# LATENT STRUCTURE EXTRACTION FROM PATIENT SATISFACTION SURVEYS USING R: A CASE STUDY OF MULTI-STATE CLINICS IN THE U.S.

## 1. Background

A private network of outpatient clinics across three U.S. states—Illinois, Michigan, and Ohio—conducts annual patient satisfaction surveys with 27 Likert-scale questions. The administrative team sought to identify core areas of satisfaction to simplify reporting, support benchmarking, and guide future service investments.

## 2. Objective

- To identify underlying dimensions that summarize patient satisfaction using Exploratory Factor Analysis (EFA) in R
- To validate the structure statistically and assess reliability
- To guide the redesign of their annual feedback form based on grouped factor constructs

## 3. Data Used

**Source**: Internal patient satisfaction survey (Q4 FY2023) across 22 outpatient centers

**Sample Size**: 2,300 patient responses **Variables**: 27 Likert-scale items (1 = Strongly Disagree to 5 = Strongly Agree) **Themes Covered**: Staff behavior, appointment experience, billing, wait time, facility, follow-up, communication, provider confidence

## 4. Methodology

### 4.1 Data Preparation

- Removed incomplete records (na.omit())
- Checked for sufficient sampling adequacy using:

```
library(psych)

KMO_result <- KMO(patient_data)
```

KMO = 0.89 → **Very good** Bartlett's test $p < 0.001$ → **Suitable for factor analysis**

### 4.2 Factor Extraction

- Used psych::fa() with principal axis factoring
- Scree plot and parallel analysis suggested 4-factor solution

```
fa_result <- fa(r = patient_data, nfactors = 4, rotate = "varimax", fm = "pa")
```

**4.3 Rotation & Interpretation**

- Applied varimax for orthogonal rotation
- Factors interpreted as:
    1. **Care Experience** (e.g., confidence in doctor, listened attentively)
    2. **Operational Efficiency** (e.g., wait times, online booking ease)
    3. **Administrative Process** (e.g., billing clarity, check-in procedure)
    4. **Facility Comfort** (e.g., cleanliness, privacy, environment)

**4.4 Reliability Analysis**

- Calculated Cronbach's alpha for each extracted factor (all > 0.78)

```
alpha(patient_data[factor1_items])
```

# 5. Results Summary

| Factor Name | # Items | Cronbach's α | Sample Items |
|---|---|---|---|
| Care Experience | 9 | 0.88 | "Doctor listened", "Felt respected" |
| Operational Efficiency | 6 | 0.84 | "Easy scheduling", "Short wait time" |
| Administrative Process | 7 | 0.81 | "Billing clarity", "Check-in smooth" |
| Facility Comfort | 5 | 0.78 | "Clean lobby", "Noise-free environment" |

- Factors explained **71.3% of total variance**
- No cross-loadings > 0.3 observed → good factor distinction

# 6. Interpretation and Recommendations

- Original 27-item form can be reduced to ~12 representative items by selecting top-loading questions under each factor
- Clinics should **benchmark scores** by factor rather than item for better comparability

- Suggested aligning **staff KPIs** with “Care Experience” factor as it explains the most variance
- Recommend revisiting the structure annually using R script provided

## 7. Deliverables

- **Factor Analysis Report**: PDF (26 pages) with:
  - Scree plots, loading matrices, alpha tables
  - Explanation of extraction method
  - Factor-to-item mappings
- **Code Modules**:
  - clean_survey_data.R
  - efa_analysis.R
  - reduced_form_generator.R
- **Optional Output**:
  - R Shiny prototype for uploading new survey files and auto-generating a 4-factor summary dashboard

## 8. Business Outcome

- Admin team adopted a **reduced 12-question version** of the form
- Insights led to **21% improvement** in net promoter score in the following quarter
- Facility-level scorecards created using factor outputs
- R code now reused by internal analytics team for biannual reviews