Applied Statistics and Data Science MSc Dissertation MTH7022P MSc Dissertation, 2024/25

Equality and Improvement in Mental Health Outcome Scores in East London Healthcare

Divya Gitesh Kothari, ID 240485889

Supervisor: Dr Silvia Liverani and Dr Nicolás Hernández

A thesis presented for the degree of
Master of Science in *Applied Statistics and Data Science*

School of Mathematical Sciences
Queen Mary University of London

Declaration of original work

This declaration is made on August 31, 2025.

Student's Declaration: I, Divya Gitesh Kothari, hereby declare that the work in this thesis is my original work. I have not copied from any other students' work, work of mine submitted elsewhere, or from any other sources except where due reference or acknowledgment is made explicitly in the text. Furthermore, no part of this dissertation has been written for me by another person, by generative artificial intelligence (AI), or by AI-assisted technologies.

Referenced text has been flagged by:

1. Using italic fonts, **and**
2. using quotation marks "...", **and**
3. explicitly mentioning the source in the text.

This work is dedicated to my sister Jiya Kothari

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Dr Silvia Liverani and Dr Nicolás Hernández, for their invaluable guidance, encouragement, and support throughout this project. I also wish to thank the team at East London NHS Foundation Trust (ELFT) for providing access to the pseudonymised patient level DIALOG dataset and for the opportunity to work on a project with real world implications for mental healthcare.I am also thankful to the faculty and staff of the School of Mathematical Sciences, Queen Mary University of London for providing the necessary resources and environment to conduct this research. Special thanks to the Unit for Social and Community Psychiatry at Queen Mary University of London for their development of the DIALOG tool, which served as the foundation for this analysis. Finally, I am grateful to my friends and family for their constant support during this project.

Abstract

This project explores whether mental health treatment outcomes in East London are experienced equally by patients from different demographic backgrounds. Using the DIALOG scale(Priebe et al., 2007), which collects patient-reported satisfaction across areas such as mental and physical health, relationships, and treatment, the data was analysed from over 6,000 records collected each year. Each patient had scores recorded at multiple time points(before and after), allowing for the measurement of change over the course of treatment.

The aim was to examine whether there are differences in how various gender and ethnic groups begin treatment, how their scores change over time, and whether any groups show consistently better or worse outcomes. Data cleaning and preprocessing was carried out, followed by clustering methods to identify patterns and potential outliers.

The findings suggest that while many patients show improvement, some groups differ in how they report their experiences and outcomes. These insights can support ongoing work to make mental health services in East London more equitable, helping ensure all patients receive fair and effective care.

Contents

Chapter 1

Introduction

Background: Mental Health and Fairness

East London is a diverse area with people from many different backgrounds, cultures, and communities. Due to this, mental health services here face unique challenges. One of the biggest concerns is making sure that everyone gets fair and equal treatment regardless of their gender, ethnicity, or background. To help track how people feel about their mental health and the care they receive, the East London NHS Foundation Trust (ELFT) uses a tool called DIALOG. This is a patient-reported outcome measure that asks patients to rate their satisfaction across several life areas and aspects of their treatment on a 7-point scale. With over 20,000 DIALOG scores recorded every year for patients in East London, there is a valuable dataset that can help us understand patient outcomes better.

Motivation:

Tracking patient satisfaction and quality of life during treatment is important. It helps clinicians and policymakers see which services are working well and which may need improvement. If certain groups of patients based on gender or ethnicity are not seeing the same improvements, this could point to gaps or inequalities in the care system. By identifying these differences, we can help guide changes that make mental health services more inclusive and effective for everyone. This not only improves individual outcomes but also builds a better and fairer healthcare system where all patients have equal chances of recovery and well-being.

Project Objective

This project looks at patient satisfaction using DIALOG scores taken before and after treatment, together with demographic details. The aim is to understand levels of improvement, overall patient experiences, and how these experiences may differ across groups.

The key aims are to:

- Cluster patients using combined DIALOG scores from before and after treatment. This will help identify hidden subgroups that reflect overall satisfaction profiles over time. Unsupervised learning methods, including K-Means, Gaussian Mixture Models (GMM), and Hierarchical Clustering, will be applied.
- Investigate demographic influences, particularly gender and ethnicity, to see whether different groups experience care differently. Also examine other characteristics such as CPA status, IMD decile, and age.
- Compare clustering scenarios using different inputs: before-only scores, after-only scores, combined scores, and scores plus demographics. Vary the number of clusters to assess stability and interpretability.
- Conduct statistical tests: Chi-squared tests to examine demographic differences between clusters, and Kruskal–Wallis tests to assess variations in specific satisfaction domains.
- Visualise clustering results using PCA (2D plots) to illustrate cluster separation and make patterns easier to interpret.
- Identify disparities or patterns that may indicate gaps or unfairness in care delivery.

  The findings will support the East London NHS Foundation Trust in making decisions to improve mental health care quality and fairness across their services.

Achievements

The project produced a cleaned dataset and applied multiple clustering approaches to identify meaningful subgroups of patients. Statistical tests revealed significant demographic influences on treatment experiences, particularly by gender and ethnicity.
Visualisations highlighted disparities in satisfaction trajectories, offering evidence of inequities in outcomes. These results provide insights that can support ELFT in developing fairer and more effective mental health services.

Chapter 2

Data Description

2.1 Dataset Overview

This project is based on a dataset provided by the East London NHS Foundation Trust (ELFT). The data was collected using the DIALOG assessment tool, a validated Patient Reported Outcome Measure (PROM) designed to evaluate the satisfaction of mental health service users across multiple aspects of life and treatment. The tool was developed by the WHO Collaborating Centre for Mental Health Service Development at Queen Mary University of London. The dataset contains **6,281** anonymised patient records and **35** variables, collected using the DIALOG assessment tool. The dataset follows a longitudinal structure, meaning each patient was assessed at two time points referred to as Time 1 (initial) and Time 2 (follow up) covering assessments conducted between **June 2023 and May 2025.** This format allows to examine changes over time in how patients perceive different aspects of their life and treatment.

Each row in the dataset corresponds to a single patient's response at either time point, with a unique identifier (PatientKey) allowing us to match assessments for the same individual across both time points and carry out paired analyses.

2.1.1 DIALOG Scoring System

The DIALOG tool measures satisfaction across various life and treatment domains using a 7 point Likert scale. The scores are the main outcome variables of interest, ranging from **1** (Very Dissatisfied) to **7** (Very Satisfied). Each score reflects the patient's satisfaction with **8 Life Domains (outcome variables)**:

- **Mental Health**: Emotional well-being and ability to cope
- **Physical Health**: General health, diet, exercise
- **Job Situation**: Employment or daily activities

- **Accommodation**: Living conditions and housing stability
- **Leisure Activities**: Hobbies and social involvement
- **Relationships**: Interactions with family or partner
- **Friendships**: Social circle and support outside family
- **Personal Safety**: Feeling safe from harm (self/others)

Each of these is scored twice, once at each assessment time point, to observe how it changes over time. Scores below 4 indicate dissatisfaction, scores around 4 are neutral, and scores above 4 reflect satisfaction (Priebe et al., 2007).

2.2 Data Variables

The dataset includes 35 variables, of which 15 are categorical type and 20 are numeric. The variables fall into two main categories:

2.2.1 Variable Categories

Demographic and Identifying Information:

- **PatientKey**: A unique pseudonymised ID assigned to each patient.
- **Sex**: Patient's gender (Male, Female, Other/Unknown; 0 missing values).
- **EthnicGroup**: Ethnic background (6 values: White, Black, Asian, Mixed, Other, Missing; 381 missing entries (6.07%))
- **IMD Decile**(Index of Multiple Deprivation): Socioeconomic status; 1 (most deprived) to 10 (least deprived) (2.26% missing)

Assessment Metadata:

- **PeriodCovered**: Date range of the data extracted (identical for all rows).
- **CPAStatus**: Whether the patient is under the Care Programme Approach (CPA), used to manage complex mental health needs (Yes/No).
- **AssessmentDate1/2** : Timestamps of each assessment.
- **AgeAtAssessment1, AgeAtAssessment2**: Patient age at each assessment (Patients ranged from 12 to 90 years old).
- **MeetingType**: Type of clinical meeting (Assessment, Review, Discharge).

- **TeamName, TeamNationalDesc, Directorate, Area**: Name of the clinical team, national level description of the team's function(e.g., Adult Mental Health, CAMHS) overseeing the service, directorate and service type (inpatient or outpatient). (These exhibit 12%–33% missingness, retained for optional subgroup analysis)

2.2.2 Missing Data

Most numeric variables are complete. However, several categorical variables have notable missing values as detailed below:

- `TeamNationalDesc1` and `TeamNationalDesc1/2` each have 33.21% missing values
- `Directorate2` and `TeamName2` each have approximately 20% missing values
- `EthnicGroup` is missing in **381 entries (6.07%)**

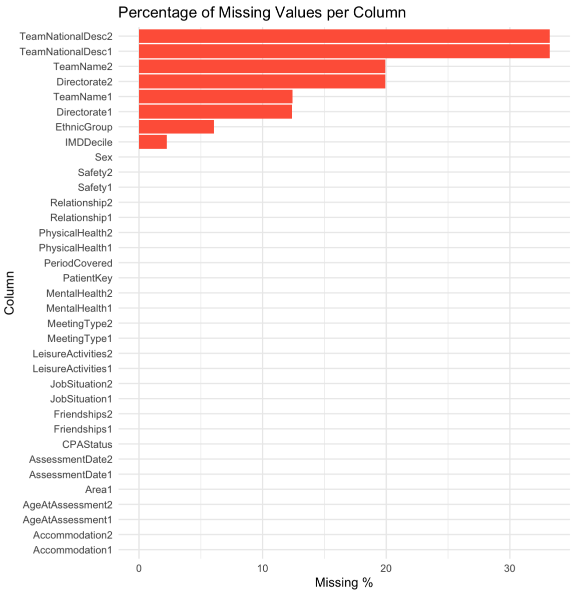


Figure 2.1: Missing data pattern across key variables

2.2.3 Age Distribution

Patient age at each assessment ranges from 12 to 90 years, reflecting a wide span of service users across adolescence, adulthood, and older age. The age distribution is summarised below:

- **Time 1:** Mean = 38.8 years, Median = 37 years, Inter Quartile Range (IQR) = 27–48 years
- **Time 2:** Mean = 39.7 years, Median = 37 years, Inter Quartile Range (IQR) = 28–49 years

The central tendency and spread remain consistent across Time 1 and Time 2, with most patients concentrated in their late 30s to early 40s. The similarity in quartiles and median suggests that age distribution remained stable over time, without significant shifts in the population structure between assessments.

2.2.4 Domain Scores (paired)

Box plots comparing pre and post treatment scores shows that scores after treatment are generally higher across most domains, with median values shifting upward.
In some areas, the scores also became less spread out, meaning patients had more similar, positive experiences. A few low scores remain after treatment, showing that not everyone improved equally. Overall, the plots suggest that most patients felt better after treatment.
These suggest a general improvement across all domains between the two assessments.

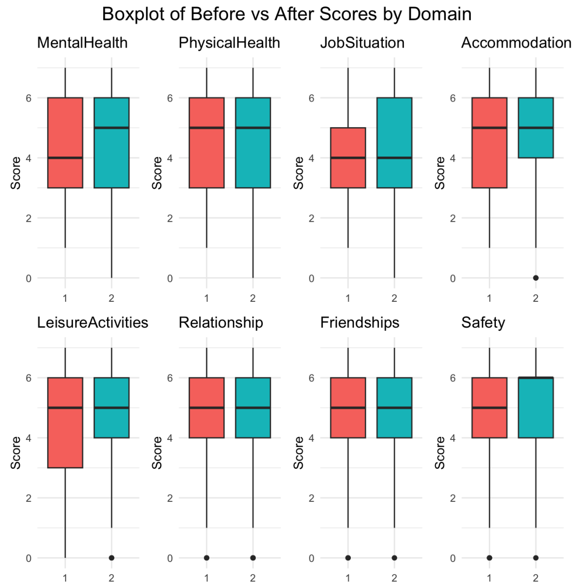


Figure 2.2: Box plots of domain scores before and after treatment

2.3 Data Pre processing

Inclusion Criteria (filtering): Patients were included in the analysis only if they had completed DIALOG assessments at both Time 1 and Time 2. This ensured the availability of paired domain scores for each individual.

Derived Metrics: The following metrics were calculated for exploratory and statistical analysis :

- **Change Scores:** Calculated for each domain as the difference between post and pre treatment scores (`MentalHealth2 - MentalHealth1`).
- **Time Between Assessments:** Computed from the `Length_Of_Time` variable and categorized into five groups: ≤1 month, 1–3 months, 3–6 months, 6–12 months, and >1 year.
- **Change Categories:** Each domain was labeled as *Improved*, *No Change*, or *Declined* based on the direction of score change.

Missing Data: No imputation was applied to missing values in demographic variables. All analyses were conducted using complete case data only.

2.3.1 Subgroup Analyses (Visual Insights)

A. Treatment Duration

- The figure shows the average score change for each domain across five treatment duration categories.
- The **6 to 12 months** group shows the highest improvements in most domains, particularly in *MentalHealth*, *Safety*, and *JobSituation.*
- Shorter durations (especially ≤1 month) have minimal or no improvement across most domains.
- For domains like *Relationship* and *Friendships*, even moderate durations (1–3 or 3–6 months) show notable improvements.
- In some domains (PhysicalHealth and LeisureActivities), longer durations beyond one year seem to have less impact.

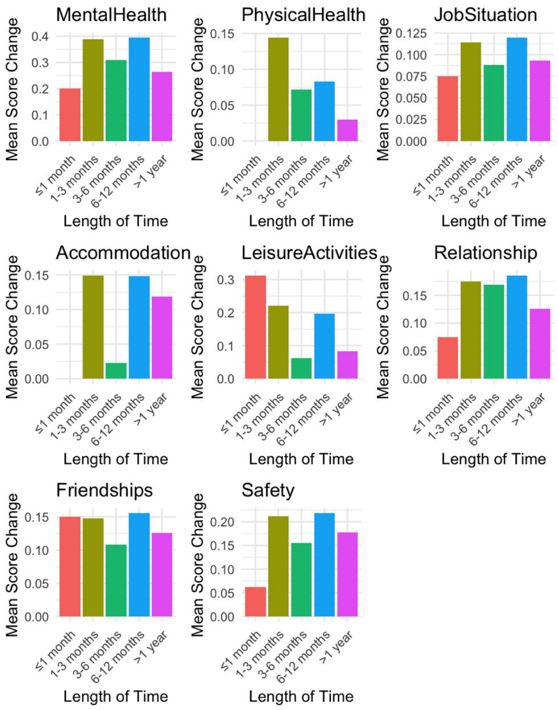


Figure 2.3: Average Change in Scores by Length of Treatment Duration

B. CPA Status

- *MentalHealth1* scores (Assessment 1) compared by CPA status.
- Patients under CPA had a higher average score (5) than those not under CPA (4).
- This pattern is unexpected, as CPA is typically assigned to patients with more complex mental health needs.
- The plot highlights a clear difference in baseline mental health perceptions between CPA and non-CPA groups.

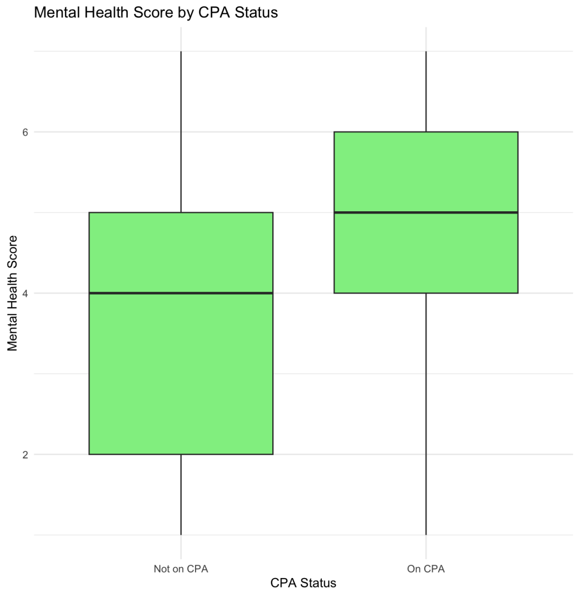


Figure 2.4: Baseline Mental Health Scores by CPA Status

C. Gender-Based Score Analysis

Baseline Scores by Gender

- Males had slightly higher baseline scores than females in most domains, particularly in *MentalHealth*, *LeisureActivities*, *PhysicalHealth*, and *Safety*.
- *JobSituation* showed minimal difference between genders at baseline.
- Females consistently reported lower well being scores before treatment across domains.
- This suggests that males and females may begin treatment from slightly different self perceived mental and social states.

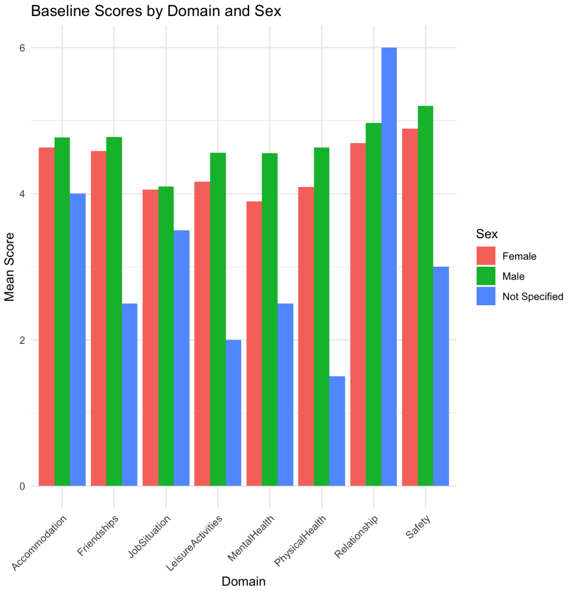


Figure 2.5: Baseline Scores by Gender

Score Change by Gender

- Females showed greater average improvement across most domains, especially in *MentalHealth* (r=0.38 vs. 0.28), *Friendships*, and *Safety*.
- The only domain without significant gender difference in improvement was *JobSituation*.
- Heatmap results indicate that female patients responded slightly better to treatment overall.
- This suggests a potential gender-based variation in treatment response patterns.

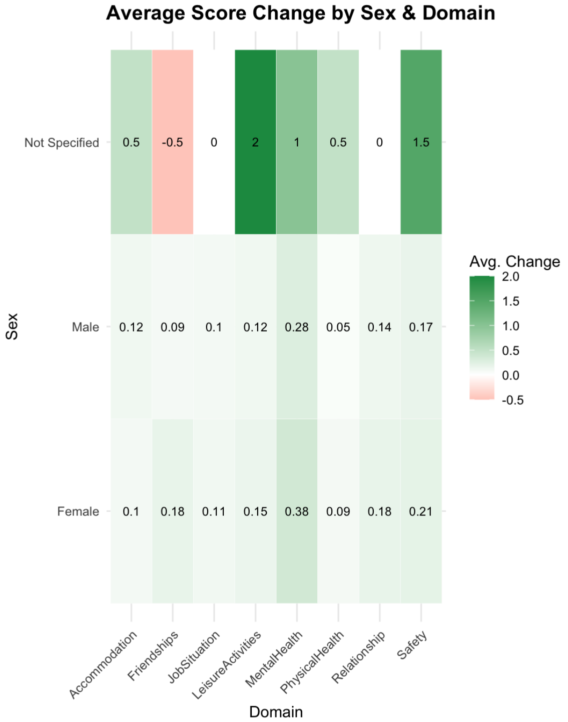


Figure 2.6: Score Change by Gender and Domain (Heatmap)

Statistical Findings

- Paired t-tests revealed statistically significant improvements in females for **7 out of 8 domains**. **Job Situation** was the only domain with no significant gender difference in change. In all domains with significant differences, **females scored higher than males** post-treatment.

D. Ethnic Group Analysis

Baseline Scores by Ethnic Group

- The largest ethnic group is **White**.
- *MentalHealth, LeisureActivities* and *PhysicalHealth* scores showed noticeable differences across ethnicities.
- **Black** patients reported the highest average scores in *Safety*, while **Mixed** ethnicity patients scored highest in *Friendships*.
- Domains like *Accommodation* and *Relationships* had similar medians across groups, suggesting comparable living and relational conditions.
- Overall, baseline domain scores showed modest variation, highlighting differences in initial wellbeing between ethnic groups.

Score Change by Ethnic Group

- The highest improvement in *MentalHealth* was observed in the **"Not specified"** group (+0.61), followed by **Other** (+0.35) and **Asian** (+0.32).
- The **Black** group showed the least improvement overall, especially in *MentalHealth* and *PhysicalHealth*.
- The "Not specified" group showed strong improvement across multiple domains, highlighting the importance of including underreported groups in the analysis.
- This suggests differences in treatment outcomes, showing the need for better approaches across ethnic groups.

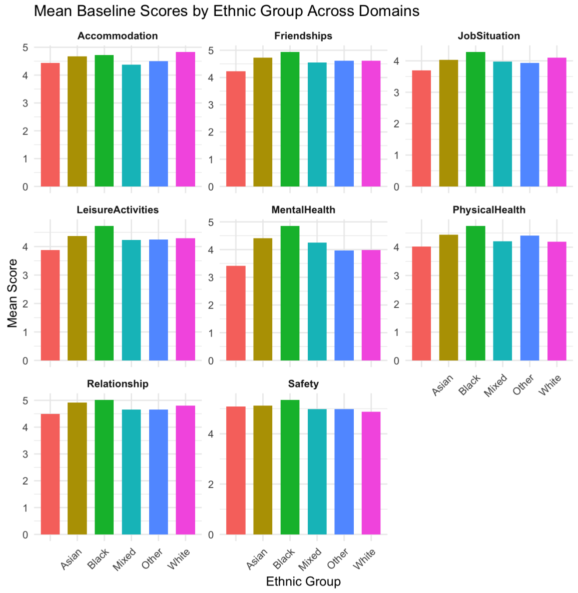


Figure 2.7: Baseline Domain Scores by Ethnic Group

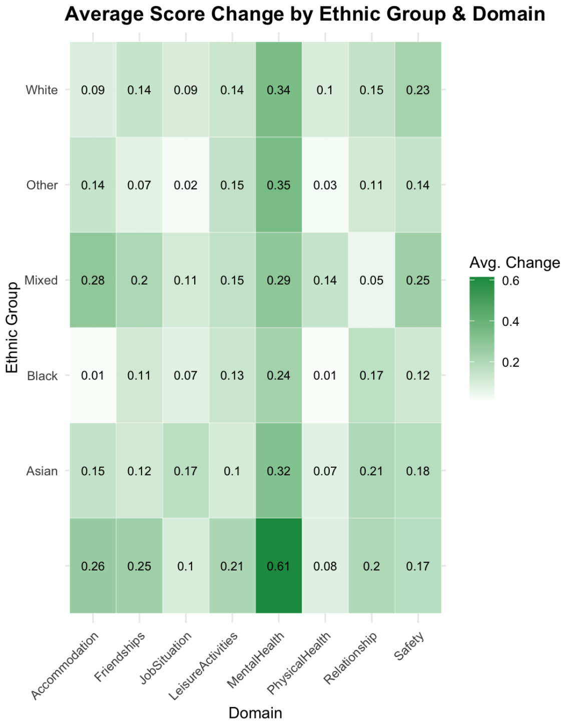


Figure 2.8: Score Change by Ethnic Group and Domain (Heatmap)

D.Socioeconomic Context(IMD Decile)

- Reveals insights into how socioeconomic status may affect recovery.
- The bar plot shows average score change across domains for three IMD bands: **Most deprived**, **Moderately deprived**, and **Least deprived**.
- Patients from the **least deprived** regions reported the greatest improvements in domains such as *Mental Health*, *Friendships*, and *Physical Health*.
- Those from **most deprived** areas showed more moderate improvements in these clinically sensitive domains.
- Domains like *Accommodation*, *Safety*, *Job Situation*, and *Leisure Activities* showed more balanced improvement across all IMD bands.
- While all groups benefited, service effectiveness may not be uniformly distributed across deprivation levels.

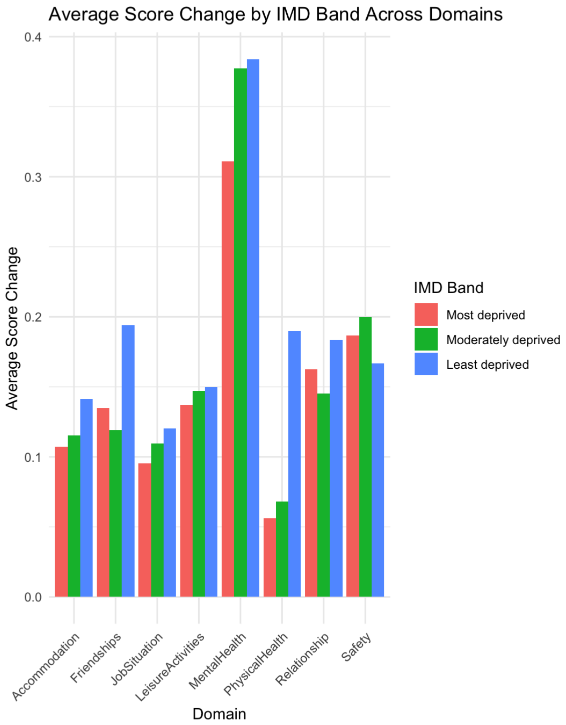


Figure 2.9: Average Score Change by IMD Band Across Domains

E. Correlation Between Domains

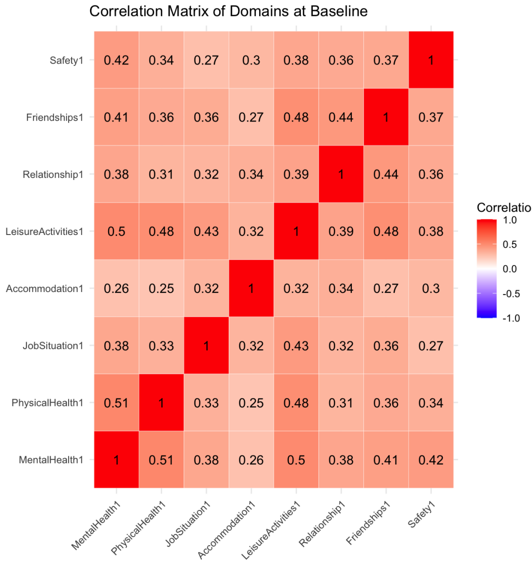
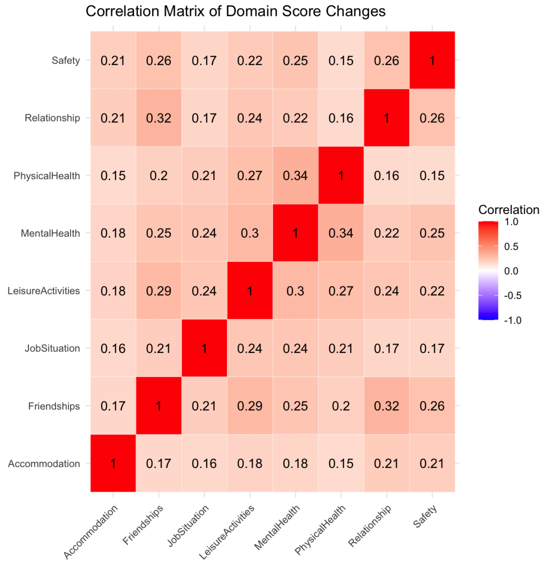


Figure 2.10: Correlation matrix at baseline.



Correlation matrix of domain score changes.

1. Domains at Baseline

- All domains show **positive correlations**, indicating they are closely linked at baseline.
- **Mental Health** is strongly linked with **Physical Health** (r = 0.51) and **Leisure Activities** (r = 0.50).
- **Friendships** and **Leisure Activities** also show a strong connection (r = 0.48).
- Social and lifestyle areas like **Job Situation**, **Friendships**, and **Relationships** are moderately related.
- Suggests that doing well in one area often results in better outcomes in others.

2. Domain Score Changes

- Score change correlations are **weaker but still positive**(r= 0.15 to 0.34).
- Strongest improvement overlap is between **Mental Health** and **Physical Health** (r = 0.34).
- **Relationships** and **Friendships** tend to improve together (r = 0.32), reflecting social recovery links.
- Minimal co-improvement is seen in domains like **Accommodation** and **Job Situation**.
- This shows recovery differs by area, so support should be personalised.

Combined Insight

- At the start of treatment, different areas of lives are closely linked. After treatment, improvements are less connected.
- Domains like **Mental and Physical Health** tend to improve together, while others like **Job Situation** and **Accommodation** show little overlap.
- This shows that people recover in different ways depending on their individual circumstances.

Chapter 3

Methods

Methods Overview
This study combines unsupervised learning techniques and statistical hypothesis testing to analyze DIALOG domain scores of mental health patients before and after treatment. Clustering algorithms were used to identify subgroups of patients with similar response profiles, while T-tests were used to assess the significance of changes in scores and group differences. Chi-squared tests were used to examine demographic differences between clusters, and Kruskal-Wallis tests to assess variations in specific satisfaction domains.

3.1 Clustering

Unsupervised learning is a branch of machine learning that deals with unlabeled data. It is designed to find patterns and relationships within the data without any prior knowledge of the labeled outputs or target variable. **Clustering** is an unsupervised machine learning technique which groups similar data points together based on patterns, similarity in the data or distance metrics (like Euclidean distance). The main goal is to organise the data into subgroups called clusters, where items in the same cluster are more alike than those in other clusters. It helps reveal hidden patterns or structures in unlabeled data. It can be used to reduce dimensionality of large datasets because observations within a cluster can be summarized by its center. This study applies three clustering methods: K-Means (centroid-based), Gaussian Mixture Models (model-based), and Agglomerative Hierarchical clustering.

3.2 K - means Clustering(Centroid-based)

K-Means is a form of hard clustering, which means that each data point is assigned to exactly one cluster. It is a partition based unsupervised machine learning algorithm used to classify a set of observations into K distinct non-overlapping clusters, where each observation belongs to the cluster with the nearest mean (centroid). The algorithm aims to identify natural groupings in the data by minimizing within-cluster variance.

Objective: Given observations $X = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$, where each $x_i \in \mathbb{R}^d$, K-Means partitions them into K clusters $C = \{C_1, C_2, \ldots, C_K\}$ to minimize the within-cluster sum of squares (WCSS), defined as:

$$\mathbf{WCSS} = \sum_{j=1}^{K} \sum_{x_i \in C_j} \|x_i - \boldsymbol{\mu}_j\|^2$$

x_i: the i^{th} data point $\quad$ C_j: the j^{th} cluster

$\|\cdot\|$: Euclidean norm (L2 distance); μ_j: centroid (mean vector) of cluster C_j, calculated as:
$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

Steps:

1. **Initialization:** Select K initial centroids (randomly or using a method such as the Elbow or Silhouette method).

2. **Assignment Step:** For each data point x_i, compute the squared Euclidean distance $\|x_i - \mu_j\|^2$ to each centroid μ_j, and assign x_i to the cluster with the nearest centroid.

3. **Update Step:** For each cluster j, update the centroid as the mean of all points assigned to that cluster:
$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

4. Repeat Steps 2 and 3 until cluster assignments no longer change or a maximum number of iterations is reached. This means the algorithm has converged to a solution.

Selecting the Number of Clusters: Selecting too few clusters can oversimplify the data, while too many can overfit the noise. Two common methods to help decide the optimal K are the Elbow Method and the Silhouette Method.

- **Elbow Method:** evaluates the Within Cluster Sum of Squares (WCSS) for different values of K. K-Means clustering is run for a range of K values (e.g., from 1 to n). For each K, the corresponding WCSS is calculated. These values are then plotted with K on the x-axis and WCSS on the y-axis. The optimal number of clusters are identified by locating the point on the plot where the rate of decrease in WCSS slows down significantly, forming a visible "elbow" or bend. This elbow represents a balance between minimizing WCSS and avoiding overly complex models with too many clusters.

- **Silhouette Score:** measures how similar a data point is to its own cluster compared to other clustersby balancing two key aspects.
  - **Cohesion** $a(i)$**:** The average distance between a data point i and all other points in the same cluster.
  - **Separation** $b(i)$**:** The average distance between i and all points in the nearest neighboring cluster.

  The Silhouette Score for a single point i is given by $s(i) = \dfrac{b(i) - a(i)}{\max\{a(i),\, b(i)\}}$.

  A higher mean Silhouette Score (close to 1) indicates well-separated and compact clusters. Scores near 0 or negative suggest overlapping clusters or misclassified points. To select the optimal number of clusters k, run clustering for different K values, compute the mean Silhouette Score for each, and choose the k with the highest average score.

Implementation in R: K-Means clustering is applied using R's built-in **kmeans()**function, which partitions data points into clusters by minimizing the total within-cluster sum of squares (WCSS). The algorithm iteratively assigns each data point to the nearest cluster centroid and updates centroids until convergence.

3.3 Gaussian Mixture Model (Model-based)

Latent Variable Models and Latent Profile Analysis:
Latent Variable Models (LVMs) are statistical models that assume observed data are influenced by hidden (latent) variables or factors that cannot be directly measured. Latent Profile Analysis (LPA) is a specific type of LVM designed for continuous observed variables. It aims to identify distinct subgroups or profiles within the data. Each profile represents a group of observations that share similar characteristics, with the assumption that these groups come from different probability distributions. Gaussian Mixture Models (GMMs) are a form of LPA that model profiles as mixtures of Gaussian distributions.
Gaussian Mixture Models:
Finite mixture models (FMMs) provide the statistical framework for model-based clustering by assuming that observed data are generated from a finite mixture of probability distributions. **Gaussian mixture models (GMMs)** are a type of FMMs which specifically assume that each of the underlying distributions is a Gaussian distribution. The data within each cluster are normally distributed, but with potentially different means and covariance matrices. GMMs provide *soft clustering*, giving probabilities of membership to each cluster.

A Gaussian Mixture Model (GMM) assumes the data is generated from K Gaussian distributions, each with its own parameters: mean vector $\boldsymbol{\mu}_k$ (center of the distribution), covariance matrix $\mathbf{\Sigma}_k$ (spread and orientation), and mixing proportion π_k (weight of that cluster):

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \mathbf{\Sigma}_k),$$

where the mixing proportions satisfy $\sum_{k=1}^{K} \pi_k = 1$.

Here, $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \mathbf{\Sigma})$ denotes the multivariate Gaussian distribution whose probability density function is:

$$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{d/2} \, |\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right),$$

where d is the dimensionality of the data, and $|\Sigma|$ is the determinant of the covariance matrix.

Clusters modeled by a GMM are ellipsoidal in shape. Their characteristics like volume, shape, and orientation are determined by the covariance matrices $\Sigma_1, \ldots, \Sigma_K$. These covariance matrices can be parameterized using an eigen-decomposition:

$$\mathbf{\Sigma}_k = \lambda_k \, \mathbf{U}_k \, \mathbf{\Delta}_k \, \mathbf{U}_k^\top,$$

where $\lambda_k = |\mathbf{\Sigma}_k|^{1/d}$ controls the volume, $\mathbf{\Delta}_k$ is a diagonal matrix of normalized eigenvalues ($|\mathbf{\Delta}_k| = 1$) controlling the shape, and $\mathbf{U}_k$ is an orthogonal matrix of eigenvectors controlling the orientation.

Covariance models are coded by three letters describing how volume, shape, and orientation vary across clusters. **E**: parameter identical across clusters, **V**: parameter varies by cluster, **I**: covariance is isotropic (equal variance). This coding allows GMMs to model clusters with varying size and shape.

Parameter Estimation via the EM Algorithm:

Given a random sample $\{x_1, x_2, \ldots, x_n\}$ in d dimensions, the goal is to estimate the set of parameters:

$$\theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}$$

that best explains the observed data. This is done by maximizing the log-likelihood function:

$$\ell(\theta) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \pi_k \, \phi_d(x_i; \mu_k, \Sigma_k) \right),$$

where $\theta = (\pi_{1:K-1}, \mu_{1:K}, \Sigma_{1:K})$, $\phi_d(\cdot)$ denotes the d-dimensional Gaussian density. Maximising this likelihood directly is difficult because of the summation inside the logarithm, which makes the optimisation problem nonlinear and complex. The EM algorithm solves this by introducing latent variables indicating cluster membership and iteratively updating parameters.

- **E-step (Expectation):** Compute the posterior probability that observation x_i belongs to cluster k:
$$\tau_{ik} = \frac{\pi_k \, \phi_d(x_i; \mu_k, \Sigma_k)}{\sum_{g=1}^{K} \pi_g \, \phi_d(x_i; \mu_g, \Sigma_g)}.$$
Here, τ_{ik} measures the probability that observation i belongs to cluster k, and it lies between 0 and 1.

- **M-step (Maximization):** Update parameters using these responsibilities τ_{ik}:
$$\pi_k = \frac{1}{n}\sum_{i=1}^{n} \tau_{ik}, \quad \boldsymbol{\mu}_k = \frac{\sum_{i=1}^{n} \tau_{ik}\mathbf{x}_i}{\sum_{i=1}^{n} \tau_{ik}}, \quad \boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{n} \tau_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top}{\sum_{i=1}^{n} \tau_{ik}}.$$

The EM algorithm alternates between these two steps until convergence is achieved, meaning the parameter estimates stabilize or the maximum number of iterations is reached.

Implementation in R:

The **mclust** package provides an easy interface for GMM clustering. If the number of clusters (G) or the covariance model (modelNames) is not specified, it fits all possible combinations and selects the model with the highest BIC. BIC and ICL can be computed directly with the functions **mclustBIC()** and **mclustICL()**. The results are visualized through plots which show an "elbow point".

Model Selection:

When applying model based clustering, one important step is deciding which model best fits the data. This involves choosing both the number of clusters and the covariance structure. Increasing the number of clusters generally improves model fit but can lead to overfitting, so careful model selection is necessary. Two of the most used tools for this are the **Bayesian Information Criterion (BIC)** and the **Integrated Complete Likelihood (ICL)**. Both criteria balance model fit (how well the model explains the data) against model complexity (number of parameters).

Bayesian Information Criterion (BIC):For a given model M,

$$\mathrm{BIC}_M = 2\,\ell_M(\hat{\theta}) - \nu_M \log(n),$$

where $\boldsymbol{\ell}_M(\hat{\boldsymbol{\theta}})$ is the maximized log-likelihood, $\boldsymbol{\nu}_M$ is the number of independent parameters estimated in model M, and $\boldsymbol{n}$ is the sample size.

Integrated Complete Likelihood(ICL):

$$\mathrm{ICL}_M = \mathrm{BIC}_M + 2\sum_{i=1}^{n}\sum_{k=1}^{K} c_{ik}\log(\tau_{ik}),$$

where $\boldsymbol{c}_{ik} = 1$ if observation x_i is assigned to cluster C_k, and 0 otherwise.

To select the best model:

1. Fit multiple models (GMMs) with different numbers of clusters K and covariance structures.
2. Calculate BIC and ICL for each model. Choose the model with the highest BIC or ICL value, as this indicates the best balance between data fit and model simplicity.
3. If no single model stands out, plot the criterion values against the number of clusters. Choose the point (elbow) where improvement slows down.

3.4 Hierarchical Clustering

Hierarchical clustering is a technique that builds a hierarchy of clusters without requiring the number of clusters to be specified in advance. The method produces a tree-like diagram called a dendrogram, which visually represents the nested grouping of observations and the order in which clusters are merged or split. There are two main approaches:

- Agglomerative (bottom-up) – starts with each observation as its own cluster and iteratively merges the closest clusters until all points belong to a single cluster.
- Divisive (top-down) – starts with all observations in one cluster and recursively splits clusters into smaller groups.

Distance Metrics The similarity (or dissimilarity) between points is measured using a distance metric such as:

Euclidean distance

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{m=1}^{p} (x_{im} - x_{jm})^2}$$

Manhattan distance

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{m=1}^{p} |x_{im} - x_{jm}|$$

Here, $\mathbf{x}_i \in \mathbb{R}^p$ is the i-th observation in p-dimensional space, and $d(\mathbf{x}_i, \mathbf{x}_j)$ is the distance between observations i and j.

Linkage Criteria The linkage method determines how distances between clusters are computed. Common approaches include:

- **Single linkage:** The distance between two clusters is the distance between the closest points in each cluster.This can make long, stretched out clusters.

$$D(C_a, C_b) = \min_{x_i \in C_a,\, x_j \in C_b} d(x_i, x_j)$$

- **Complete linkage:** The distance between two clusters is the distance between the farthest points in each cluster. This usually makes compact, tight clusters.

$$D(C_a, C_b) = \max_{x_i \in C_a,\, x_j \in C_b} d(x_i, x_j)$$

- **Average linkage:** Average distance between all pairs of points from two different clusters, measuring how far apart the clusters are.

$$D(C_a, C_b) = \frac{1}{|C_a||C_b|} \sum_{x_i \in C_a} \sum_{x_j \in C_b} d(x_i, x_j)$$

- **Ward's method:** Merges clusters that minimize the increase in total within-cluster variance:

$$\Delta(C_a, C_b) = \frac{|C_a||C_b|}{|C_a| + |C_b|} \|\bar{x}_a - \bar{x}_b\|^2,$$

  where $\bar{x}_a$ is the centroid of cluster C_a.

where C_k is the k-th cluster, $d(x_i, x_j)$ is the distance between observations i and j, and $D(C_a, C_b)$ is the distance between clusters C_a and C_b according to the linkage criterion.

Steps of Agglomerative Hierarchical Clustering:

1. **Initialize clusters:** Treat each data point as its own cluster.
2. **Calculate and update distances:** Compute distances between clusters and after merging the closest pair, update the distances using a linkage method(e.g. complete, Ward, etc).
3. **Merge clusters iteratively:** Repeat merging the closest clusters and updating distances until all points form one cluster or a stopping condition is reached.
4. **Visualize results:** Use a dendrogram to display the cluster hierarchy and help decide the cluster structure.

3.5 Statistical Tests:

In addition to clustering, statistical tests were used to check how mental health scores changed over time and whether clusters were linked to patient demographics. These tests added evidence to support the clustering results, showing real differences between groups and changes over time.

Paired t-test

The paired t-test is a parametric test that evaluates whether the mean difference between two related measurements is significantly different from zero. It is particularly appropriate when the same participants are measured at two time points, as in this study where mental health scores were collected before and after treatment. The test is based on the differences $d_i = X_i - Y_i$ between paired observations. The test statistic is calculated as:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

where $\bar{d}$ is the mean of the differences, s_d the standard deviation of the differences, and n the number of paired cases. Under the null hypothesis (no mean difference), t follows a Student's t distribution.

Kruskal–Wallis Test:

The test is a non-parametric alternative to one-way ANOVA. It does not require normally distributed data, making it well-suited for psychological measures such as satisfaction or mental health scores, which often show skewness. Instead of comparing means, the test ranks all observations and compares the sum of ranks between groups.
The test statistic is:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \bar{R}_i^2 - 3(N+1)$$

where N is the total number of observations, n_i the size of group i, and $\bar{R}_i$ the mean rank of group i. Under the null hypothesis that all groups come from the same distribution, H follows a chi-square distribution with $k-1$ degrees of freedom.

Chi-Square Tests of Independence:

The Chi-Square test of independence is used to examine whether two categorical variables are associated. In this study, it was applied to test whether demographic characteristics such as sex and ethnicity were unevenly distributed across clusters. The test statistic is:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where O_{ij} is the observed frequency in cell (i, j), and E_{ij} is the expected frequency under the null hypothesis of independence. The statistic approximately follows a chi-square distribution with $(r-1)(c-1)$ degrees of freedom.

Chapter 4

Results

4.1 Descriptive Statistics and Improvement

Paired t-test: Mental Health

A paired t-test was conducted to compare Mental Health scores before and after treatment:

- Test Statistic (t): -17.56
- p-value: $< 2.2 \times 10^{-16}$
- Mean Difference: -0.33 (95% CI: [-0.37, -0.29])
- *Interpretation*: The negative t-value indicates higher post-treatment scores. The result is highly statistically significant, confirming a meaningful improvement in Mental Health post-intervention.

Improvement from Low to High Scores

Number of patients whose scores improved from 1–3 (pre-treatment) to 4–7 (post-treatment):

Table 4.1: Count of patients improving from low (1–3) to high (4–7) scores by domain

Domain	Count Improved
MentalHealth	878
PhysicalHealth	604
JobSituation	684
Accommodation	589
LeisureActivities	620
Relationship	541
Friendships	541
Safety	493

Correlation Between Domains

Correlation analysis between domains at baseline (Time 1) revealed:

Table 4.2: Pearson correlation coefficients among baseline domains

Var1	Var2	r-score
PhysicalHealth1	MentalHealth1	0.5143
LeisureActivities1	MentalHealth1	0.4982
LeisureActivities1	PhysicalHealth1	0.4830
Friendships1	LeisureActivities1	0.4800
Friendships1	Relationship1	0.4415
LeisureActivities1	JobSituation1	0.4308
Safety1	MentalHealth1	0.4225
Friendships1	MentalHealth1	0.4108

Interpretation: Baseline domains are moderately correlated (r = 0.41 to 0.51), indicating interdependence across wellbeing areas.This suggests that clustering (if present) will likely be defined by patterns across multiple domains rather than a single variable.

4.2 Gaussian Mixture Model (GMM) Clustering

Before reporting results, it is important to define poor cluster separation. Poor separation occurs when clusters overlap, meaning that people placed in different groups actually share very similar profiles. Statistically, this shows up as silhouette scores close to zero (sometimes even negative). This indicates that patient satisfaction profiles are highly individualised, continuous and do not fall into distinct groups based on demographics.

1. GMM on Pre-Treatment Scores (Time 1)

A GMM using the `mclust` package was fitted to the 8 pre-treatment domains. The **VVE** model (Variable volume, Variable shape, Equal orientation) with **6** components was selected based on BIC:

- Log-likelihood: -61093.12
- n = 6281, df = 129
- **BIC = -123314.4, ICL = -126845.9**
- The six clusters sizes: 808, 854, 1078, 767, 2236, 538

Observation: Clusters did not meaningfully separate by sex or ethnicity. This suggests that pre-treatment variation is primarily in domain scores rather than demographics.Poor separation indicates overlapping profiles among participants.

2. GMM on Post-Treatment Scores (Time 2)

For post-treatment scores, a **VVE** GMM with **6** components yielded:

- Log-likelihood: -60715.28
- **BIC = -122558.7, ICL = -125130.1**
- Cluster sizes: 748, 632, 2651, 485, 1628, 137

Observation: Similar to pre-treatment, clusters contained a mix of sexes and ethnicities.

3. GMM Default Model (All Domains)

The default GMM selected **9 clusters (VEV model)**, based on the Bayesian Information Criterion (BIC):

- Variable volume, Equal shape, Variable orientation.
- Log-likelihood:$-156{,}986.4$; BIC $= -324{,}956.9$
- Cluster sizes: 871, 687, 733, 890, 381, 622, 487, 623, 987

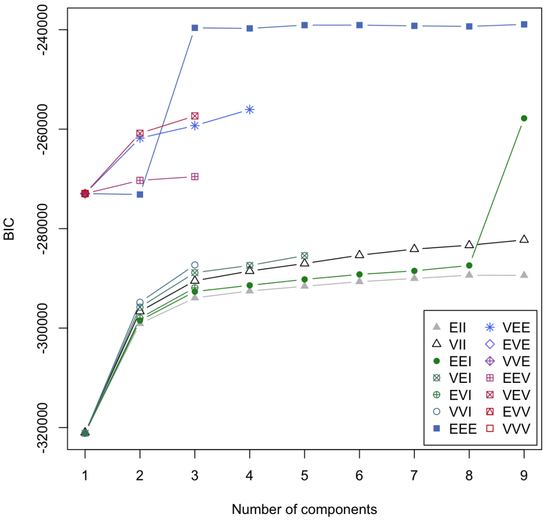


Figure 4.1: BIC values for different numbers of clusters in the GMM, showing that the 9-cluster VEV model was optimal.

Observation: The results indicate that the 9-cluster VEV model provided the best fit according to the BIC values. Examination of the cluster composition showed that both sex and ethnicity were mixed within clusters, suggesting that demographic characteristics did not determine the grouping. The clusters instead reflected variation in DIALOG domain scores, meaning that individuals were grouped based on differences in their reported experiences and satisfaction across domains. This highlights that the model captured patient-reported outcomes rather than demographic factors, offering a more meaningful basis for understanding distinct patient profiles.

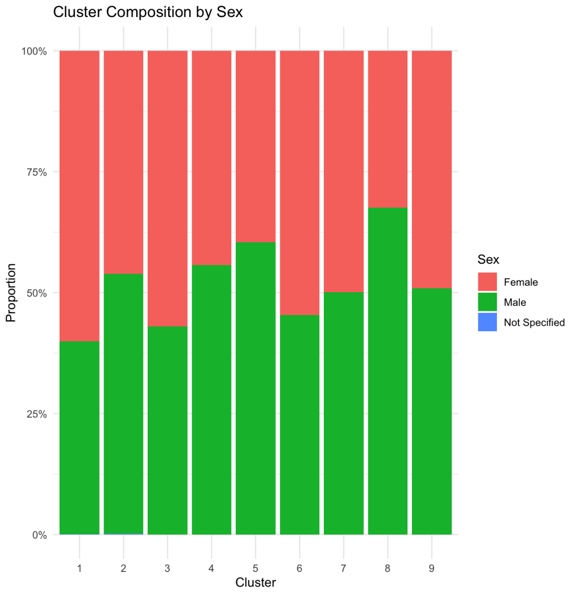


Figure 4.2: Distribution of sex across clusters.

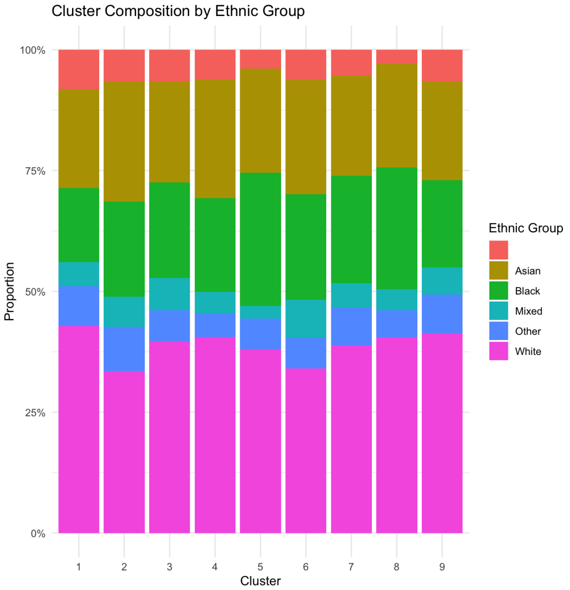


Figure 4.3: Distribution of ethnicity across clusters.

4. GMM with Increased Cluster Numbers (20 and 30)

The default `Mclust` implementation in R considers a maximum of **9 clusters** by default. In the earlier run, the optimal solution returned 9 clusters, which raised the question: was 9 selected because it was genuinely optimal, or simply because it was the maximum allowed? To investigate, the allowed number of clusters was increased to 20 and 30 to test whether more clusters would yield better model fit according to BIC.
For the **20-cluster run**, the BIC selected a **VEV** model with **19 components**.

- Log-likelihood: −141,524.6, BIC = −306,101.8, ICL = −307,662.7
- Cluster sizes ranged from **41** to **1,169**.

For the **30-cluster run**, the BIC selected a **VEV** model with **17 components**.

- Log-likelihood: −147,843.1, BIC = −316,325.2, ICL = −317,896.9
- Cluster sizes ranged from **52** to **894**.

Observation: In the expanded analyses, the clusters were unstable even with the random seed set, the best number of clusters kept changing. This shows that the 9-cluster solution from the default run was chosen by the model itself, not limited by the settings, since allowing more clusters didn't consistently improve the BIC or stability.

5. GMM with only Sex as Input

Including sex as a variable led to clear sex separation:

- 9 components selected by BIC (EEE model)
- Cluster sizes: 846, 99, 565, 110, 313, 406, 1814, 1316, 812
- Each cluster predominantly contained a single sex, but ethnicity remained mixed.

6. GMM with only Ethnicity as Input

Ethnicity inclusion resulted in partial separation:

- 9 clusters selected (VEV model)
- Cluster sizes: 222, 278, 614, 768, 334, 1212, 500, 777, 1576
- Each cluster had a mix of ethnicities, though some clusters were enriched for specific groups, sex remained mixed.

7. GMM with Sex and Ethnicity Together

When both `Sex` and `Ethnicity` variables were included in the Gaussian Mixture Model, the optimal model selected was the EEE model with $G = 9$ clusters.

Model Statistics:

- Equal volume, Equal shape, and Equal orientation
- Log-likelihood: -117972.6
- BIC: -238927.3
- Cluster sizes: 880, 955, 164, 227, 754, 667, 344, 1659, 631
- BIC Plot: Shows model selection criterion used to choose $G = 9$.

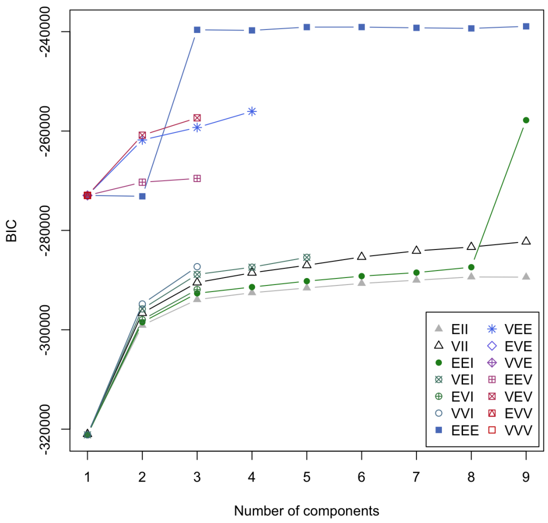


Figure: BIC values with both `Sex` and `Ethnicity`.

Clusters were examined by sex, ethnicity, age, and IMD band to understand how demographic factors aligned with the groupings.

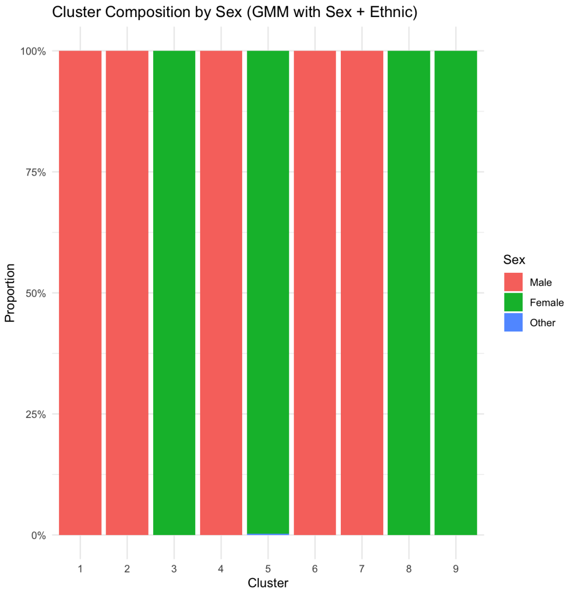


Figure 4.4: Composition by `Sex`.

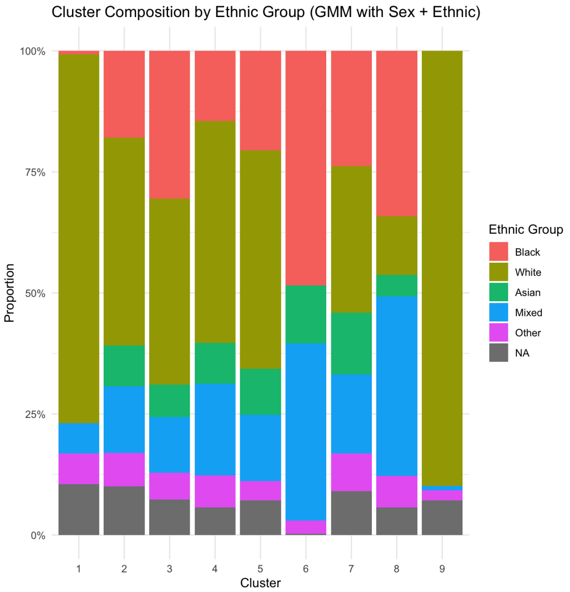


Figure 4.5: Composition by `Ethnicity`.

Observation: Adding demographic information helped separate the clusters by sex, which was clearly distinguished. For ethnicity, most clusters showed a similar distribution of ethnic groups, except for Cluster 1 and Cluster 9, where more than 75% of individuals were White. This indicates only partial separation by ethnicity, with most clusters being ethnically mixed and some overlap between clusters remaining. This suggests that the satisfaction profiles do not perfectly follow demographic groups and are influenced by a mix of domain scores and demographic factors.

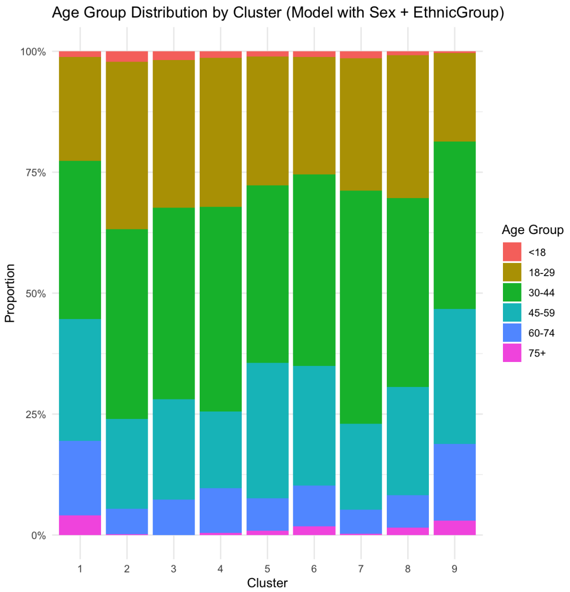


Figure 4.6: Cluster distribution by `Age`.

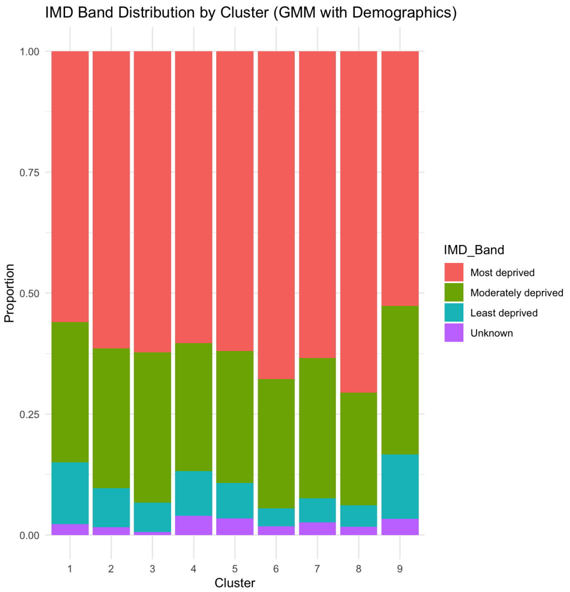


Figure 4.7: Cluster distribution by IMD.

For age, the clusters were not separated: individuals aged 18 to 44 made up roughly 25% of each cluster, and the remaining age groups had similar proportions across clusters. This even distribution suggests age had little influence on cluster formation. A similar pattern was observed for IMD, where more than half of the individuals in every cluster came from the most deprived category, again showing minimal variation.
Overall, only sex showed clear separation, ethnicity showed limited differences (with the exception of Clusters 1 and 9), and age and IMD did not contribute meaningfully to differentiation. This suggests that the clusters primarily capture variation in DIALOG domain scores rather than being driven by demographic factors.

8. Silhouette Analysis and Cluster Evaluation

For $n = 6,281$ units, silhouette analysis was conducted for 9 clusters. Average silhouette widths (0.0077) were low across all cases, indicating poor separation. Cluster-specific widths for the 9-cluster model were:

$$\{-0.0355,\ -0.0020,\ -0.0291,\ 0.0266,\ -0.0161,\ 0.0504,\ 0.0608,\ 0.0069,\ 0.0063\}$$

Observation: Several negative values and near-zero averages suggest considerable overlap, consistent with GMM results indicating complex, non-well-separated structures.
Cluster stability measured using the **Adjusted Rand Index (ARI)** via bootstrapping, was moderate for the 9-cluster model (ARI = 0.356) indicating that clusters are somewhat reproducible but not highly robust.

9. Statistical Tests

Kruskal–Wallis Test: Mental Health (Pre-Treatment)

The test was used to compare baseline mental health scores across clusters. The results showed:

$$\chi^2 = 604.77, \quad df = 6, \quad p < 2.2 \times 10^{-16}.$$

Interpretation: This highly significant result confirms that baseline mental health scores differ between clusters. The clustering is therefore not random, but captures real differences in participants' initial mental health status.

Chi-Squared Tests: Demographics vs. Cluster

Chi-squared tests were conducted to examine whether cluster membership was associated with key demographic variables, namely sex and ethnicity. The results are shown in table 4.3. Both tests indicated highly significant associations.

Test	χ^2	*df*	p-value
Sex vs. Cluster	6286.2	16	$< 2.2 \times 10^{-16}$
Ethnicity vs. Cluster	12709	40	$< 2.2 \times 10^{-16}$

Table 4.3: Results of chi-squared tests assessing associations between demographic variables and cluster membership.

Both sex and ethnicity are significantly associated with cluster membership, indicating different distributions across clusters. However, when these demographic variables were not included in the clustering process, clusters remained demographically mixed. This suggests that sex and ethnicity do not naturally define the clusters, but can influence their formation when explicitly incorporated into the model.

Summary of GMM Findings

Across multiple GMM specifications (pre-, post-treatment, and all domains), BIC consistently favoured high-component solutions (typically G = 9, VEV/EEE variants). However, silhouette widths were near zero and stability was only moderate (ARI = 0.36). Clusters were mixed by sex and ethnicity unless these were forced into the model, in which case sex separated clearly but ethnicity only partially. Allowing more clusters (up to 30) did not yield more stable solutions. Overall, GMM results show that patient satisfaction profiles reflect continuous variation in DIALOG domain scores rather than distinct, well-separated subgroups.

4.2.1 K-Means Clustering

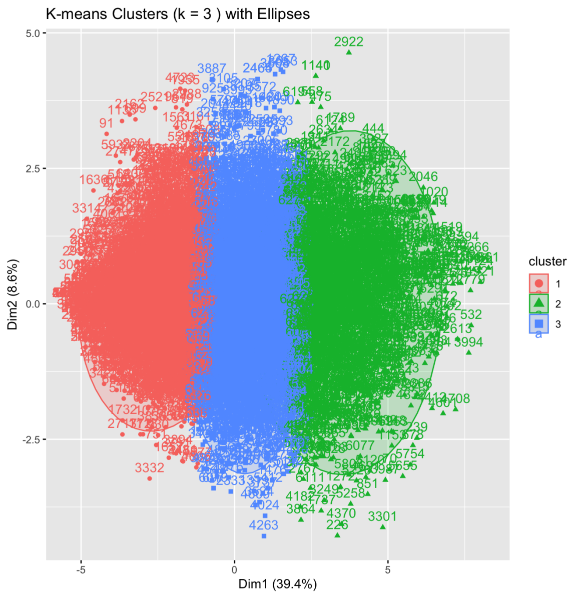
The K-Means algorithm was applied to the dataset to explore alternative clustering structures.

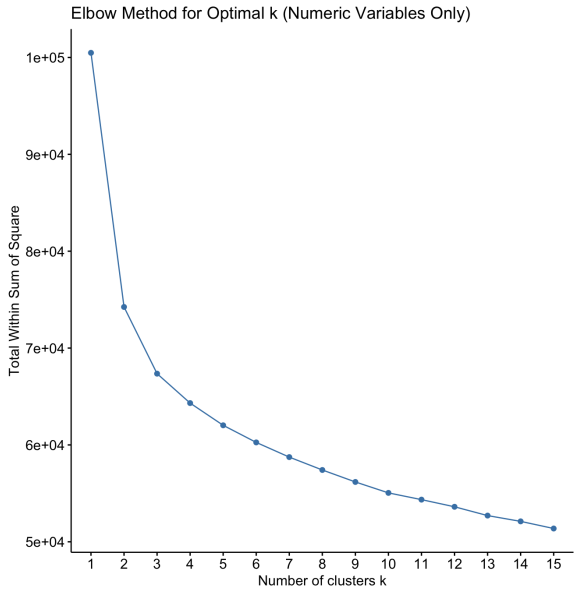
Optimal Number of Clusters

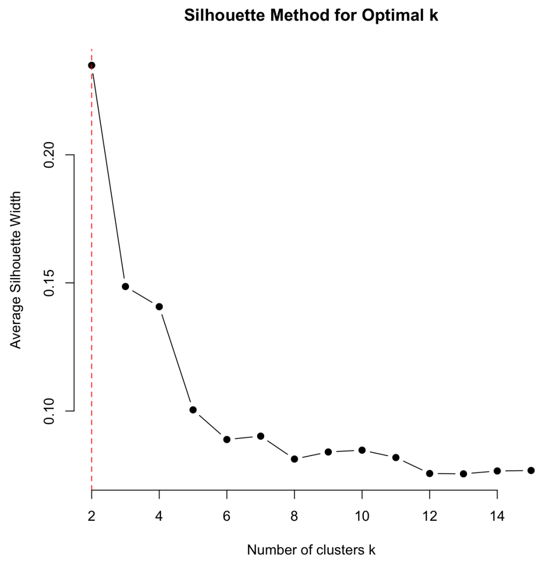
Elbow method suggested 3 clusters; **Silhouette analysis** suggested 2 clusters (Silhouette score: 0.2349). This represents moderate separation between clusters.

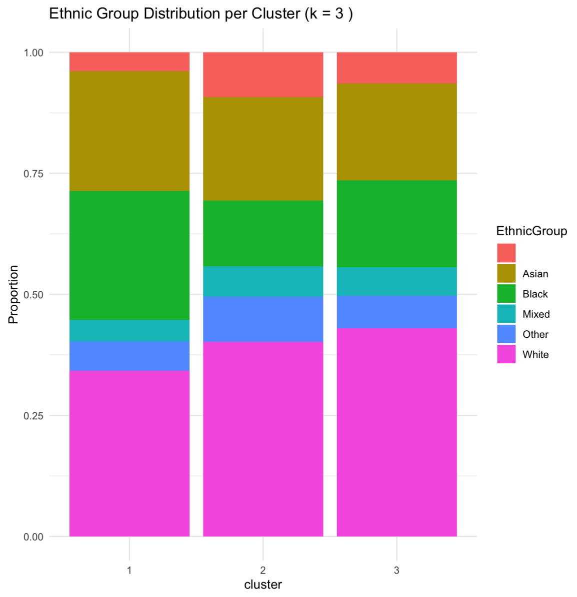
Cluster Sizes

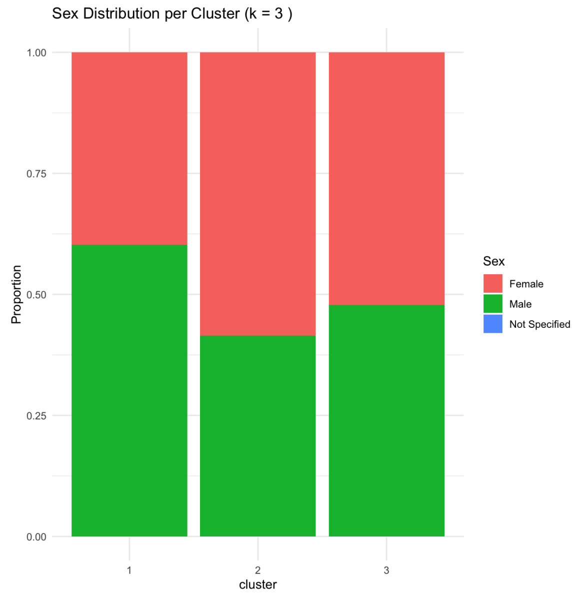
- **k = 3:** Cluster sizes = 2321, 1322, 2638. Visual inspection of the plot showed that clusters were well separated in the reduced-dimensional space and did not appear to overlap. However, demographic analysis revealed that sex and ethnicity were still mixed within each cluster.
- **k = 9:** Similar results, with mixed sex and ethnicities.











Inclusion of Demographics in Clustering

The K-Means clustering procedure was repeated with the inclusion of demographic variables (sex and ethnicity) as additional features. The Elbow method continued to indicate an optimal number of clusters at k=3, with a Silhouette score of 0.2349.

Adding demographics did not make the clusters more distinct in terms of sex or ethnicity, each cluster still had a mix of both.

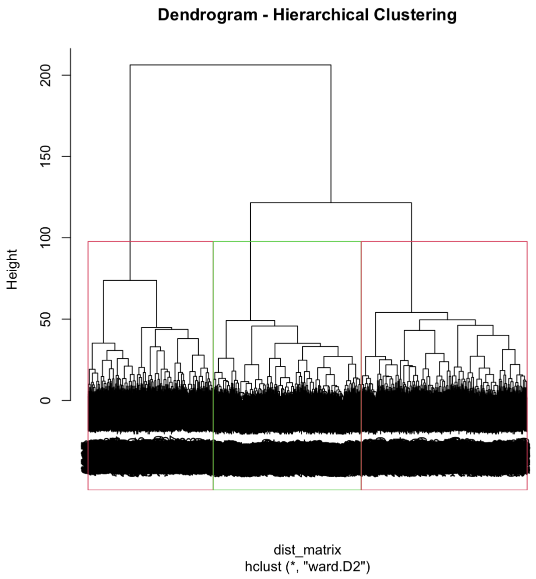
Higher k-values:

A higher number of clusters (k=9) was tested to see if smaller groups might separate demographics better. The clusters were still clearly separate when plotted, but the distribution of sex and ethnicity was the same as before. This means that K-Means could not create groups based on demographics, even when the number of clusters was changed or demographic information was added to the model.

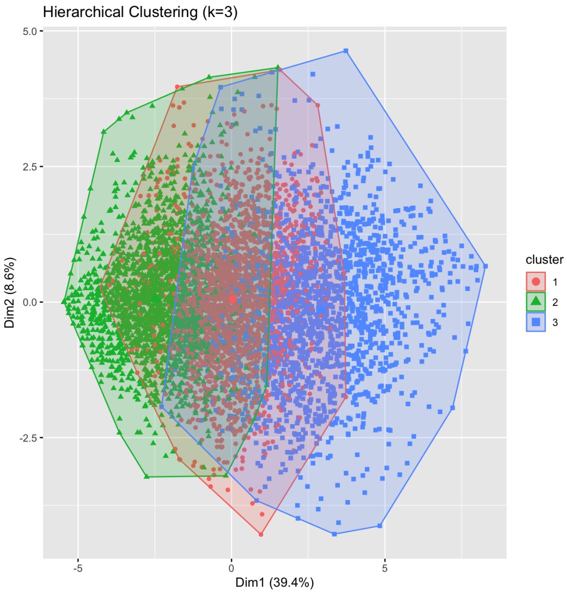
4.2.2 Hierarchical Clustering

Agglomerative hierarchical clustering was applied to the full dataset using Ward's method with Euclidean distance. Inspection of the dendrogram suggested that **three** clusters were the most suitable cut. The resulting clusters were very similar in size, composition, and demographic distribution to those obtained from K-Means. In particular, sex and ethnicity remained mixed across all clusters, and adding demographic variables or increasing the number of clusters did not change this pattern, mirroring the findings from K-Means.

Observation: Hierarchical clustering results are similar to K-Means findings, confirming that clusters reflect DIALOG score variation rather than demographic structure.



Dendrogram (Ward's Method)



Hierarchical Clustering (k = 3)

Chapter 5

Conclusion

5.1 Study Objectives

The primary aim of this project was to identify patterns in patient satisfaction and quality of life using DIALOG assessment data from the East London NHS Foundation Trust (ELFT). Specifically, the study focused on:

- Detecting latent patient subgroups based on satisfaction profiles using clustering methods (GMM, K-Means, Hierarchical Clustering).
- Assessing whether these subgroups differ by demographics (sex, ethnicity, age, IMD decile) or clinical characteristics.
- Evaluating improvements in satisfaction domains after treatment and differences in improvement across subgroups.

5.2 Summary of Key Findings

5.2.1 Overall Treatment Effects

The results showed that treatment was associated with statistically significant improvements across all eight DIALOG domains, with a very strong level of statistical confidence $p < 2.2 \times 10^{-16}$. The largest improvements were seen in mental health, personal safety, and job situation, three areas that are particularly central to patient well-being. An important observation was the shift many patients made from dissatisfaction, defined as ratings between one and three, to higher satisfaction levels, defined as scores between four and seven. To give some specific examples, 878 patients improved their mental health ratings, 604 showed improvement in physical health, and 620 reported higher satisfaction in leisure activities after treatment.

The duration of treatment also appeared to play an important role in outcomes. Patients who received treatment for between six and twelve months tended to show the greatest improvements, while those with very short or very long treatment durations demonstrated comparatively fewer gains. This suggests that there may be an optimal period of treatment in which patients benefit the most. Another interesting and somewhat unexpected finding was that patients under the Care Programme Approach (CPA) began treatment with higher baseline mental health scores than patients not under CPA.

5.2.2 Gender Differences

Gender-based differences were also evident in the data. Men tended to report slightly higher baseline satisfaction across most domains at the start of treatment. However, women demonstrated greater improvements following treatment in seven out of the eight domains. The gains for women were particularly marked in mental health, friendships, and safety. These differences were visible and also statistically significant, which indicates that men and women may respond differently to treatment and that gender could be an important factor to consider in the design and evaluation of mental health interventions.

5.2.3 Inter-Domain Correlations

The results showed that different areas of life are connected, not separate. For example, physical and mental health were linked (r = 0.5143), meaning changes in one often affect the other. The strongest links were between emotional well-being, physical health, and leisure. This shows that improving one area can also improve others, so it's important to look at the whole picture when supporting patients.

5.3 Clustering Analysis

5.3.1 Gaussian Mixture Models (GMM)

Clustering analyses provided further insight into patterns of patient satisfaction. Using Gaussian Mixture Models, the optimal solutions identified six clusters both before and after treatment. These clusters contained patients from mixed demographic backgrounds, suggesting that satisfaction profiles were not determined by demographics alone. When pre- and post-treatment scores were combined, the model identified a nine cluster solution with a VEV structure and a Bayesian Information Criterion (BIC) value of -324,956.9. This indicated that there was meaningful variation in the satisfaction profiles that could not be explained by simple categories.

Demographic factors influenced cluster formation differently. Sex was the strongest determinant of group separation, while ethnicity had a smaller, more specific effect. Age and socioeconomic status (IMD) showed minimal influence, indicating that these factors did not strongly define satisfaction patterns.

The clusters were of moderate quality. The average silhouette score was 0.0077, meaning there was quite a bit of overlap between groups, and the Adjusted Rand Index (ARI) was 0.356, showing the model was moderately stable. This overlap suggests that while the clusters reveal some meaningful differences, patient experiences are still complex and can't be neatly divided into separate groups.

5.3.2 K-Means Clustering

The K-Means analysis gave a slightly different view of the data. Using the elbow method, three clusters were identified as optimal. The silhouette score of 0.2349 indicated moderate separation, meaning the groups were clearer and had less overlap than the GMM clusters. Similar to the GMM results, each cluster contained a mix of demographic groups. This shows that patient satisfaction cannot be fully explained by age, gender, ethnicity, or social and economic background alone, and that experiences are still complex and varied.

5.3.3 Hierarchical Clustering

The hierarchical clustering analysis gave results very similar to the K-Means approach. Using Ward's method, three clusters were identified as the best solution. The clusters were roughly equal in size and included a mix of different demographic groups. Gender and ethnicity were spread across all clusters, and adding demographic information or increasing the number of clusters did not change this pattern. This confirms that the clusters mainly reflect differences in patient satisfaction scores rather than demographic factors, just like in the K-Means results.

5.3.4 Interpretation of Clustering

The clustering analyses show that patient satisfaction is complex and cannot be fully explained by simple demographic factors. Across all clustering methods (GMM, K-Means, and hierarchical), gender was the most influential factor in determining how patients were grouped. Ethnicity had a smaller effect, mainly affecting specific clusters, while age and social or economic background (IMD) had very little influence. The clusters also showed some overlap, meaning that patient experiences are not sharply divided and many people share similar satisfaction patterns across groups. This indicates that patient satisfaction is multi-dimensional, involving many factors beyond basic demographics. These findings highlight the importance of taking a personalized and careful approach when interpreting patient satisfaction data. Interventions and support strategies should consider the full range of experiences rather than assuming that demographic characteristics alone can explain differences in satisfaction.

5.4 Implications for Practice

- Optimal treatment duration is 6–12 months; very short treatment limits improvements.
- Paying attention to gender can improve outcomes, especially for women starting with lower satisfaction.
- Services should be fair for everyone, even though ethnic differences were minimal.
- CPA status should be further investigated due to unexpectedly higher baseline satisfaction.
- Interventions should focus on each patient's needs, not just demographics.

5.5 Limitations

- Self-reported data may not fully reflect objective outcomes.
- Missing data (33%) limited subgroup and demographic analyses.
- Cluster overlap indicates patient experiences cannot be perfectly categorized.
- IMD may not fully capture individual-level socioeconomic differences.

5.6 Recommendations

- Design interventions that consider gender, especially to support women who start with lower satisfaction by giving them targeted care
- Explore whether differences in outcomes come from the care provided, unmet needs, or patient expectations.
- Improve completeness of categorical data for more robust analyses.
- In future clustering, include more variables (like CPA status, age, deprivation index) to explore if any underlying cluster emerges.

- Target support to groups with lower improvement, especially in Mental Health, Accommodation, and Safety domains
- Combine Different Types of Data along with DIALOG scores, including other health records or patient surveys to get a fuller picture of satisfaction.
- Monitor patients over time to track long-term trends and identify those needing extra support early.

Bibliography

[1] Priebe, S.B. and Bird, V., 2019. *DIALOG scale – analytical framework for mental health services.* East London NHS Foundation Trust. Available at: https://www.elft.nhs.uk/sites/default/files/DIALOG%20Analytical%20Framework.pdf (Accessed: 10 June 2025).

[2] Scrucca, L., Saqr, M., López-Pernas, S. and Murphy, K., 2023. *An introduction and tutorial to model-based clustering in education via Gaussian mixture modelling. arXiv preprint.* Available at: https://arxiv.org/pdf/2306.06219 (Accessed: 24 June 2025).

[3] Cook, D. and Laa, U., 2025. *Introduction to clustering – interactively exploring high-dimensional data and models in R.* In: *Mulgar Book* [online]. Available at: https://dicook.github.io/mulgar_book/6-intro-clust.html (Accessed: 09 July 2025).

[4] Scrucca, L., Fop, M., Murphy, T.B. and Raftery, A.E., 2016. *mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. The R Journal.* Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC5096736/ (Accessed: 01 July 2025).

[5] Manoj, 2021. *Simple explanation to understand K-means clustering. Analytics Vidhya.* Available at: https://www.analyticsvidhya.com/blog/2021/02/simple-explanation-to-understand-k-means-clustering/ (Accessed: 08 August 2025).

[6] Benhur, S., 2023. *Hierarchical clustering: Agglomerative and divisive explained. Built In.* Available at: https://builtin.com/machine-learning/agglomerative-clustering (Accessed: 10 August 2025).