstances, an aspect that may be more intuitively understood by both patients and healthcare providers.

This secondary analysis followed a similar structural framework as the primary model but focused on mastery as the sole mediator between SES and the three health outcomes. As in the primary analysis, we began by testing the individual associations between SES, mastery, and each outcome to determine whether the conditions for mediation modeling were met. Once confirmed, we specified and estimated the structural mediation model using the WLSMV estimator, consistent with our previous analytical approach. Results are reported in standardized form, including direct, indirect, and total effects, as well as the proportion of the SES effect mediated by mastery.

Details on individual path analysis are discussed in Subsection 8.5. The Structural Equation Model results for this section are presented as path diagrams and the regression tables are provided in the Supplementary material Tables 12; 14; 13. Detailed results for the outcomes Professional severity score; patient global assessment of health and the number of self-reported comorbidities are found in Subsections 4.4.1; 4.4.2; 4.4.3 respectively.

4.4.1 Mastery as mediator: Professional Severity Score Outcome

Following analysis of individual paths in Table 15, a mediation analysis was carried out with the mastery of health as the only mediator. In this simplified mediation model, SES had a significant positive effect on mastery of health, with a one standard deviation (SD) increase in SES associated with a 0.425 SD increase in mastery ($p < 0.001$). In turn, mastery of health was significantly associated with lower professional severity scores, such that a one SD increase in mastery corresponded to a 0.243 SD decrease in professional severity ($p < 0.001$). The indirect effect of SES on severity through mastery was statistically significant (unstandardized = -0.409, $p < 0.001$; standardized = -0.103). The direct effect of SES on professional severity was not statistically significant ($c' = -0.090$, $p = 0.116$), suggesting that most of the effect of SES on the outcome occurs through mastery of health. The total effect of SES on severity was significant (-0.766, $p < 0.001$), and the proportion of this effect mediated by mastery was 53.4%, calculated as the ratio of the indirect to total effect ($-0.409/-0.766 = 0.534$). Detailed estimates are presented in the Table 12 in the Supplementary material.

Model fit indices indicated that RMSEA (0.076 [90% CI: 0.059, 0.094]) and SRMR (0.071) were within commonly accepted thresholds, while CFI (0.861) and TLI (0.776) fell below the conventional cutoff

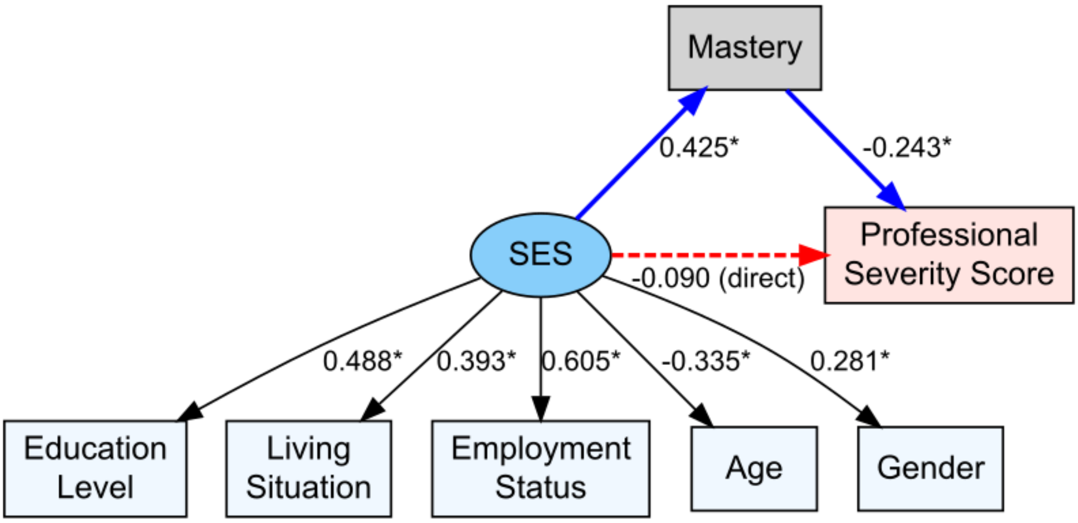


Figure 6: Path diagram for the outcome professional severity score. The blue solid lines represent the indirect effect of SES to Professional severity score through Mastery. Significant standardised path coefficients are indicated with an asterisk (*) (p <0.05). SES and Mastery are scored such that higher values reflect more favorable outcomes that is lower severity scores

of 0.90. The chi-square test was significant, ($\chi^2_{WLSMV}(13) = 71.368$, $p < 0.001$).

4.4.2 Mastery as mediator: Patient global assessment of health

Following significant results of the individual paths in Table 19, a mediation analysis was carried out with the mastery of health as the mediator. Path analysis indicated that higher SES was significantly associated with greater Mastery of Health ($a = 0.428$, $p < 0.001$), which in turn predicted better Patient Health outcomes ($b = 0.375$, $p < 0.001$). These standardized coefficients indicate the standard deviation change in outcome per one standard deviation increase in the predictor. The indirect effect of SES on patient health score through mastery of health was statistically significant (unstandardized $= 0.448$, standardized $= 0.161$, $p < 0.001$). The direct effect of SES on patient health score was not significant ($c' = 0.080$, $p = 0.0113$), indicating that the effect of SES on the patient health scores is mainly through mastery of health. Approximately 66.7% of the total effect of SES on the patient health score was mediated through mastery of health, indicating that mastery represents a key mechanism which links SES to health outcomes. Full regression estimates are presented in Table 13.

Model fit was showed that the Chi-square test was significant, $\chi^2_{WLSMV}(13) = 75.186$, $p < 0.001$. The CFI (0.882) and TLI (0.810) fell below the conventional thresholds for good fit, whereas the RMSEA (0.073; 90% CI: 0.058, 0.090) and SRMR (0.068) were within acceptable ranges.

4.4.3 Mastery as mediator: Number of comorbidities outcome

Mediation analysis was conducted following significant individual paths as detailed on Table 17. Results from the path analysis show that higher SES were significantly associated with greater sense

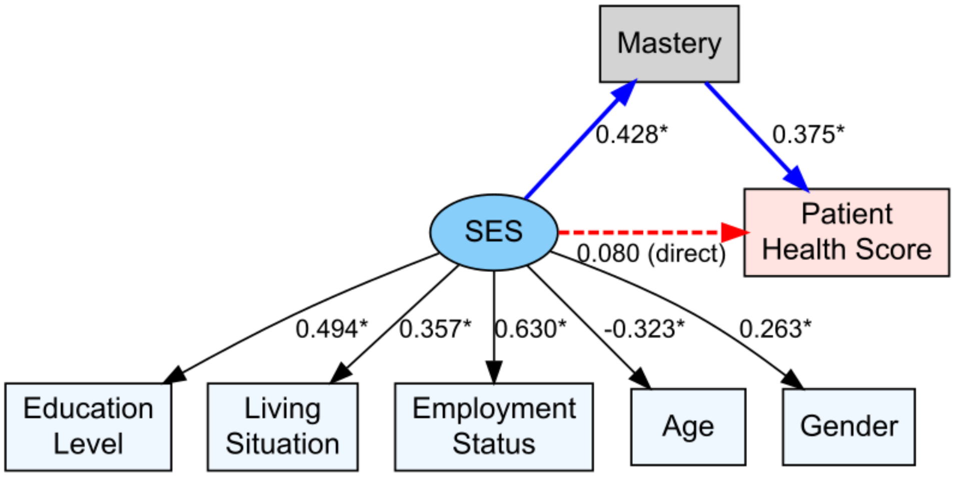


Figure 7: Path diagram for the patient health score outcome with mastery as the mediator. The blue solid lines represent the indirect effect of SES to patient health score through Mastery. Significant standardised path coefficients are indicated with an asterisk (*) (p <0.05). SES and Mastery are scored such that higher values reflect more favorable outcomes that is higher reported scores indicating better health

of mastery of health ($a = 0.430$, $p < 0.001$), and greater sense of mastery was associated with fewer comorbidities ($b = -0.138$, $p = 0.001$). The indirect effect of SES on comorbidities through mastery of health was statistically significant (standardized $= -0.059$, $p = 0.001$). The direct effect of SES on comorbidities was significant ($c' = -0.316$, $p < 0.001$), explaining most of the relationship. Only 15.8% of the total effect was mediated through mastery of health, indicating that SES has a direct effect on comorbidities reported with minimal mediation through mastery. Full regression estimates are reported in Table 14.

The model fit showed that the Chi-square was significant, $\chi^2_{WLSMV}(13) = 90.245$, $p < 0.001$, and CFI = 0.849, TLI = 0.756 fit indices fell below the conventional thresholds. The RMSEA = 0.082 [90% CI: 0.066, 0.098], and SRMR = 0.075 were within acceptable ranges.

4.5 Secondary objective (2): Investigating whether health literacy and mastery of health mediates the relationship between socio-economic factors and the three outcomes

Building on the earlier models in which health literacy and mastery were examined separately as mediators, we moved on to exploring a combined model that included both mediators simultaneously. This sequential mediation approach allowed us to assess whether health literacy may influence mastery, and whether together they mediate the relationship between socioeconomic status (SES) and health outcomes. This step served as a logical extension of the individual models and was motivated by the potential for these two modifiable mediators to act together along the SES-health outcome pathway.[10]

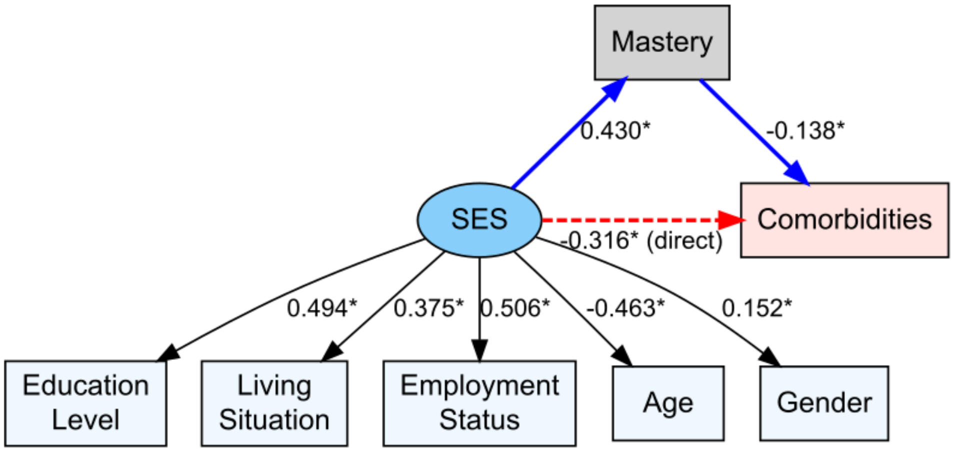


Figure 8: Path diagram for the number of comorbidities outcome with mastery as the mediator. The blue solid lines represent the indirect effect of SES to number of comorbidities through Mastery. Significant standardised path coefficients are indicated with an asterisk (*) (p <0.05). SES and Mastery are scored such that higher values reflect more favorable outcomes that is fewer number of comorbidities

As in the previous analyses, the analytic procedure remained consistent, with individual path analysis for the three outcomes ; professional severity score, patient global assessment of health and number of comorbidities showing significant paths. In addition, structural regression paths are also shown in Subsubsections 8.6.1, 8.6.3, 8.6.2 respectively in the Supplementary material. For the mediation analysis, standardized coefficients were used to report all direct, indirect, and total effects. The model was estimated using the WLSMV estimator. Particular attention was given to the sequential indirect effects, where the effect of SES on the outcomes is through both health literacy and mastery in a stepwise manner. Results for these are outlined in Subsections 4.5.1; 4.5.2; 4.5.3;

4.5.1 Health literacy and mastery mediators: Professional Severity Score outcome

Figure 9 illustrates the structural equation model estimated to examine the pathways linking socioeconomic status to professional health severity through health literacy and mastery of health as sequential mediators. The model demonstrated adequate fit to the data ($\chi^2_{WLSMV}(100) = 659.280$, $p < 0.001$; CFI = 0.934, TLI = 0.921, with RMSEA = 0.085 [90% CI: 0.079, 0.091], SRMR = 0.083), being slightly above the proposed conventional thresholds.

Analysis of the structural pathways revealed that a one standard deviation increase in socioeconomic status (SES) was associated with a 0.355 standard deviation increase in health literacy ($p < 0.001$), indicating that individuals with higher SES tend to possess stronger health-related knowledge and competencies. In turn, health literacy was positively associated with mastery of health (0.393, $p < 0.001$), which was negatively associated with professional severity scores (–0.282, $p < 0.001$), suggesting that greater perceived control over one's health corresponds with reduced disease severity scores.

The sequential mediation pathway yielded a statistically significant indirect effect ($a \times f \times e = -0.223$,

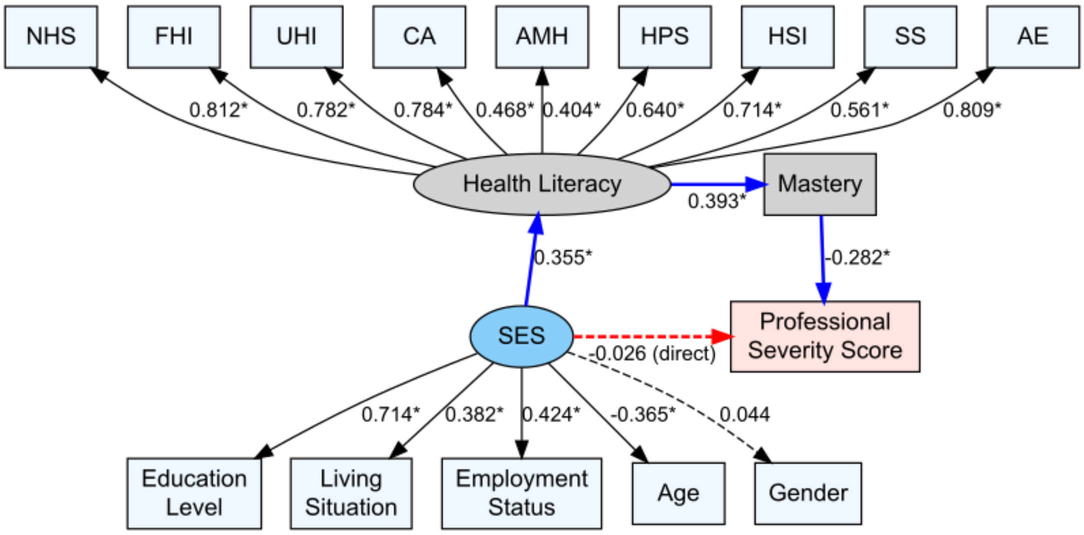


Figure 9: Path diagram for the professional severity score outcome. The blue solid lines represent the indirect effect of SES to Professional severity score through Healthy liteарcy and Mastery. Significant standardised path coefficients are indicated with an asterisk (*) (p <0.05). SES, HL Mastery are scored such that higher values reflect more favorable outcomes that is lower severity scores

$p < 0.001$), establishing health literacy and mastery of health as meaningful mediators in the SES severity relationship. The total effect of SES on professional severity was -0.749, with the sequential pathway accounting for approximately 29.7% of this association ($p = 0.024$). Full structural estimates are reported in Table 16.

4.5.2 Health literacy and mastery mediators: Patient health score outcome

The hypothesized sequential mediation model (Figure 10) demonstrated acceptable fit. The Chi-square test was significant, $\chi^2_{WLSMV}(100) = 646.5873$, $p < 0.001$; CFI $= 0.937$, and TLI $= 0.925$ were within acceptable ranges and RMSEA $= 0.084$ (90% CI: 0.082, 0.094), and SRMR $= 0.082$ were slightly above the conventional thresholds.

Path analysis revealed a sequential pattern linking SES to self reported patient health. Higher socioeconomic status was associated with higher levels of health literacy ($a = 0.356$, $p < 0.001$), and higher levels of health literacy was associated with a greater sense of mastery of health ($f = 0.392$, $p < 0.001$). Consequently, a greater sense of mastery of health was positively associated with self reported patient health ($e = 0.368$, $p < 0.001$). All coefficients are standardized, representing the expected change (in standard deviations) in the outcome for a one standard deviation change in the predictor.

The sequential indirect effect from SES to health through health literacy and mastery was statistically significant ($a \times f \times e = 0.051$, $p < 0.001$). The direct effect of SES on self reported patient health was small and non-significant ($c = 0.027$, $p = 0.597$). The sequential mediation pathway explained 25.5% of the total effect of SES on patient reported health (0.808). While this demonstrates the importance

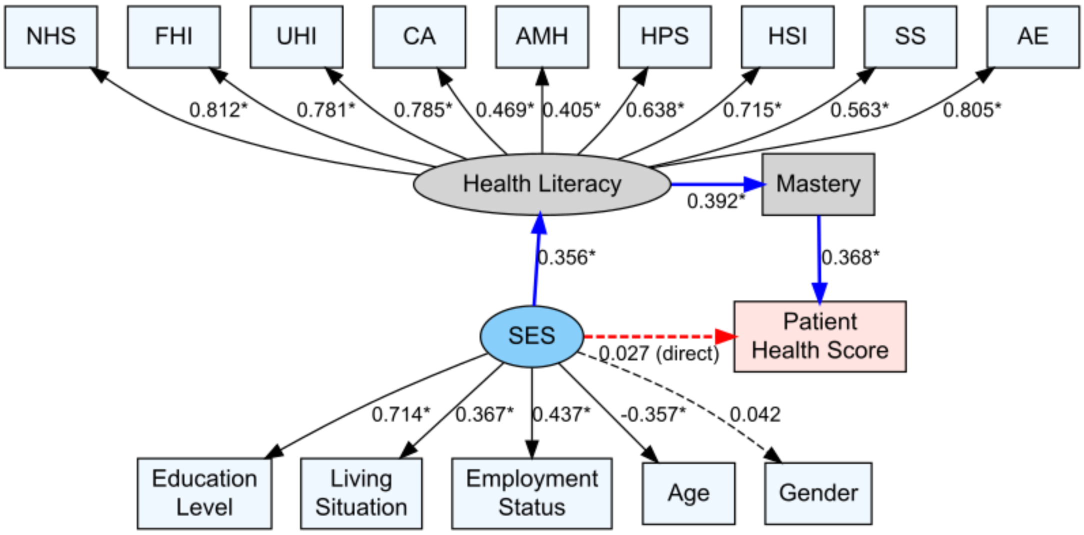


Figure 10: Path diagram for the patient health score outcome with health literacy and mastery as the mediators. The blue solid lines represent the indirect effect of SES to patient health score through health literacy and Mastery. Significant standardised path coefficients are indicated with an asterisk (*) (p <0.05). SES , health literacy and Mastery are scored such that higher values reflect more favorable outcomes that is higher reported scores indicating better health

of health literacy and mastery as mediating mechanisms, it also suggests that additional pathways exists which also contributes to the remaining proportion in this relationship. Full standardized estimates are reported in Table 20.

4.5.3 Health literacy and mastery as mediators: Number of comorbidities

The sequential mediation model fit indices were as follows: CFI = 0.936, TLI = 0.923, with RMSEA = 0.084 (90% CI: 0.082, 0.094), and SRMR = 0.083 being slightly above the conventional thresholds. The $\chi^2_{WLSMV}(100) = 653.902$, $p < 0.001$, was significant

The path diagram Figure 11 shows that higher SES is associated with higher health literacy levels($a = 0.365$, $p < 0.001$), which in turn predicts greater mastery of health ($f = 0.388$, $p < 0.001$). Mastery of health is associated with fewer comorbidities ($e = -0.134$, $p = 0.001$). All effects are standardized, indicating changes in standard deviations of outcomes per standard deviation change in predictors. The indirect effect of SES on comorbidities through health literacy and mastery of health was small but significant (standardized indirect effect $= -0.019$, $p = 0.006$). The direct effect of SES on comorbidities remained significant and larger in magnitude ($c = -0.311$, $p < 0.001$), demonstrating partial mediation. While SES influences comorbidities through these mediators, much of its effect is directly from SES rather than through Health Literacy and mastery of health. The mediation pathway accounted for only 5.0% of the total effect of SES on comorbidities . Full standardized estimates are detailed in Table 18.

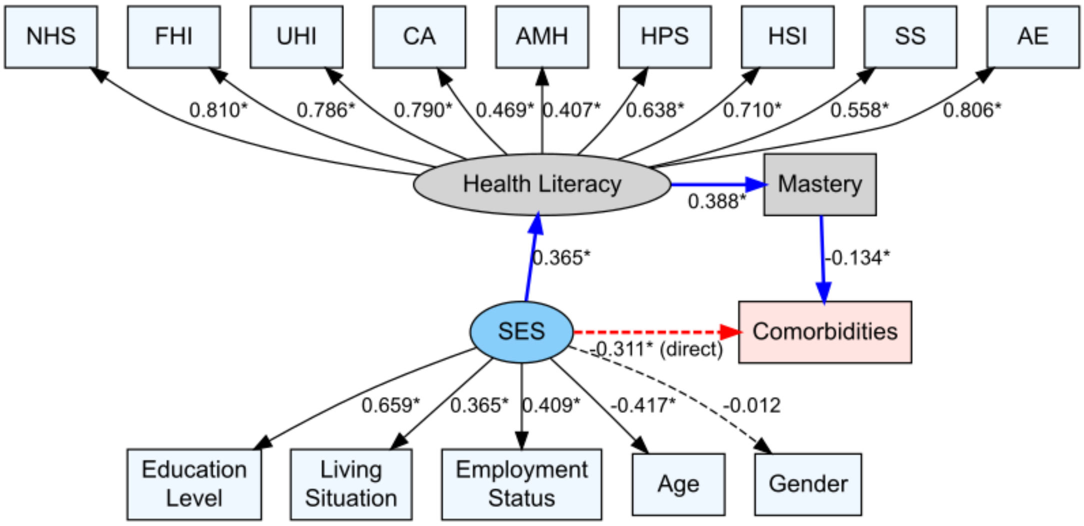


Figure 11: Path diagram for the number of comorbidities outcome with mastery as the mediator. The blue solid lines represent the indirect effect of SES to number of comorbidities through health literacy and Mastery. Significant standardised path coefficients are indicated with an asterisk (*) (p <0.05). SES , health literacy and Mastery are scored such that higher values reflect more favorable outcomes that is fewer number of comorbidities

5 Discussion

The main aim of this study was to investigate whether health literacy (HL) mediates the relationship between socioeconomic status (SES) and three key health outcomes in individuals with inflammatory rheumatic diseases. The outcomes assessed were: patients' global assessment of health, healthcare professionals' global assessment of disease impact on functioning and health, and comorbidities. In addition, the study examined whether mastery serves either as an independent mediator or as a subsequent mediator through which HL further influences these outcomes.

Overall, mastery demonstrated a stronger mediating effect than health literacy in the relationship between socioeconomic status (SES) and health outcomes. However, including both mediators simultaneously did not significantly enhance the model beyond using health literacy alone. For self-reported comorbidities, SES had a direct effect with little or no evidence of mediation, unlike the other health outcomes, where mediation was observed.

Socio-economic status and health literacy as latent variables
To pave the way for the main analysis, confirmatory factor analyses (CFA) were conducted to validate the measurement models for socioeconomic status (SES) and health literacy (HL). All nine proposed HL domains were retained in the final measurement model. The strongest contributors to health literacy were "Navigating the Healthcare System" (0.799) and "Ability to Engage with Providers" (0.789). Information-related domains 'Having Sufficient Information", "Finding Health Information", and "Understanding Health Information" had a contribution of around 0.74. The least contributing domains were "Social Support" (0.544), "Appraisal of Health Information" (0.482), and "Actively

Managing Health" (0.417).

Of the six hypothesized SES indicators, migration background showed non-significant factor loadings and was removed from the model, likely due to limited variability given the high proportion of native patients (83%). Gender contributed significantly to the SES construct, with results indicating that men tended to have higher socioeconomic status than women. However, this contribution was relatively lower (0.266) compared to other indicators (all ≥ 0.4), suggesting a weaker role in defining SES. The final SES model also revealed a negative relationship between age and SES, indicating that older participants tended to have lower socioeconomic status, a pattern consistent with the study's older population (mean age = 61 years).

Health literacy as a mediator between SES and outcome
When health literacy was included as a mediator in the structural equation model examining patients' global assessment of health, the direct path from SES to self-reported health became statistically non-significant. A significant indirect effect emerged through health literacy, accounting for 43% of the total effect. This suggests that health literacy may serve as a pathway through which SES is associated with subjective health perceptions, possibly by supporting patients in understanding health information, navigating the healthcare system, and effectively engaging with providers.

In contrast, the indirect effect of health literacy on the health professional-rated severity score was smaller (28%), indicating a weaker mediating role. Although 'global assessment of health' and 'severity of health' (also rated on a 0–10 scale) are closely related concepts, the divergence in associations likely reflects different perspectives between patients and health professionals and a more limited role of patients' health literacy when health professionals rate their patients' health. It is also plausible that health professionals have limited visibility into their patients' health literacy, which may explain why health literacy had a stronger mediating effect on patient-reported outcomes.

Finally, the analysis of comorbidities found no significant mediating role of health literacy in the relationship between SES and the number of self-reported comorbidity conditions. This suggests that SES influences comorbidity burden directly, independent of health literacy. One plausible explanation is that patients completed the Health Literacy Questionnaire (HLQ) immediately after their rheumatology consultation, which may have shaped their responses to reflect their experiences with rheumatic care specifically. Consequently, their answers may not capture their capacity to manage other chronic conditions, such as diabetes or chronic lung disease. This interpretation is consistent with the view that health literacy is situational and context-dependent, rather than a stable personal trait. In this context, the HLQ likely assessed disease-specific health literacy rather than general health literacy across comorbidities, highlighting the need to consider timing and disease focus when interpreting HL measures in populations with multi-morbidity.

Mastery as a mediator between SES and outcome
Introducing mastery as a mediator did not substantially alter the overall structure of the SES measurement model, which remained largely consistent with the original CFA results. Gender continued to contribute meaningfully to the SES construct, and age maintained its negative contribution. However,

employment status contributed more strongly than in previous models, becoming the most prominent indicator of SES, while the contribution of education decreased. This shift may reflect the relevance of financial security or occupational engagement in fostering a sense of control over life and health, an element central to the mastery construct. The difference with the CFA for SES in the HL model suggests that different psycho-social mediators are associated with distinct facets of socioeconomic status. The variation in indicator contributions across models highlights that the operationalization of SES is not fixed, but may shift depending on the conceptual pathway being modeled. While the overall latent construct remains coherent, the relative weight of individual indicators appears to flexibly align with whether the mediator emphasizes personal control (as in mastery) or cognitive and social access to skills and resources (as in health literacy) .

Mastery mediated 67% of the effect of SES on patient-reported health, which is approximately 20% more than with health literacy as a mediator. This suggests that individuals' perceived control over their health may play a particularly strong role in shaping how they rate their own health. For the professional severity score, 53% of the total effect was mediated through mastery, compared to 28% through health literacy. In contrast to the two other outcomes, mastery played a minimal role in mediating the relationship between SES and comorbidities. Only 16% of the effect was mediated through mastery, while the remaining 84% was a direct effect of SES. This suggests that the number of self-reported comorbidities may be more strongly influenced by the aspects of SES, such as long-term access to healthcare or cumulative exposure to risk factors, rather than by personal control or informational resources.

Sequential mediation

Building on the findings from the individual mediation models, we tested a more complex sequential mediation pathway in which health literacy (HL) and mastery were included as mediators between SES and the outcomes. This model captured both the individual indirect effects of each mediator and the indirect path running sequentially through both HL and mastery. The results from seperate analysis showed that the sequentially mediated effects were present but relatively lower. For example, the proportion of the total effect mediated for the patient global assessment was 26%, compared to 43% for HL alone and 67% for mastery alone. Similarly, for the professional-rated severity score, the sequential mediation accounted for 30% of the effect, compared to 28% for HL alone and 53% for mastery alone. We also note that model fit varied depending on the pathway used. While the SES-to-mastery model did not improve fit relative to SES alone, including HL as a mediator resulted in better model fit and rendered the effect of gender non-significant. This suggests that HL may partially compensate for structural limitations in the SES variable. However, given the complexity and constraints of sequential mediation models, the additional mediation captured here should be interpreted cautiously.

Clinical relevance

From a broader perspective, these findings reaffirm that both socioeconomic status (SES) and health literacy (HL) are multi-dimensional constructs that interact differently across health outcomes. The presence or absence of mediation appeared highly sensitive to the nature of the outcome. For instance, the absence of a mediation effect for comorbidities was somewhat unexpected, particularly in light of syndemic models that emphasize the clustering of health burdens among socioeconomically disadvantaged groups. One plausible explanation is that the Health Literacy Questionnaire (HLQ) was administered shortly after a rheumatology consultation, prompting participants to respond primarily in relation to their arthritis care. This situational focus may have reduced the relevance of comorbid conditions in their responses. Such findings reinforce the view that health literacy is not a fixed personal trait, but a context-dependent resource shaped by individuals' interaction with specific healthcare environments.

Clinically, this distinction has practical implications. Health literacy (HL), a set of personal and social skills, emerges as a more actionable target for intervention compared to constructs like mastery, which are likely more trait-like and less amenable to short-term change. HL can be improved through interventions that support patients in accessing, understanding, and using health information, navigating the healthcare system, and communicating effectively with providers. These are modifiable dimensions that health professionals can directly influence. In contrast, enhancing patients' general sense of mastery or self-efficacy may require more intensive, long-term strategies, and may not be easily modifiable at all. Taken together, our findings highlight the importance of tailoring health interventions to both the specific outcome of interest and the particular domains of HL most relevant in the clinical context.

6 Methodological limitations

Since this study involved secondary data analysis, several statistical challenges arose. The `lavaan` package was used for structural equation modeling (SEM) due to accessibility and familiarity, despite being less advanced than software like Stata or Mplus. One key limitation is that `lavaan` does not account for overdispersion in count data. Overdispersion was confirmed in our count variable number of comorbidities, with a mean of 1.85 and variance of 2.45, indicating greater variability than expected under a Poisson distribution. To address violations of normality and the ordinal nature of many variables, we used the WLSMV estimator, which treats counts as ordinal data and is less sensitive to distributional assumptions, though it does not fully resolve overdispersion. As such, we recommend use of more advanced SEM software capable of handling count data and overdispersion, such as Mplus or Stata,so as to improve modeling of variables like comorbidities. Future research may benefit from exploring Bayesian SEM approaches or other software that better accommodate the data characteristics.[31]

Secondly, the retained measurement model for the SES construct, comprising five indicators, demonstrated mixed fit. Approximate fit indices were reasonable, with a Root Mean Square Error of Approximation (RMSEA) of 0.085 and a Standardized Root Mean Square Residual (SRMR) of 0.070. However, incremental fit indices, including the Comparative Fit Index (CFI) at 0.857 and the Tucker-Lewis Index (TLI) at 0.714, indicated room for improvement, falling below conventional thresholds of 0.90 to 0.95. These results suggest that the selected indicators may not fully capture the complexity of the multidimensional SES construct. Nonetheless, due to the theoretical coherence of the five retained indicators, we deemed the measurement model adequate for subsequent analyses. In addition, given the complex and multidimensional nature of socioeconomic status, future studies should explore alternative or additional indicators to better explain the health outcomes.

Thirdly, although structural equation modeling (SEM) allows us to test complex relationships and potential mediation effects, the cross-sectional design of our study means we cannot confirm cause-and-effect. SEM can demonstrate that variables are related in ways consistent with theory, but it cannot prove the direction of causality since all data were collected at a single time point. Therefore, the mediation results from our analyses should be interpreted as associations that align with the proposed model, rather than definitive evidence of causal pathways. To establish causality, future studies using longitudinal or experimental designs are needed.

Finally, we recommend conducting subgroup analyses to examine whether the mediation relationships differ across the three iRMDs (rheumatoid arthritis (RA), spondyloarthritis (SpA), and gout). Subsequent descriptive analyses highlighted distinct demographic differences among the diseases: RA patients were generally older, predominantly female, and had higher rates of work disabled and comorbidities; SpA patients were younger, more often employed, and showed a more balanced gender distribution; gout patients were mostly older males with the highest comorbidity burden. Investigating these differences would be important to determine if the pathways from SES through health literacy and mastery to health outcomes would be consistent across disease groups.

7 Ethical Considerations, Societal Relevance, and Stakeholder Awareness

This project builds upon a strong ethical foundation established by the original multi-center study, which was approved by the Medical Ethics Review Committee at Maastricht UMC+ (reference: 2018-0327), along with additional approvals from the ethics committees at each participating hospital (Maastricht: 18-4-037; Rotterdam: L2018057; Enschede: KH18-23). Importantly, a patient research partner (MdW) was involved throughout the original research process, ensuring that patient perspectives informed the study from the outset. For this secondary analysis, only fully anonymized data were used. All participants had provided written informed consent, and no identifiable personal information was accessed, meaning that no further ethical approval was required.

The ethical motivation for this research arises from the persistent challenge of health disparities rooted in socioeconomic status (SES). This study advances our understanding by focusing on health literacy a person's ability to find, understand, and use health information as a key factor linking SES to health outcomes. Importantly, health literacy is not simply a marker of existing social inequalities; it is a modifiable skill that can be enhanced through targeted education and support. This distinction matters because it opens a practical pathway for intervention. By improving health literacy, healthcare providers and policymakers can empower individuals, enabling better navigation of healthcare systems, more effective communication with professionals, and ultimately, improved health outcomes. Such efforts have the potential to reduce health disparities more effectively than approaches focused solely on socioeconomic factors, which are often less amenable to change. Therefore, prioritizing health literacy interventions offers a promising strategy to make healthcare more equitable and responsive to the needs of socioeconomically disadvantaged populations.

In addition to health literacy, this study also examined mastery, or individuals' perceived control over their life circumstances, as a secondary mediator. Including mastery acknowledges that managing one's health depends also on emotional and psychological resources. Together, these two factors health literacy and mastery offer a more complete view of how SES influences health, helping to identify where support might be most impactful.

The societal relevance of these findings lies in their potential to inform more equitable healthcare strategies. Rather than viewing low SES as an unchangeable factor, this study paves way to intervene by improving health literacy and strengthening a sense of personal control. This shift in perspective encourages us to see individuals not as passive recipients of care, but as capable people who may simply need better tools and support to manage their health effectively.

These insights are relevant to multiple stakeholders. For healthcare providers, the findings highlight the importance of tailoring communication and care strategies to patients' varying needs, supporting the implementation of health literacy responsive care where healthcare systems actively support persons in vulnerable positions and/or with limitations in health literacy to make optimal health decisions. For public health professionals and policymakers, the study supports the development of integrated approaches that build both health knowledge and personal empowerment. While this secondary analysis did not involve direct community stakeholder engagement, the interpretation of the results was carried out in close collaboration with the original study investigators. This ensured that the work remained aligned with the patient-centered goals of the broader research.

References

[1] van der Heijde, et.al. (2018). Common language description of the term rheumatic and musculoskeletal diseases (RMDs) for use in communication with the lay public, healthcare providers and other stakeholders endorsed by the European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR). Annals of the Rheumatic Diseases, 77(6), 829–832. https://doi.org/10.1136/annrheumdis-2018-213328

[2] Arthritis Foundation. The Spondyloarthritis family? https://www.arthritis.org/diseases/spondyloarthritis

[3] World Health Organization. (2023). Rheumatoid arthritis. https://www.who.int/news-room/fact-sheets/detail/rheumatoid-arthritis

[4] Centers for Disease Control and Prevention. (2023). *Gout*. Retrieved May 4, 2025, from https://www.cdc.gov/arthritis/gout/index.html

[5] van der Heide I, Wang J, Droomers M, Spreeuwenberg P, Rademakers J, Uiters E. The relationship between health, education, and health literacy: results from the Dutch Adult Literacy and Life Skills Survey. J Health Commun. 2013;18 Suppl 1(Suppl 1):172-84. doi: 10.1080/10810730.2013.825668. PMID: 24093354; PMCID: PMC3814618.

[6] Svendsen, M. T., et al. (2020). Associations of health literacy with socioeconomic position, health risk behavior, and health status: a large national population-based survey among Danish adults. BMC Public Health, *20*(1), 565. https://doi.org/10.1186/s12889-020-08498-8

[7] Bakker, M. M.,et al (2021). Addressing health literacy needs in rheumatology: Which patient health literacy profiles need the attention of health professionals? Arthritis Care and Research, *73*(1), 100–109. https://doi.org/10.1002/acr.24480

[8] Berete F, et al. (2024) Does health literacy mediate the relationship between socioeconomic status and health related outcomes in the Belgian adult population? BMC Public Health. 2024 Apr 27;24(1):1182. doi: 10.1186/s12889-024-18676-7. PMID: 38678179; PMCID: PMC11055376.

[9] Stormacq C, Van den Broucke S, Wosinski J. Does health literacy mediate the relationship between socioeconomic status and health disparities? Integrative review. Health Promot Int. 2019 Oct 1;34(5):e1-e17. doi: 10.1093/heapro/day062. PMID: 30107564.

[10] Paasche-Orlow, Michael and Wolf, Michael. (2007). The Causal Pathways Linking Health Literacy to Health Outcomes. American journal of health behavior. 31 Suppl 1. S19-26. 10.5993/AJHB.31.s1.4.

[11] Pearlin, L.I., and Schooler, C. (1978). The structure of coping. Journal of Health and Social Behavior, 19, 2-21 https://www.jstor.org/stable/2136319

[12] Mark M Bakker, et.al, Exploring discordance between Health Literacy Questionnaire scores of people with RMDs and assessment by treating health professionals, Rheumatology, Volume 62, Issue 1, January 2023, Pages 52–64, https://doi.org/10.1093/rheumatology/keac248

[13] International Standard Classification of Education Fields of Training and Education 2013 (ISCEDF 2013) by UNESCO: https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/onderwijs-en-beroepen/standaard-onderwijsindeling–soi–

[14] Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health. 2013;13:658. doi:10.1186/1471-2458-13-658

[15] Hair, J.F. et.al (2021). An Introduction to Structural Equation Modeling. In: Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R. Classroom Companion: Business. Springer, Cham. https://doi.org/10.1007/978-3-030-80519-7_1

[16] Hair, J. F., et al (2010). Multivariate Data Analysis (7th ed.). Pearson.

[17] Beran, T.N., Violato, C. Structural equation modeling in medical research: a primer. BMC Res Notes 3, 267 (2010). https://doi.org/10.1186/1756-0500-3-267

[18] Kline, R. B. (2016). Principles and Practice of Structural Equation Modeling (4th ed.). The Guilford Press (Sample size requirements pg 15-16) .

[19] Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1–36. https://doi.org/10.18637/jss.v048.i02

[20] Braveman, P. et.al (2005). Socioeconomic disparities in health in the United States: What the patterns tell us. American Journal of Public Health, 95(4), 716–724.

[21] Baron, R. M., and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173–1182.

[22] Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford Press.

[23] Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

[24] Beaujean, A. A. (2014). Latent Variable Modeling Using R: A Step-by-Step Guide. Routledge.

[25] Flora DB, Curran PJ. (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods. 2004 Dec;9(4):466-91. doi: 10.1037/1082-989X.9.4.466. PMID: 15598100; PMCID: PMC3153362.

[26] MacKinnon, D. (2012). Introduction to statistical mediation analysis. Routledge.

[27] MacKinnon DP. Contrasts in multiple mediator models. In: Rose JS, Chassin L, Presson CC, Sherman SJ, editors. Multivariate Applications in Substance Use Research: New Methods for New Questions. Mahwah, NJ: Erlbaum; 2000. pp. 141–60. [Google Scholar]

[28] Berete, F., Gisle, L., Demarest, S. et al. Does health literacy mediate the relationship between socioeconomic status and health related outcomes in the Belgian adult population?. BMC Public Health 24, 1182 (2024). https://doi.org/10.1186/s12889-024-18676-7

[29] Bauer, D. J., Preacher, K. J., Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. Psychological Methods, 11, 142-163

[30] Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 1). J Epidemiol Community Health. 2006 Jan;60(1):7-12. doi: 10.1136/jech.2004.023531. PMID: 16361448; PMCID: PMC2465546.

[31] Muthén B, Asparouhov T. Bayesian structural equation modeling: a more flexible representation of substantive theory. Psychol Methods. 2012 Sep;17(3):313-35. doi: 10.1037/a0026802. PMID: 22962886.

[32] Anita Peerson, Margo Saunders, Health literacy revisited: what do we mean and why does it matter?, Health Promotion International, Volume 24, Issue 3, September 2009, Pages 285–296, https://doi.org/10.1093/heapro/dap014

[33] Lastrucci, V., Lorini, C., Caini, S., Bonaccorsi, G. (2019). Health literacy as a mediator of the relationship between socioeconomic status and health: A cross-sectional study in a population-based sample in Florence. PLOS ONE, 14(12), e0227007. https://doi.org/10.1371/journal.pone.0227007

8 Supplementary material

8.1 Education Levels as per Dutch System

Table 11: Dutch Educational Levels

Low (Laag)	Medium (Middelbaar)	High (Hoog)
1 Basisonderwijs, vmbo, mbo1	2 Havo, vwo, mbo2-4	3 Hbo, wo
11 Basisonderwijs	21 Havo, vwo, mbo	31 Hbo-, wo-bachelor
111 Basisonderwijs	211 Mbo2 en mbo3	311 Hbo-, wo-bachelor
1111 Basisonderwijs gr1-2	2111 Mbo2	3111 Hbo-associate degree
1112 Basisonderwijs gr3-8	2112 Mbo3	3112 Hbo-bachelor
12 Vmbo, havo-, vwo-onderbouw, mbo1	212 Mbo4	3113 Wo-bachelor
121 Vmbo-b/k, mbo1	2121 Mbo4	32 Hbo-, wo-master, doctor
1211 Praktijkonderwijs	213 Havo, vwo	321 Hbo-, wo-master, doctor
1212 Vmbo-b/k	2131 Havo-bovenbouw	3211 Hbo-master
1213 Mbo1	2132 Vwo-bovenbouw	3212 Wo-master
122 Vmbo-g/t, havo-, vwo-onderbouw		3213 Doctor
1221 Vmbo-g/t		
1222 Havo-, vwo-onderbouw		

8.2 Measurement Model for Health Literacy (HL)

The latent variable Health Literacy (HL) was specified using observed indicators from the nine domain scores of the Health Literacy Questionnaire (HLQ). The factor loading for the reference indicator was fixed to 1.0 to identify the model. The measurement model is expressed as follows:

$$\text{Feeling understood and supported by healthcare providers (HPS)} = 1.0 \times \xi_{HL} + \delta_1 \quad \text{(reference indicator)} \tag{12}$$

$$\text{Having sufficient information to manage health (HSI)} = \lambda_2 \xi_{HL} + \delta_2 \tag{13}$$

$$\text{Actively managing health (AMH)} = \lambda_3 \xi_{HL} + \delta_3 \tag{14}$$

$$\text{Social support for health (SS)} = \lambda_4 \xi_{HL} + \delta_4 \tag{15}$$

$$\text{Critical appraisal of health information (CA)} = \lambda_5 \xi_{HL} + \delta_5 \tag{16}$$

$$\text{Ability to actively engage with healthcare providers (AE)} = \lambda_6 \xi_{HL} + \delta_6 \tag{17}$$

$$\text{Navigating the healthcare system (NHS)} = \lambda_7 \xi_{HL} + \delta_7 \tag{18}$$

$$\text{Ability to find good health information (FHI)} = \lambda_8 \xi_{HL} + \delta_8 \tag{19}$$

Each domain score represents the mean score within its HLQ domain. The standardized factor loadings (λ_i) indicate the strength of association between each domain and the latent construct ξ_{HL}, while δ_i represents the measurement error for each indicator.

8.3 Mediation analysis steps

8.3.1 Mediation Analysis: One mediator (Mastery)

The mediation analysis steps applied to the one-mediator models (e.g., with mastery or health literacy as mediators) followed the same procedure described for the primary analysis in the main manuscript. These included assessing individual path significance prior to conducting formal mediation, followed by estimation of direct, indirect, and total effects using the WLSMV estimator in structural equation modeling. While the analytical approach remained consistent, the path diagrams for each model differ and are presented here to illustrate the distinct mediation structures.

$$Y = b_{01} + cX + e_5 \tag{20}$$

$$M_2 = b_{02} + dX + e_6 \tag{21}$$

$$Y = b_{03} + eM_2 + e_7 \tag{22}$$

$$Y = b_{04} + c'X + eM_2 + e_8 \tag{23}$$

8.3.2 Mediation Analysis: Health Literacy and Mastery in Sequence

To evaluate a sequential mediation model, in which SES influences health literacy (M_1), which in turn affects mastery (M_2), and subsequently the outcome (Y), the following set of equations was specified:

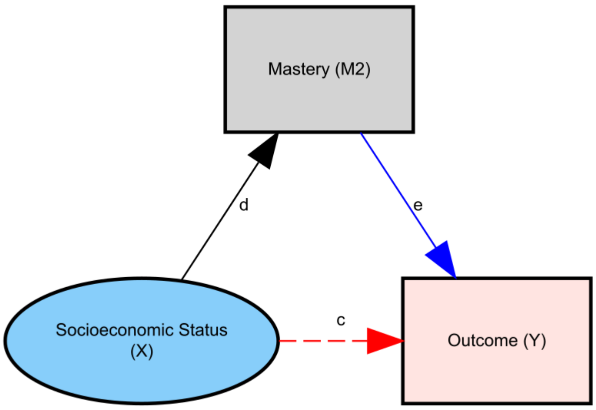
$$M_2 = b_{08} + fM_1 + e_9 \tag{24}$$

$$Y = b_{09} + c'X + fM_1 + eM_2 + e_{10} \tag{25}$$

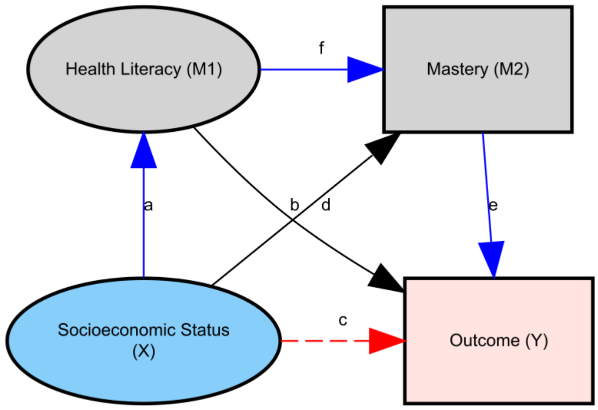
This model builds on earlier equations for the mediation path SES $\rightarrow$ Health Literacy (M_1), and includes:

- Equation 23: the effect of health literacy on mastery.
- Equation 24: the combined effects of SES, health literacy, and mastery on the outcome.

For this sequential mediation to hold, the following paths should be significant: SES $\rightarrow M_1$ (Equation 9), $M_1 \rightarrow M_2$ (Equation 23), and $M_2 \rightarrow Y$ (Equation 24). The overall indirect effect is the product $a \times f \times e$ (product of the sequential model). The total effect is $c = c' + ab + de + afe$ where (ab is the indirect effect through health literacy); (de is the indirect effect through mastery)



(a) Model with health literacy as sole mediator



(b) Model with Health literacy and mastery of health as sole mediator

Figure 12: Path diagrams illustrating two mediation models tested: (a) mastery of health alone, and (b) both health literacy and mastery of health sequentially mediating the relationship between socioeconomic status and health outcomes among individuals with inflammatory arthritis.

8.4 Measurement model assumptions

8.4.1 Assessing the multivariate assumption for the health literacy nine domains

Prior to conducting the confirmatory factor analysis (CFA) for health literacy, the distributional properties of the observed domain scores were examined to assess multivariate normality. Each domain was constructed by averaging responses to four or five Likert-type items, producing continuous-looking indicators hence it was important to test whether the assumption of multivariate normality held. Mardia's test indicated substantial violations of this assumption (skewness = 1512.14, $p < 0.001$; kurtosis = 45.04, $p < 0.001$), and these results were also consistent with what was observed by visual inspection of histograms Figure 13 in the Supplementary material. These findings confirmed that the observed indicators did not follow a multivariate normal distribution, justifying the use of the WLSMV estimator, which is robust to such violations and well-suited for analyzing models with ordinal or non-normally distributed data.

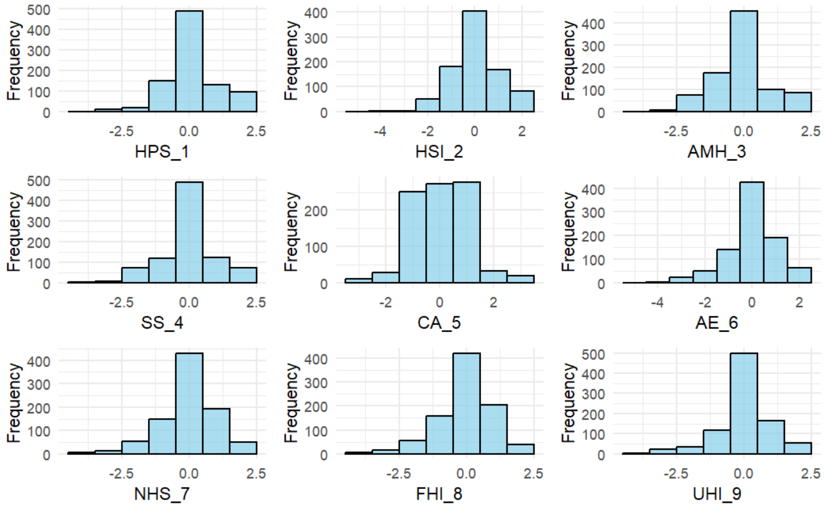


Figure 13: Histograms of the score distributions across the different health literacy domains. The x-axis displays z-standardised domain scores (mean = 0, SD = 1), allowing comparison across domains with different numbers of items by placing them on a common scale.

8.5 Structural Model tables (Mastery as the only mediator)

The regression models described in this sub-section used the same tables that were applied in the sequential individual path analysis. Therefore, they are presented in the section on sequential mediation analysis, where the detailed results for those tables are also discussed. However, it is important to note that in all the path analyses conducted, the significance of individual paths required to perform mediation analysis were met.

8.5.1 SES-Mastery-Professional Severity Score

Table 12: Structural Equation Modeling with mediation analysis results: Professional severity score

Latent Variable: SES	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Employment Status	1.000	–	–	0.605
Education Level	0.807	0.118	< 0.001	0.488
Living Situation	0.649	0.128	< 0.001	0.393
Age	-7.710	1.294	< 0.001	-0.335
Gender	0.465	0.107	< 0.001	0.281
Regressions	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Mastery $\sim$ SES (d)	2.417	0.378	< 0.001	0.425
Professional severity score$\sim$ SES (c')	-0.357	0.227	0.116	-0.090
Professional severity score $\sim$ Mastery (e)	-0.169	0.029	< 0.001	-0.243
Defined Effects	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Indirect Effect (d $\times$ e)	-0.409	0.088	< 0.001	-0.103
Total Effect	-0.766	0.218	< 0.001	-0.194
Proportion Mediated	0.534	–	–	0.534

8.5.2 SES-Mastery-Patient health score

Table 13: Structural Equation Modelling with mediation analysis results: Patient health score

Latent Variable: SES	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Employment Status	1.000	–	–	0.630
Education Level	0.785	0.117	< 0.001	0.494
Living Situation	0.567	0.122	< 0.001	0.357
Age	-7.143	1.241	< 0.001	-0.323
Gender	0.418	0.103	< 0.001	0.263
Regressions	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Mastery $\sim$ SES (d)	2.335	0.372	< 0.001	0.428
Patient health score $\sim$ SES (c')	0.223	0.141	0.113	0.080
Patient health score $\sim$ Mastery (e)	0.192	0.018	< 0.001	0.375
Defined Effects	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Indirect Effect (d $\times$ e)	0.448	0.076	< 0.001	0.161
Total Effect	0.672	0.153	< 0.001	0.241
Proportion Mediated	0.667	–	–	0.667

8.5.3 SES-Mastery-Comorbidities

Table 14: Structural Equation Modelling with mediation analysis results: Number of comorbidities

Latent Variable: SES	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Employment Status	1.000	–	–	0.506
Education Level	0.977	0.142	< 0.001	0.494
Living Situation	0.741	0.146	< 0.001	0.375
Age	-12.745	1.906	< 0.001	-0.463
Gender	0.301	0.119	0.011	0.152
Regressions	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Mastery $\sim$ SES (d)	2.922	0.451	< 0.001	0.430
Comorbidities $\sim$ SES (c')	-0.969	0.199	< 0.001	-0.316
Comorbidities $\sim$ Mastery (e)	-0.062	0.019	0.001	-0.138
Defined Effects	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Indirect Effect (d $\times$ e)	-0.182	0.057	0.001	-0.059
Total Effect	-1.151	0.186	< 0.001	-0.375
Proportion Mediated	0.158	–	–	0.158

8.6 Individual paths and Structural Model tables (Both HL and Mastery as mediators)

8.6.1 SES-HL-Mastery-Professional severity score outcome

Table 15: Individual path analysis for the outcome professional severity score

Regression Path	Estimate	SE	P-value	Std. All
SES → Health Literacy (a)	0.676	0.137	< 0.001	0.357
Health Literacy →Mastery of health(f)	2.681	0.258	< 0.001	0.494
SES →Professional Severity Score(c)	-0.822	0.241	< 0.001	-0.190
Mastery of health → Professional Severity Score (e)	-0.196	0.024	< 0.001	-0.281
SES → Mastery of health (d)	2.429	0.378	< 0.001	0.404

For the Professional Severity Score outcome, all individual paths SES to Health Literacy (a), Health Literacy to Mastery of Health (f), SES directly to Professional Severity Score (c), and Mastery of Health to Professional Severity Score (e) were statistically significant ($p < 0.001$). Additionally, the direct path from SES to Mastery of Health (d) was also significant ($p < 0.001$). These findings, based on data from 778 participants, fulfill the conditions necessary for conducting a mediation analysis. As a result, both the sequential mediation model and the model with Mastery of Health as a standalone mediator were further examined through formal mediation analysis.

Table 16: Structural Equation Modelling with Sequential Mediation Results: Professional severity score

Latent Variable: SES	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Employment Status	1.000	–	–	0.424
Education Level	1.685	0.300	< 0.001	0.714
Living Situation	0.902	0.223	< 0.001	0.382
Age	-12.009	2.537	< 0.001	-0.365
Gender	0.104	0.167	0.533	0.044
Latent Variable: Health Literacy	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Healthcare Providers Support (HPS_1)	1.000	–	–	0.640
Having Sufficient Information (HSI_2)	1.122	0.050	< 0.001	0.714
Actively Managing Health (AMH_3)	0.626	0.050	< 0.001	0.404
Social Support (SS_4)	0.858	0.045	< 0.001	0.561
Appraisal of Health (CA_5)	0.717	0.052	< 0.001	0.468
Ability to Engage (AE_6)	1.245	0.063	< 0.001	0.809
Navigating Health System (NHS_7)	1.265	0.067	< 0.001	0.812
Finding Health Information (FHI_8)	1.221	0.070	< 0.001	0.782
Understanding Health Information (UHI_9)	1.207	0.068	< 0.001	0.784
Regressions	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Health Literacy ∼ SES (a)	0.548	0.107	< 0.001	0.355
Mastery ∼ Health Literacy (f)	2.065	0.211	< 0.001	0.393
Mastery ∼ SES (d)	2.181	0.479	< 0.001	0.269
Professional Severity Score ∼ SES (c')	-0.145	0.313	0.644	-0.026
Professional Severity score ∼ Mastery(e)	-0.197	0.029	< 0.001	-0.282
Professional Severity score ∼ Health Literacy(b)	0.086	0.153	0.576	0.023
Defined Effects	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Indirect Effect (a × f × e)	-0.223	0.054	< 0.001	-0.039
Total Effect ($c' + ab + de + afe$)	-0.749	0.292	0.010	-0.133
Proportion Mediated (a × f × e / Total)	0.297	–	–	0.297

8.6.2 SES-HL-Mastery-Comorbidities outcome

Table 17: Individual path analysis for the outcome self reported comorbidities

Regression Path	Estimate	SE	P-value	Std. All
Health Literacy ∼ SES (a)	0.676	0.137	< 0.001	0.357
Mastery ∼ Health Literacy (f)	2.690	0.202	< 0.001	0.502
Comorbidities ∼ SES (c)	-2.062	0.366	< 0.001	-0.450
Comorbidities ∼ Mastery (e)	-0.125	0.015	< 0.001	-0.271
Mastery ∼ SES (d)	2.429	0.378	< 0.001	0.404

For the comorbidities outcome, all individual paths were statistically significant (p ¡ 0.001) that is from SES to health literacy (a), from health literacy to mastery of health (f), from SES directly to comorbidities (c), and from mastery of health to comorbidities (e). The direct path from SES to mastery of health (d) was also significant ($p < 0.001$). These findings are based on data from 894 participants. Given that all relevant paths met the criteria for mediation, both the sequential mediation model and the model with mastery of health as a stand alone mediator were further examined through formal mediation analysis.

Table 18: Structural Equation Modeling with Sequential Mediation Results: Comorbidities

Latent Variable: SES	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Employment Status	1.000	–	–	0.409
Education Level	1.612	0.274	< 0.001	0.659
Living Situation	0.893	0.219	< 0.001	0.365
Age	-14.207	2.797	< 0.001	-0.417
Gender	-0.029	0.168	0.862	-0.012
Latent Variable: Health Literacy	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Healthcare Providers Support (HPS)	1.000	–	–	0.638
Having Sufficient Information (HSI)	1.119	0.050	< 0.001	0.710
Actively Managing Health (AMH)	0.632	0.051	< 0.001	0.407
Social Support (SS)	0.856	0.045	< 0.001	0.558
Appraisal of Health (CA)	0.721	0.053	< 0.001	0.469
Ability to Engage (AE)	1.245	0.064	< 0.001	0.806
Navigating Health System (NHS)	1.266	0.068	< 0.001	0.810
Finding Health Information (FHI)	1.231	0.071	< 0.001	0.786
Understanding Health Information (UHI)	1.221	0.069	< 0.001	0.790
Regressions	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Health Literacy ∼ SES (a)	0.582	0.112	< 0.001	0.365
Mastery ∼ Health Literacy (f)	2.045	0.214	< 0.001	0.388
Mastery ∼ SES (d)	2.152	0.491	< 0.001	0.253
Comorbidities ∼ SES (c')	-1.182	0.268	< 0.001	-0.311
Comorbidities ∼ Mastery (e)	-0.060	0.019	0.001	-0.134
Comorbidities∼ Health Literacy (b)	0.020	0.094	0.836	0.008
Defined Effects	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Indirect Effect (a × f × e)	-0.072	0.026	0.006	-0.019
Total Effect ($c' + ab + de + afe$)	-1.424	0.256	< 0.001	-0.375
Proportion Mediated (a × f × e / Total)	0.050	–	–	0.050

8.6.3 SES-HL-Mastery-Patient health score outcome

Table 19: Individual path analysis for the outcome self reported patient health score

Regression Path	Estimate	SE	P-value	Std. All
SES → Health Literacy (a)	0.676	0.137	< 0.001	0.357
Health Literacy → Mastery of health(f)	2.690	0.202	< 0.001	0.502
SES→ Patient health score (c)	0.622	0.148	< 0.001	0.228
Mastery of health → Patient health score(e)	0.211	0.016	< 0.001	0.409
SES→ Mastery of health(d)	2.429	0.378	< 0.001	0.404

For the Patient Health Score outcome, all individual paths SES to Health Literacy (a), Health Literacy to Mastery of Health (f), SES directly to Patient Health Score (c), and Mastery of Health to Patient Health Score (e) were statistically significant ($p < 0.001$). Additionally, the direct path from SES to Mastery of Health (d) was also significant ($p < 0.001$). These results are based on data from 894 participants as there was one missing data from the mastery of health score. The results satisfy the criteria for conducting a mediation analysis. As such, mediation analysis were conducted for both the sequential mediation model and the model with mastery of health as a standalone mediator.

Table 20: Structural Equation Modelling with Sequential Mediation Results: Patient reported health

Latent Variable: SES	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Employment Status	1.000	–	–	0.437
Education Level	1.635	0.289	< 0.001	0.714
Living Situation	0.841	0.213	< 0.001	0.367
Age	-11.369	2.424	< 0.001	-0.357
Gender	0.095	0.162	0.556	0.042
Latent Variable: Health Literacy	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Healthcare Providers Support (HPS_1)	1.000	–	–	0.638
Having Sufficient Information (HSI_2)	1.126	0.051	< 0.001	0.715
Actively Managing Health (AMH_3)	0.628	0.051	< 0.001	0.405
Social Support (SS_4)	0.863	0.045	< 0.001	0.563
Appraisal of Health (CA_5)	0.720	0.053	< 0.001	0.469
Ability to Engage (AE_6)	1.243	0.064	< 0.001	0.805
Navigating Health System (NHS_7)	1.267	0.068	< 0.001	0.812
Finding Health Information (FHI_8)	1.224	0.071	< 0.001	0.781
Understanding Health Information (UHI_9)	1.212	0.069	< 0.001	0.785
Regressions	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Health Literacy ∼ SES (a)	0.532	0.104	< 0.001	0.356
Mastery∼ Health Literacy (f)	2.062	0.211	< 0.001	0.392
Mastery ∼SES (d)	2.010	0.447	< 0.001	0.254
Patient Health ∼ SES (c')	0.109	0.206	0.597	0.027
Patient Health ∼ Mastery(e)	0.188	0.019	< 0.001	0.368
Patient Health ∼ Health literacy (b)	0.145	0.102	0.157	0.052
Defined Effects	**Estimate**	**Std. Error**	**P-value**	**Std. All**
Indirect Effect (a × f × e)	0.206	0.045	< 0.001	0.051
Total Effect ($c' + ab + de + afe$)	0.808	0.215	< 0.001	0.201
Proportion Mediated (a × f × e / Total)	0.255	–	–	0.255

9 Codes

````
```{r}
library(haven)
library(dplyr)
library(lavaan)
library(ggplot2)
library(DiagrammeR)
library(GGally)
newdata <- read_dta("C:/Users/ayiem/OneDrive/Desktop/Master Thesis/Dataset Ayleen/UHasselt_dataset.dta")
head(newdata)
```

#####Ordering variables for SEM
```{r}

newdata4$PT_Gender   <- as.ordered(newdata4$PT_Gender)
newdata4$PT_EDU_CAT    <- as.ordered(newdata4$PT_EDU_CAT)
newdata4$COB_class   <- as.ordered(newdata4$COB_class)
newdata4$Living_Situation    <- as.ordered(newdata4$Living_Situation)
newdata4$Employment_Status   <- as.ordered(newdata4$Employment_Status)

```

#### Final CFA Model:SES
```{r}
ses_model <- '
  SES =~ PT_EDU_CAT + Employment_Status + Living_Situation + PT_Age+ PT_Gender
'

# Fit the model
fit1 <- cfa(ses_model, data = newdata4,
            estimator = "WLSMV",
            ordered = c("Employment_Status", "PT_EDU_CAT", "Living_Situation", "PT_Gender"))

summary(fit1, standardized = TRUE, fit.measures = TRUE)

```

````

```
##### Standardisation
#### Histograms
```{r}
indicators <- c("HPS_1", "HSI_2", "AMH_3", "SS_4", "CA_5", "AE_6",
   "NHS_7", "FHI_8", "UHI_9")


plot_list <- list()


for (indicator in indicators) {
  p <- ggplot(newdata4, aes(x = .data[[indicator]])) +
    geom_histogram(binwidth = 1, fill = "skyblue", color = "black",
       alpha = 0.7) +
    labs(x = indicator,
         y = "Frequency") +
    theme_minimal() +
    theme(plot.title = element_blank())
  plot_list[[indicator]] <- p
}

grid.arrange(grobs = plot_list, ncol = 3)


```

```{r}
mvn_result <- mvn(newdata4[, indicators],
                  mvnTest = "mardia",
                  univariateTest = "SW",
                  univariatePlot = "qqplot",
                  multivariatePlot = "qq",
                  showOutliers = TRUE)

print(mvn_result$multivariateNormality )
```


```{r}

HLmodel <- '
```
```

```
  Healthliteracy =~ HPS_1 + HSI_2 + AMH_3 + SS_4 + CA_5 + AE_6 +
     NHS_7 + FHI_8 + UHI_9
'
fitHL <- cfa(HLmodel, data = newdata3,
              estimator = "WLSMV")

summary(fitHL, standardized = TRUE, fit.measures = TRUE)
```


#####mediation model for patient reported health
```{r}

mediation_model1a <- '
  # Latent variables
  SES =~ Employment_Status + PT_EDU_CAT + Living_Situation + PT_Age
     + PT_Gender
  Healthliteracy =~ HPS_1 + HSI_2 + AMH_3 + SS_4 + CA_5 + AE_6 +
     NHS_7 + FHI_8 + UHI_9

  # Mediation paths
  Healthliteracy ~ a*SES
PT_Health ~ b*Healthliteracy + c*SES

  # Indirect, total, proportion mediated
  indirect := a * b
  total := c + (a * b)
  prop := indirect / total
'


fit_mediation1a <- sem(mediation_model1a,
                    data = newdata4,
                    estimator = "WLSMV",
                    ordered = c( "PT_EDU_CAT",
                                "Living_Situation",
                                   "Employment_Status",  "COB_class"
                                    ))


summary(fit_mediation1a, standardized = TRUE, fit.measures = TRUE)


```

```

```‘


```{r}

library(DiagrammeR)
library(lavaan)

create_sem_diagram <- function() {
  dot_code <- '
digraph SEM {
    rankdir=BTLR;
    bgcolor="white";
    splines=true;
    compound=true;

    // Default styles
    node [shape=box, style="filled", color="black",
       fillcolor="aliceblue", fontname="Arial", fontsize=14];
    edge [color="black", fontname="Arial", fontsize=12];

    // Latent variables
    SES [shape=ellipse, fillcolor="lightskyblue", label="SES"];
    HL  [shape=ellipse, fillcolor="lightgray", label="Health
       Literacy"];

    // SES indicators
    Education  [label="Education\\nLevel"];
    LivingSit  [label="Living\\nSituation"];
    Employment [label="Employment\\nStatus"];
    Age        [label="Age"];
    Gender     [label="Gender"];

    // HL indicators (now light green)
    NHS [label="NHS", fillcolor="palegreen"];
    FHI [label="FHI", fillcolor="palegreen"];
    UHI [label="UHI", fillcolor="palegreen"];
    CA  [label="CA",  fillcolor="palegreen"];
    AMH [label="AMH", fillcolor="palegreen"];
    HPS [label="HPS", fillcolor="palegreen"];
    HSI [label="HSI", fillcolor="palegreen"];
    SS  [label="SS",  fillcolor="palegreen"];
```

```
    AE  [label="AE",  fillcolor="palegreen"];

    // Outcome
    PT_Health [shape=box, fillcolor="mistyrose", label="Patient
       Health\\nScore"];

    // Rank alignment
    { rank=same; SES; HL; }
    { rank=max; NHS; FHI; UHI; CA; AMH; HPS; HSI; SS; AE; }
    { rank=min; PT_Health; }

    // SES measurement model
    SES -> Employment [label="0.379*"];
    SES -> Education  [label="0.679*"];
    SES -> LivingSit  [label="0.422*"];
    SES -> Age        [label="-0.353*"];
    SES -> Gender     [label="0.006" , color="black",
       style="dashed"];

    // HL measurement model
    HL -> HPS [label="0.645*"];
    HL -> HSI [label="0.713*"];
    HL -> AMH [label="0.418*"];
    HL -> SS  [label="0.545*"];
    HL -> CA  [label="0.485*"];
    HL -> AE  [label="0.796*"];
    HL -> NHS [label="0.802*"];
    HL -> FHI [label="0.776*"];
    HL -> UHI [label="0.769*"];

    // Structural model
    SES -> HL         [label="0.354*", color="blue", penwidth=2.0];
    HL  -> PT_Health  [label="0.210*", color="blue", penwidth=2.0];
    SES -> PT_Health  [label="0.097 (direct)", color="red",
       style="dashed", penwidth=2.0];
  }
  '
  gtt <- grViz(dot_code)

  DiagrammeRsvg::export_svg(gtt) %>%
    charToRaw() %>%
    rsvg::rsvg_png("sem_diagramtt.png", width = 800)

```

```
  return(gtt)
}

# Run it
create_sem_diagram()


```


```{r}
create_sem_diagram <- function() {
  dot_code <- '
digraph SEM {
    rankdir=BTLR;
    bgcolor="white";
    splines=true;
    compound=true;

    // Default styles
    node [shape=box, style="filled", color="black",
       fillcolor="aliceblue", fontname="Arial", fontsize=14];
    edge [color="black", fontname="Arial", fontsize=12];

    // Latent variables
    SES [shape=ellipse, fillcolor="lightskyblue", label="SES"];
    HL  [shape=ellipse, fillcolor="lightgray", label="Health
       Literacy"];

    // SES indicators
    Education   [label="Education\\nLevel"];
    LivingSit   [label="Living\\nSituation"];
    Employment [label="Employment\\nStatus"];
    Age         [label="Age"];
    Gender      [label="Gender"];

    // HL indicators (now light green)
    NHS [label="NHS", fillcolor="palegreen"];
    FHI [label="FHI", fillcolor="palegreen"];
    UHI [label="UHI", fillcolor="palegreen"];
    CA  [label="CA",  fillcolor="palegreen"];
    AMH [label="AMH", fillcolor="palegreen"];
    HPS [label="HPS", fillcolor="palegreen"];
```

```
    HSI [label="HSI", fillcolor="palegreen"];
    SS  [label="SS",  fillcolor="palegreen"];
    AE  [label="AE",  fillcolor="palegreen"];

    // Outcome (renamed)
    severity  [shape=box, fillcolor="mistyrose", label="Health Professional\\nSeverity Score"];

    // Rank alignment
    { rank=same; SES; HL; }
    { rank=max; NHS; FHI; UHI; CA; AMH; HPS; HSI; SS; AE; }
    { rank=min; severity; }

    // SES measurement model
    SES -> Employment [label="0.370*"];
    SES -> Education  [label="0.754*"];
    SES -> LivingSit  [label="0.398*"];
    SES -> Age        [label="-0.367*"];
    SES -> Gender     [label="0.026", color="black", style="dashed"];

    // HL measurement model
    HL -> HPS [label="0.652*"];
    HL -> HSI [label="0.716*"];
    HL -> AMH [label="0.419*"];
    HL -> SS  [label="0.558*"];
    HL -> CA  [label="0.484*"];
    HL -> AE  [label="0.795*"];
    HL -> NHS [label="0.803*"];
    HL -> FHI [label="0.779*"];
    HL -> UHI [label="0.775*"];

    // Structural model
    SES -> HL         [label="0.357*", color="blue", penwidth=2.0];
    HL  -> severity  [label="-0.093*", color="blue", penwidth=2.0];
    SES -> severity  [label="-0.087 (direct)", color="red", style="dashed", penwidth=2.0];
  }
  '
  gtf <- grViz(dot_code)

  DiagrammeRsvg::export_svg(gtf) %>%
    charToRaw() %>%
    rsvg::rsvg_png("sem_diagramtf.png", width = 800)
```

```

  return(gtf)
}

# Run it
create_sem_diagram()

```


```{r}
create_sem_diagram <- function() {
  dot_code <- '
digraph SEM {
    rankdir=BTLR;
    bgcolor="white";
    splines=true;
    compound=true;

    // Default styles
    node [shape=box, style="filled", color="black",
       fillcolor="aliceblue", fontname="Arial", fontsize=14];
    edge [color="black", fontname="Arial", fontsize=12];

    // Latent variables
    SES [shape=ellipse, fillcolor="lightskyblue", label="SES"];
    HL  [shape=ellipse, fillcolor="lightgray", label="Health
       Literacy"];

    // SES indicators
    Education  [label="Education\\nLevel"];
    LivingSit  [label="Living\\nSituation"];
    Employment [label="Employment\\nStatus"];
    Age        [label="Age"];
    Gender     [label="Gender"];

    // HL indicators (light green)
    NHS [label="NHS", fillcolor="palegreen"];
    FHI [label="FHI", fillcolor="palegreen"];
    UHI [label="UHI", fillcolor="palegreen"];
    CA  [label="CA",  fillcolor="palegreen"];
    AMH [label="AMH", fillcolor="palegreen"];
    HPS [label="HPS", fillcolor="palegreen"];
```

```
    HSI [label="HSI", fillcolor="palegreen"];
    SS  [label="SS",  fillcolor="palegreen"];
    AE  [label="AE",  fillcolor="palegreen"];

    // Outcome
    Comorbidity [shape=box, fillcolor="mistyrose",
       label="Comorbidities"];

    // Rank alignment
    { rank=same; SES; HL; }
    { rank=max; NHS; FHI; UHI; CA; AMH; HPS; HSI; SS; AE; }
    { rank=min; Comorbidity; }

    // SES measurement model
    SES -> Employment [label="0.353*"];
    SES -> Education  [label="0.603*"];
    SES -> LivingSit  [label="0.414*"];
    SES -> Age        [label="-0.423*"];
    SES -> Gender     [label="-0.054", color="black",
       style="dashed"];

    // HL measurement model
    HL -> HPS [label="0.645*"];
    HL -> HSI [label="0.708*"];
    HL -> AMH [label="0.422*"];
    HL -> SS  [label="0.540*"];
    HL -> CA  [label="0.487*"];
    HL -> AE  [label="0.795*"];
    HL -> NHS [label="0.799*"];
    HL -> FHI [label="0.780*"];
    HL -> UHI [label="0.774*"];

    // Structural model
    SES -> HL           [label="0.364*", color="blue", penwidth=2.0];
    HL  -> Comorbidity  [label="-0.041", color="blue", penwidth=2.0];
    SES -> Comorbidity  [label="-0.387* (direct)", color="red",
       style="dashed", penwidth=2.0];
  }
  '
  gtt <- DiagrammeR::grViz(dot_code)

  DiagrammeRsvg::export_svg(gtt) %>%
    charToRaw() %>%
```

````
    rsvg::rsvg_png("sem_diagram_comorbidity.png", width = 800)

  return(gtt)
}

# Run it
create_sem_diagram()

```

```{r}
newdata5 <- newdata4[!is.na(newdata4$SOM_Total), ]
dim(newdata5)
dim(newdata4)

```

#####Both Mastery and Health literacy
```{r}

# Complete model with all possible indirect paths
mastery_complete <- '
  # Latent variable for Health Literacy
Healthliteracy =~ HPS_1 + HSI_2 + AMH_3 + SS_4 + CA_5 + AE_6 + NHS_7 + FHI_8 + UHI_9

# Latent variable for Socioeconomic Status
SES =~ Employment_Status + PT_EDU_CAT + Living_Situation + PT_Age + PT_Gender

# All possible paths
Healthliteracy ~ a*SES                    # Path a: SES -> Healthliteracy
SOM_Total ~ b*Healthliteracy + e*SES      # Path b: Healthliteracy -> SOM_Total, Path e: SES -> SOM_Total
PT_Health ~ c*SES + d*SOM_Total + f*Healthliteracy  # All direct effects

# Define all individual indirect effects
indirect_abd := a*b*d                     # SES -> Healthliteracy -> SOM_Total ->
indirect_af := a*f                        # SES -> Healthliteracy -> PT_Health
````

```
indirect_ed := e*d                          # SES -> SOM_Total -> PT_Health

# Total indirect and total effects
total_indirect := indirect_abd + indirect_af + indirect_ed
total := c + total_indirect

# Proportion mediated by each pathway
prop_abd := indirect_abd / total          # Proportion via sequential path
prop_af := indirect_af / total            # Proportion via SES -> HL -> PT_Health
prop_ed := indirect_ed / total            # Proportion via SES -> SOM -> PT_Health

# Additional useful calculations
prop_direct := c / total                  # Proportion of direct effect

'

# Fit the model
fit_complete <- sem(
  model = mastery_complete,
  data = newdata5,
  estimator = "WLSMV",
  ordered = c("PT_EDU_CAT", "Living_Situation", "Employment_Status", "PT_Gender"))

summary(fit_complete, standardized = TRUE, fit.measures = TRUE)

```


```{r}
set.seed(10)
library(lavaan)
# Complete model with all possible indirect paths
mastery_complete1 <- '
  # Latent variable for Health Literacy
Healthliteracy =~ HPS_1 + HSI_2 + AMH_3 + SS_4 + CA_5 + AE_6 + NHS_7 + FHI_8 + UHI_9

```

```
# Latent variable for Socioeconomic Status
SES =~ Employment_Status + PT_EDU_CAT + Living_Situation + PT_Age + PT_Gender

# All possible paths
Healthliteracy ~ a*SES                       # Path a: SES -> Healthliteracy
SOM_Total ~ b*Healthliteracy + e*SES         # Path b: Healthliteracy -> SOM_Total, Path e: SES -> SOM_Total
Comorbidities ~ c*SES + d*SOM_Total + f*Healthliteracy  # All direct effects

# Define all individual indirect effects
indirect_abd := a*b*d                        # SES -> Healthliteracy -> SOM_Total -> Comorbidities
indirect_af := a*f                           # SES -> Healthliteracy -> Comorbidities
indirect_ed := e*d                           # SES -> SOM_Total -> Comorbidities

# Total indirect and total effects
total_indirect := indirect_abd + indirect_af + indirect_ed
total := c + total_indirect

# Proportion mediated by each pathway
prop_abd := indirect_abd / total             # Proportion via sequential path
prop_af := indirect_af / total               # Proportion via SES -> HL -> Comorbidities
prop_ed := indirect_ed / total               # Proportion via SES -> SOM -> Comorbidities

# Additional useful calculations
prop_direct := c / total                     # Proportion of direct effect

'

# Fit the model
fit_complete1 <- sem(
  model = mastery_complete1,
  data = newdata5,
  estimator = "WLSMV",
```

```
  ordered = c("PT_EDU_CAT", "Living_Situation", "Employment_Status",
     "PT_Gender"))

# View the results
summary(fit_complete1, standardized = TRUE, fit.measures = TRUE)

```

```{r}
set.seed(11)
library(lavaan)
# Complete model with all possible indirect paths
mastery_complete2 <- '
  # Latent variable for Health Literacy
Healthliteracy =~ HPS_1 + HSI_2 + AMH_3 + SS_4 + CA_5 + AE_6 + NHS_7
   + FHI_8 + UHI_9

# Latent variable for Socioeconomic Status
SES =~ Employment_Status + PT_EDU_CAT + Living_Situation + PT_Age +
   PT_Gender

# All possible paths
Healthliteracy ~ a*SES                    # Path a: SES ->
   Healthliteracy
SOM_Total ~ b*Healthliteracy + e*SES      # Path b: Healthliteracy
   -> SOM_Total, Path e: SES -> SOM_Total
PRO_SeverityPT ~ c*SES + d*SOM_Total + f*Healthliteracy  # All
   direct effects

# Define all individual indirect effects
indirect_abd := a*b*d                     # SES -> Healthliteracy ->
   SOM_Total -> PRO_SeverityPT
indirect_af := a*f                        # SES -> Healthliteracy ->
   PRO_SeverityPT
indirect_ed := e*d                        # SES -> SOM_Total ->
   PRO_SeverityPT

# Total indirect and total effects
total_indirect := indirect_abd + indirect_af + indirect_ed
total := c + total_indirect

# Proportion mediated by each pathway
```

````
prop_abd := indirect_abd / total          # Proportion via sequential path
prop_af := indirect_af / total            # Proportion via SES -> HL -> PRO_SeverityPT
prop_ed := indirect_ed / total            # Proportion via SES -> SOM -> PRO_SeverityPT

# Additional useful calculations
prop_direct := c / total                  # Proportion of direct effect

'

# Fit the model
fit_complete2 <- sem(
  model = mastery_complete2,
  data = newdata5,
  estimator = "WLSMV",
  ordered = c("PT_EDU_CAT", "Living_Situation", "Employment_Status", "PT_Gender"))

# View the results
summary(fit_complete2, standardized = TRUE, fit.measures = TRUE)

```



```{r}
set.seed(12)
library(lavaan)
# Complete model with all possible indirect paths
mastery_complete3 <- '
  # Latent variable for Health Literacy
Healthliteracy =~ HPS_1 + HSI_2 + AMH_3 + SS_4 + CA_5 + AE_6 + NHS_7 + FHI_8 + UHI_9

# Latent variable for Socioeconomic Status
SES =~ Employment_Status + PT_EDU_CAT + Living_Situation + PT_Age + PT_Gender

# All possible paths
````

```
Healthliteracy ~ a*SES                        # Path a: SES -> Healthliteracy
SOM_Total ~ b*Healthliteracy + e*SES        # Path b: Healthliteracy -> SOM_Total, Path e: SES -> SOM_Total
Comorbidities ~ c*SES + d*SOM_Total + f*Healthliteracy  # All direct effects

# Define all individual indirect effects
indirect_abd := a*b*d                         # SES -> Healthliteracy -> SOM_Total -> Comorbidities
indirect_af := a*f                            # SES -> Healthliteracy -> Comorbidities
indirect_ed := e*d                            # SES -> SOM_Total -> Comorbidities

# Total indirect and total effects
total_indirect := indirect_abd + indirect_af + indirect_ed
total := c + total_indirect

# Proportion mediated by each pathway
prop_abd := indirect_abd / total            # Proportion via sequential path
prop_af := indirect_af / total              # Proportion via SES -> HL -> Comorbidities
prop_ed := indirect_ed / total              # Proportion via SES -> SOM -> Comorbidities

# Additional useful calculations
prop_direct := c / total                    # Proportion of direct effect

'

# Fit the model
fit_complete3 <- sem(
  model = mastery_complete3,
  data = newdata5,
  estimator = "WLSMV",
  ordered = c("PT_EDU_CAT", "Living_Situation", "Employment_Status", "PT_Gender"))

# View the results
summary(fit_complete3, standardized = TRUE, fit.measures = TRUE)
```

```

```
```{r}
create_sem_diagram <- function() {
  dot_code <- '
digraph SEM {
    rankdir=BT;
    bgcolor="white";
    splines=true;
    compound=true;

    // Default styles
    node [shape=box, style="filled", color="black",
       fillcolor="aliceblue", fontname="Arial", fontsize=14];
    edge [color="black", fontname="Arial", fontsize=12];

    // Latent variables
    SES [shape=ellipse, fillcolor="lightskyblue", label="SES"];
    HL  [shape=ellipse, fillcolor="lightgray", label="Health
       Literacy"];
 SOM  [shape=box, fillcolor="lightgray", label="Mastery"];


    // SES indicators
    Education  [label="Education\\nLevel"];
    LivingSit  [label="Living\\nSituation"];
    Employment [label="Employment\\nStatus"];
    Age        [label="Age"];
    Gender     [label="Gender"];

    // HL indicators (light green)
    NHS [label="NHS"];
    FHI [label="FHI"];
    UHI [label="UHI"];
    CA  [label="CA"];
    AMH [label="AMH"];
    HPS [label="HPS"];
    HSI [label="HSI"];
    SS  [label="SS"];
    AE  [label="AE"];

    // Outcome
```

```
    severity [shape=box, fillcolor="mistyrose", label="Professional
       \\n Severity Score"];

    // Rank alignment

  { rank=max; NHS; FHI; UHI; CA; AMH; HPS; HSI; SS; AE; }
  { rank=same; HL; SOM;}
  { rank=same; SES;severity ;}
  { rank=min; Education; LivingSit; Employment; Age; Gender; }




    // SES measurement model
    SES -> Employment [label="0.424*"];
    SES -> Education  [label="0.714*"];
    SES -> LivingSit  [label="0.382*"];
    SES -> Age        [label="-0.365*"];
    SES -> Gender     [label="0.044", color="black", style="dashed"];

    // HL measurement model
    HL -> HPS [label="0.640*"];
    HL -> HSI [label="0.714*"];
    HL -> AMH [label="0.404*"];
    HL -> SS  [label="0.561*"];
    HL -> CA  [label="0.468*"];
    HL -> AE  [label="0.809*"];
    HL -> NHS [label="0.812*"];
    HL -> FHI [label="0.782*"];
    HL -> UHI [label="0.784*"];

    // Structural model
    SES -> HL           [label="0.355*", color="blue", penwidth=2.0];
    HL  -> SOM  [label="0.393*", color="blue", penwidth=2.0];
    SES -> severity [label="-0.026 (direct)", color="red",
       style="dashed", penwidth=2.0];
    SOM-> severity  [label="-0.282* ", color="blue",  penwidth=2.0];
  }
  '
  gtt <- DiagrammeR::grViz(dot_code)

  DiagrammeRsvg::export_svg(gtt) %>%
    charToRaw() %>%
```

```
    rsvg::rsvg_png("SOM_PROSEVERITY.png", width = 800)

  return(gtt)
}

# Run it
create_sem_diagram()

```
```

Without the health literacy

```{r}
# Step 2: Model Healthliteracy -> PT_Health

mastery1 <- '
  # Latent variable for Health Literacy


  # Latent variable for Socioeconomic Status
  SES =~ Employment_Status + PT_EDU_CAT + Living_Situation + PT_Age
     + PT_Gender

  # All possible paths
  SOM_Total ~ a*SES                    # Path a: SES -> mastery
PRO_SeverityPT ~ c*SES + b*SOM_Total  # All direct effects

  # Define only the sequential indirect effect
  indirect_seq := a*b                       # SES -> Healthliteracy ->
     SOM_Total -> PRO_SeverityPT
  total :=c+(a*b)
  prop:=indirect_seq/ total
'



# Fit the model
fit1 <- sem(
  model = mastery1,
  data = newdata5,
  estimator = "WLSMV",
  ordered = c( "PT_EDU_CAT","Living_Situation", "Employment_Status",
     "PT_Gender"))
```
```

````

# View the results
summary(fit1, standardized = TRUE, fit.measures = TRUE)


```
```{r}
create_sem_diagram <- function() {
  dot_code <- '
digraph SEM {
    rankdir=BT;
    bgcolor="white";
    splines=true;
    compound=true;

    // Default styles
    node [shape=box, style="filled", color="black", fillcolor="aliceblue", fontname="Arial", fontsize=14];
    edge [color="black", fontname="Arial", fontsize=12];

    // Latent variables
    SES [shape=ellipse, fillcolor="lightskyblue", label="SES"];
    SOM  [shape=box, fillcolor="lightgray", label="Mastery"];


    // SES indicators
    Education   [label="Education\\nLevel"];
    LivingSit   [label="Living\\nSituation"];
    Employment  [label="Employment\\nStatus"];
    Age         [label="Age"];
    Gender      [label="Gender"];



    // Outcome
    PRO_SeverityPT  [shape=box, fillcolor="mistyrose", label="Professional\\nSeverity Score"];

    // Rank alignment


  { rank=same; SOM;}
  { rank=same; SES;PRO_SeverityPT  ;}
````

```
  { rank=min; Education; LivingSit; Employment; Age; Gender; }




    // SES measurement model
    SES -> Employment [label="0.605*"];
    SES -> Education  [label="0.488*"];
    SES -> LivingSit  [label="0.393*"];
    SES -> Age        [label="-0.335*"];
    SES -> Gender     [label="0.281*"];



    // Structural model
    SES -> SOM          [label="0.425*", color="blue", penwidth=2.0];
    SES -> PRO_SeverityPT  [label="-0.090 (direct)", color="red",
       style="dashed", penwidth=2.0];
    SOM-> PRO_SeverityPT   [label="-0.243* ", color="blue",
       penwidth=2.0];
  }
  '
  gtt <- DiagrammeR::grViz(dot_code)

  DiagrammeRsvg::export_svg(gtt) %>%
    charToRaw() %>%
    rsvg::rsvg_png("SOM_Proseverity1.png", width = 800)

  return(gtt)
}

# Run it
create_sem_diagram()

```


#####Without HL Patient Health
```{r}
# Step 2: Model Healthliteracy -> PT_Health
mastery1 <- '
  # Latent variable for Health Literacy
```

```


  # Latent variable for Socioeconomic Status
  SES =~ Employment_Status + PT_EDU_CAT + Living_Situation + PT_Age
     + PT_Gender

  # All possible paths
  SOM_Total ~ a*SES                    # Path a: SES -> mastery
  PT_Health ~ c*SES + b*SOM_Total  # All direct effects

  # Define only the sequential indirect effect
  indirect_seq := a*b                     # SES -> Healthliteracy ->
     SOM_Total -> Pt_Health
  total :=c+(a*b)
  prop:=indirect_seq/ total
'



# Fit the model
fit1 <- sem(
  model = mastery1,
  data = newdata5,
  estimator = "WLSMV",
  ordered = c( "PT_EDU_CAT","Living_Situation", "Employment_Status",
     "PT_Gender"))

# View the results
summary(fit1, standardized = TRUE, fit.measures = TRUE)


```

```{r}
create_sem_diagram <- function() {
  dot_code <- '
digraph SEM {
    rankdir=BT;
    bgcolor="white";
    splines=true;
    compound=true;

    // Default styles
```

```
    node [shape=box, style="filled", color="black",
       fillcolor="aliceblue", fontname="Arial", fontsize=14];
    edge [color="black", fontname="Arial", fontsize=12];

    // Latent variables
    SES [shape=ellipse, fillcolor="lightskyblue", label="SES"];
    SOM  [shape=box, fillcolor="lightgray", label="Mastery"];


    // SES indicators
    Education  [label="Education\\nLevel"];
    LivingSit  [label="Living\\nSituation"];
    Employment [label="Employment\\nStatus"];
    Age        [label="Age"];
    Gender     [label="Gender"];



    // Outcome
    Comorbidity [shape=box, fillcolor="mistyrose", label="Patient
       \\nHealth Score"];

    // Rank alignment


  { rank=same; SOM;}
  { rank=same; SES;Comorbidity ;}
  { rank=min; Education; LivingSit; Employment; Age; Gender; }




    // SES measurement model
    SES -> Employment [label="0.630*"];
    SES -> Education  [label="0.494*"];
    SES -> LivingSit  [label="0.357*"];
    SES -> Age        [label="-0.323*"];
    SES -> Gender     [label="0.263*"];



    // Structural model
    SES -> SOM           [label="0.428*", color="blue", penwidth=2.0];
```

````
    SES -> Comorbidity [label="0.080 (direct)", color="red", style="dashed", penwidth=2.0];
    SOM-> Comorbidity  [label="0.375* ", color="blue", penwidth=2.0];
  }
  '
  gtt <- DiagrammeR::grViz(dot_code)

  DiagrammeRsvg::export_svg(gtt) %>%
    charToRaw() %>%
    rsvg::rsvg_png("SOM_PT_Health2.png", width = 800)

  return(gtt)
}

# Run it
create_sem_diagram()

```



#####wITHOUT hl Comorbidities

```{r}
# Step 2: Model Healthliteracy -> PT_Health
# Complete model but only extract sequential effect
mastery1 <- '
  # Latent variable for Health Literacy


  # Latent variable for Socioeconomic Status
  SES =~ Employment_Status + PT_EDU_CAT + Living_Situation+PT_Gender+PT_Age

  # All possible paths
  SOM_Total ~ a*SES                         # Path a: SES -> mastery
  Comorbidities ~ c*SES + b*SOM_Total   # All direct effects

  # Define only the sequential indirect effect
  indirect_seq := a*b                       # SES -> Healthliteracy -> SOM_Total -> Comorbidities
  total :=c+(a*b)
````

```
  prop:=indirect_seq/ total
'



# Fit the model
fit1 <- sem(
  model = mastery1,
  data = newdata5,
  estimator = "WLSMV",
  ordered = c( "PT_EDU_CAT","Living_Situation", "Employment_Status",
     "PT_Gender"))

# View the results
summary(fit1, standardized = TRUE, fit.measures = TRUE)


```

```{r}
create_sem_diagram <- function() {
  dot_code <- '
digraph SEM {
    rankdir=BT;
    bgcolor="white";
    splines=true;
    compound=true;

    // Default styles
    node [shape=box, style="filled", color="black",
       fillcolor="aliceblue", fontname="Arial", fontsize=14];
    edge [color="black", fontname="Arial", fontsize=12];

    // Latent variables
    SES [shape=ellipse, fillcolor="lightskyblue", label="SES"];
    SOM  [shape=box, fillcolor="lightgray", label="Mastery"];


    // SES indicators
    Education  [label="Education\\nLevel"];
    LivingSit  [label="Living\\nSituation"];
    Employment [label="Employment\\nStatus"];
    Age        [label="Age"];
```

```
    Gender     [label="Gender"];



    // Outcome
    Comorbidity [shape=box, fillcolor="mistyrose",
       label="Comorbidities"];

    // Rank alignment


  { rank=same; SOM;}
  { rank=same; SES;Comorbidity ;}
  { rank=min; Education; LivingSit; Employment; Age; Gender; }




    // SES measurement model
    SES -> Employment [label="0.506*"];
    SES -> Education  [label="0.494*"];
    SES -> LivingSit  [label="0.375*"];
    SES -> Age        [label="-0.463*"];
    SES -> Gender     [label="0.152*"];



    // Structural model
    SES -> SOM          [label="0.430*", color="blue", penwidth=2.0];
    SES -> Comorbidity [label="-0.316* (direct)", color="red",
       style="dashed", penwidth=2.0];
    SOM-> Comorbidity  [label="-0.138* ", color="blue",
       penwidth=2.0];
  }
  '
  gtt <- DiagrammeR::grViz(dot_code)

  DiagrammeRsvg::export_svg(gtt) %>%
    charToRaw() %>%
    rsvg::rsvg_png("SOM_Comorbidities1.png", width = 800)

  return(gtt)
}
```

```

# Run it
create_sem_diagram()

'''
```