Final Project

Wilson Hernandez & Britte van Tiem

Contents

Contents

Objectives

1 Executive Summary

2 Background

Prison climate is widely thought to relate directly to the rehabilitation and reintegration of offenders, reflecting the broader idea that positive institutional climates improve outcomes. Over the past fifty years, researchers have made efforts to identify 'what matters' in correctional environments, and to develop and improve survey instruments that measure incarcerated people's perceptions and experience of the conditions in which they are confined. Prison climate surveys can serve to gauge how the quality of prison life differs across institutions, and be used to answer more difficult questions about whether and why those differences matter.

2.1 Description of Data

The Prison Climate Questionnaire (PCQ) is a self-report instrument developed to measure the perceived quality of prison life as part of the Dutch Life in Custody study. It was designed by researchers at Leiden University and supported by the Dutch prison service, and aims to both help the prison service in monitoring prison performance and to facilitate academic research on the quality of prison life. The PCQ measures six primary domains of prison climate (relationships in prison, safety and order, contact with the outside world, facilities, meaningful activities and autonomy) that are measured using 14 scales and 64 items. Each of the scales are composed of 3-8 items that are rated on a five-point Likert scale, with responses ranging from "strongly disagree" to "strongly agree." In addition, as an overall measure of satisfaction with the institution, the PCQ includes a one-item overall institution rating ("Generally speaking, I am satisfied with this institution") and a 3-item scale that captures how individuals experience the severity of their sentence. The survey also asks about individual characteristics not readily available through administrative data.

In early 2022, a Pennsylvania-based research team adapted the PCQ for use in the State Correctional Institution Chester in Pennsylvania (USA), a medium-security institution detaining only males.

SCI Chester held 972 individuals during data collection in the spring of 2022, spread across 14 housing units. Researchers went door-to-door during the afternoon count, when the large majority of individuals were in their cells. Every individual in their cell was given a copy of the pencil-and-paper instrument to complete while in their cell. Individuals provided their ID number on the survey to enable the payment of compensation and to facilitate administrative data linking. A total of 641 respondents completed a survey. Survey respondents provided their ID and survey data was linked to administrative data for all but twelve individuals who chose to participate anonymously. Furthermore, ten respondents had more than 10 missing answers on the prison climate items. Dropping also these 10 individuals in the analysis, leaves us with 627 respondents, amounting to 65% of the total prison population. For these 627 respondents, we imputed values for missing data using the EMB algorithm which imputes missing data from the joint distribution of all survey variables (see Annex A).

2.2 A note on Likert Scale Data

Likert scale data provide ordinal variables which are to be aggregated to scale scores for further analysis. This has implications for imputation strategies, and analysis methods.

3 Exploratory Data Analysis

The dataset we are using has data for several survey waves. For the purpose of this analysis, we work with just the data from the first survey wave. Some individuals completed more than one survey. Only the first survey for those individuals is retained.

```
load("data/processed/pcq_for_stat571.Rda")
pcq_lookup <- read_xlsx("data/raw/230207_pcq_survey_questions_NL_PA.xlsx")

pcq_lookup <- pcq_lookup %>%
  mutate(question_qno = paste0("q", question_no_pa_2022a), .after = question_no_pa_2022a)
pcq_lookup <- as.data.frame(pcq_lookup)

# Subset data to analysis based on wave 1 alone
pcq <- pcq[which(pcq$survey_wave==1),]

# Because this has just 1 individual in there, generates issues with regressions
pcq <- pcq[pcq$unit_type!="inf",]

# Delete individuals who completed the survey for a second time within wave one.
pcq$survey_no <- 1
temp <- data.frame(table(pcq$research_id))
index <- temp[temp$Freq==2,"Var1"]
pcq[pcq$research_id %in% index & pcq$block!="a","survey_no"] <- 2 # all these individuals moved from an
pcq <- pcq[pcq$survey_no==1,] # Drops the second survey of these 3 individuals.

# save full dataset
pcq_full <- pcq

# Retain only questions needed for psychometrics analysis
retain <- pcq_lookup[which(pcq_lookup$include_comparative_psych_analysis=="yes"),]$question_qno
retain <- retain[which(retain %ni% paste0("q", c(1,7,8,9)))]
pcq <- pcq[,retain]
```

```
# pcq & pcq2 - to distinguish responses that are complete but may include 'no opinion' or 'not applicab
# pcq - 999 is set to NA, no opinion = 111, not applicable = 996
# pcq2 - 999, 111, and 996 are all set to NA

pcq2 <- pcq
pcq2 <- pcq2 %>% replace_with_na_all(condition = ~.x == 111)
pcq2 <- pcq2 %>% replace_with_na_all(condition = ~.x == 996)
```

3.1 Missing data

'No opinion' answer options were provided in 23 out of 64 items on the survey. 511 out of 641 respondents (79.7%) answered all items on the prison climate scales of the PCQ. This number includes individuals who answered 'no opinion' on at least one of the items where this was an answer option but excludes individuals who left one or more questions blank or had one or more illegible answers. We will hereafter refer to these as complete questionnaires. Of the 130 individuals with incomplete questionnaires, three individuals left all prison climate items on the survey blank. Another seven individuals had more than 10 missing answers. We drop these 10 individuals in this analysis, leaving us with 631 respondents.

Table X in the annex tabulates missingness by item, as well as no opinion answers. A particularly large number of respondents selected 'no opinion' on the questions concerning visits, which is likely to reflect the fact that these individuals do not receive visits. Other individual items with high 'no opinion' prevalence are items asking about an individual's satisfaction with religious services, the dentist and the psychologist. This too, is likely to reflect the fact that individuals have not used these services. The researchers mistakenly left out 'no opinion' options for items in the frequency of contact scale, and this is reflected in disproportionate missingness on questions that ask about individual's satisfaction with how often they can see their children and lawyer.

```
# Missing data table
tab <- data.frame(question = pcq_lookup[pcq_lookup$question_qno %in% names(pcq[,1:ncol(pcq)]),"question_
                  question_qno = pcq_lookup[pcq_lookup$question_qno %in% names(pcq[,1:ncol(pcq)]),"quest
                  domain = pcq_lookup[pcq_lookup$question_qno %in% names(pcq[,1:ncol(pcq)]),"domain_no_l
                  scale = pcq_lookup[pcq_lookup$question_qno %in% names(pcq[,1:ncol(pcq)]),"scale_theory
                  missing = NA,
                  no_opinion = NA)
for(i in names(pcq[which(rowSums(is.na(pcq))<=10),1:ncol(pcq)])){
  tab[tab$question_qno==i,"missing"] <- length(which(is.na(pcq[,i])))
  tab[tab$question_qno==i,"no_opinion"] <- length(which(pcq[,i]==111))
}
tab <- tab[order(tab$domain),c("question", "missing", "no_opinion")]
names(tab) <- c("Item","Missing", "No Opinion")
print(xtable(tab, digits=c(0,0,0,0), caption="Missingness by Item"), include.rownames = FALSE, type="htm
```

Missingness by Item

Item

Missing

No Opinion

Incarcerated people on this unit treat each other respectfully

6

0

Incarcerated people on this unit are quickly accepted into the group

5

0

Incarcerated people on this unit are considerate of each other

6

0

Incarcerated people on this unit get along with each other

6

0

Incarcerated people on this unit help and support each other

5

0

Staff on this unit help me if I have problems

5

0

Staff on this unit are kind to me

4

0

Staff on this unit are there to talk to if I feel worried or sad

7

0

Staff on this unit motivate and encourage me to participate in activities

6

0

Staff on this unit treat me fairly

7

0

Staff on this unit explain their decisions to me

9

0

Staff on this unit treat me with respect

6

0

Staff on this unit give me a chance to express my views before they make decisions

7

0

I feel safe in this institution

10

0

I sometimes feel threatened by incarcerated people

7

0

There are places in this building where I feel unsafe

11

0

I am afraid of some incarcerated people

10

0

I am afraid of some staff

12

0

The visiting room is pleasant

11

164

My visitor and I can have enough physical contact during visits

13

156

The visiting hours are long enough

12

138

I have sufficient privacy during visiting hours

10

141

The staff treat my visitors nicely

12

163

The visiting hours are frequent enough

13

137

I enjoy receiving visits

13

120

After receiving a visitor, I feel good

11

120

I am satisfied with how often… I can see my family, friends or partner here

6

0

I am satisfied with how often… I can see my child(ren) here

40

0

I am satisfied with how often… I can see my laywer here

35

0

My sleep is often restless

8

0

My sleep is often disturbed

9

0

Due to poor conditions in this institution and/or my cell, I can't sleep well

11

0

I can get medical care here if I want to

8

0

Health problems are being taken care of adequately here

9

0

I am satisfied with the work of the nurse

4

34

I am satisfied with the work of the doctor

5

32

I am satisfied with the work of the dentist

8

46

I am satisfied with the work of the psychologist

6

87

I am satisfied with the range of products in the commissary

9

0

I am satisfied with the prices in the commissary

10

0

I am satisfied with the quality of the products in the commissary

7

0

I am satisfied with the recreation activities

9

11

I am satisfied with the sports

6

27

I am satisfied with the library

7

43

I am satisfied with my work in this institution

10

66

I am satisfied with the education/courses

7

51

I am satisfied with the outdoor activity

7

26

I am satisfied with the religious services

12

73

The daily program is interesting enough

9

0

I learn useful skills here

7

0

I have enough to do here

8

0

The activities here help me to develop myself

9

0

On this unit, I can prepare well for my return into society

7

9

On this unit staff encourage me to make plans for after release

8

13

On this unit I can get extra support to prepare for my return to society

6

11

On this unit I can learn things that help me to stay away from crime after release

9

12

On this unit there is much I can decide for myself

8

0

On this unit I can decide for myself on matters that are important to me

8

0

On this unit I am encouraged to arrange matters myself

15

0

On this unit I have enough freedom of movement

10

0

3.1.1 Patterns in missingness

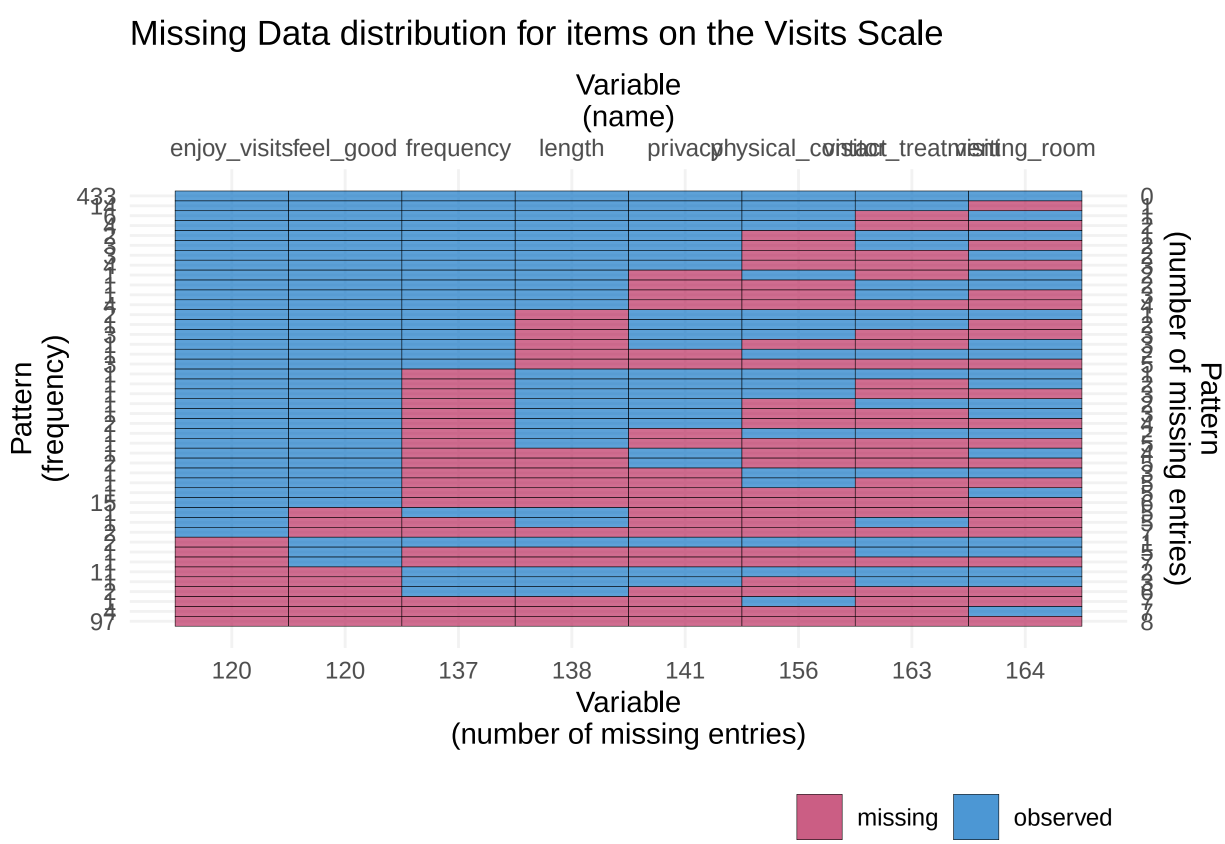
We explore “missingness”no opinion” patterns for each of the scales that have high missingness and discuss their patterns using plots to visualize missing data patterns. On the top x-axis, we the variable name. These have been shortened from the full item. On the bottom x-axis, we observe how often this variable is missing. The left y-axis indicates how often a particular missing data pattern is observed, and the right x-axis indicates how many variables are missing.

The patterns in “no opinion” patterns align with expectations about how often individuals use certain services. Individuals indicate that they have ‘no opinion’ much more frequently for types of activities that not all incarcerated indidivuals participate in. For example, many individuals do not access religious services, and many do not work or participate in education. In contrast, outdoor activities and recreation activities are available to all. This is where we observe much less use of the “no opinion” answer.

Notably, only about two thirds of respondents answer all questions on the visits scale. 15% of respondents (97 individuals) answered “no opinion” on every single visits question. This is likely explained by the fact that these individuals received no visits. Note also that individuals are more likely to respond to questions about how they feel after a visit (“I enjoy receiving visits” and “After receiving a visit, I feel good”) and about the frequency and length of the visiting hours, than about the nature of the visiting room or how their visitors were treated. This, too, makes intuitive sense: people may know how they feel after being visited, or have something to say about how frequent the visiting hours are, even if they have never set foot in the visiting room.

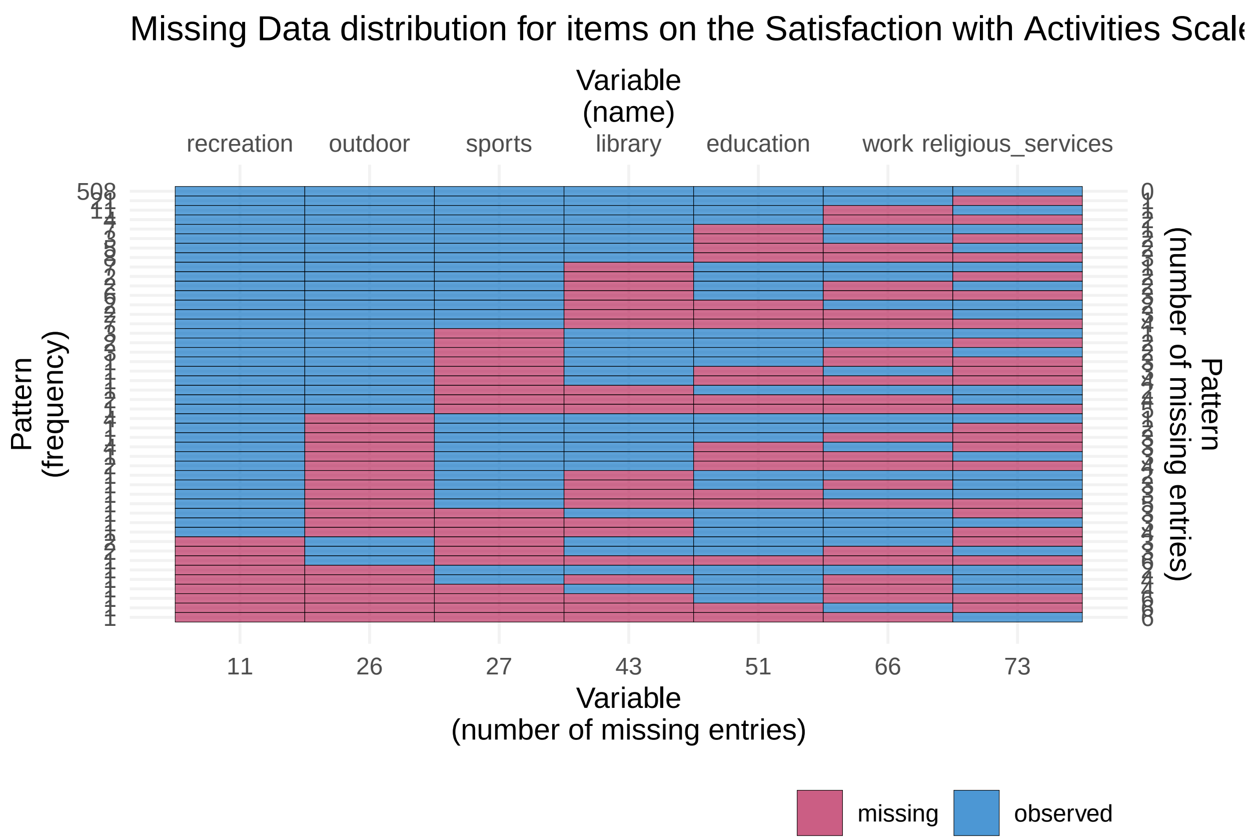
```r
pcq3 <- pcq
pcq3 <- as.data.frame(pcq3)
for(i in names(pcq3)){
k <- which(is.na(pcq3[,i]))
pcq3[k,i]  <- 999
b <- which(pcq3[,i]==111)
pcq3[b,i] <- NA
}

# Visits
vars <- paste0("q", seq(139,146,1))
temp <- pcq3[,vars]
names(temp) <- c("visiting_room", "physical_contact", "length", "privacy","visitor_treatment", "frequen
# https://amices.org/ggmice/reference/plot_pattern.html
temp <- as_tibble(temp)
plot_pattern(
  temp,
  vrb = "all") +
  ggtitle("Missing Data distribution for items on the Visits Scale")
```

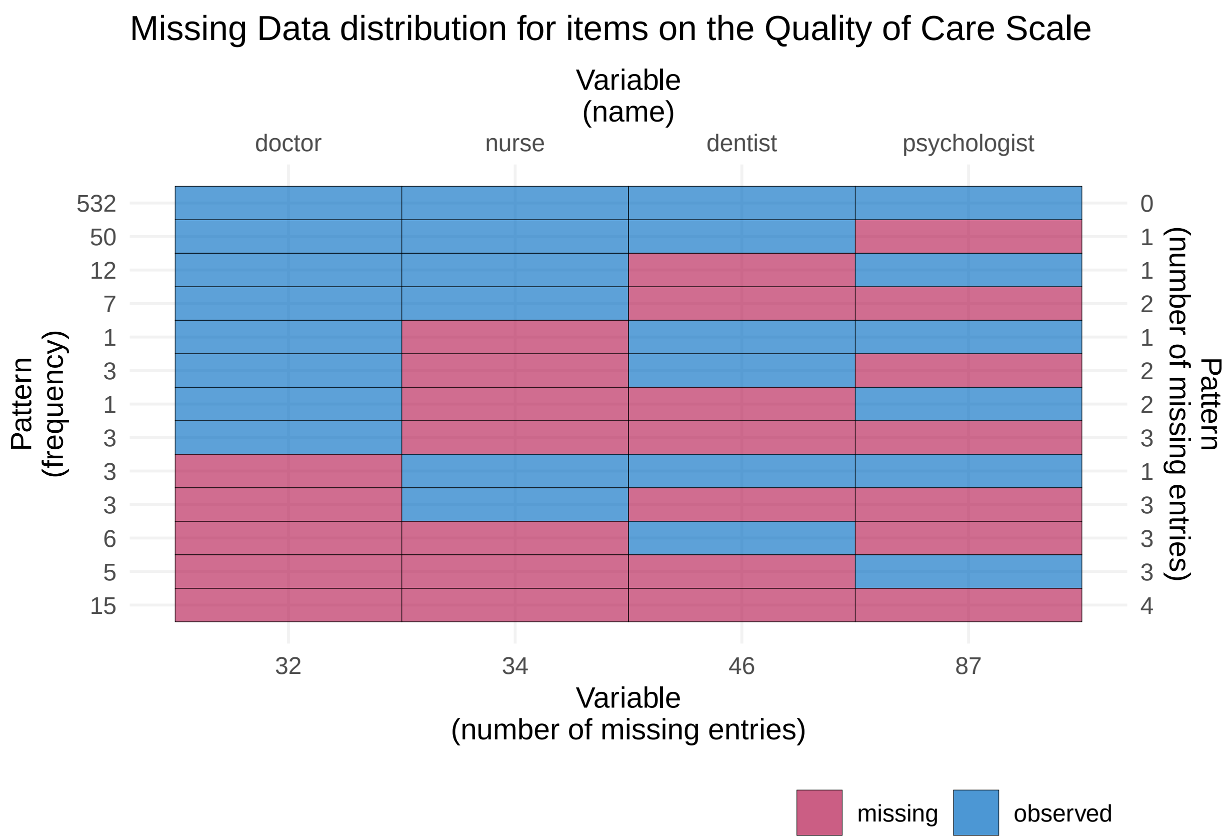



```r
# Satisfaction with Activities
vars <- paste0("q", seq(62,68,1))
temp <- pcq3[,vars]
names(temp) <- c("recreation", "sports", "library", "work", "education", "outdoor", "religious_services
# https://amices.org/ggmice/reference/plot_pattern.html
```

```
plot_pattern(
  temp,
  vrb = "all") +
  ggtitle("Missing Data distribution for items on the Satisfaction with Activities Scale")
```



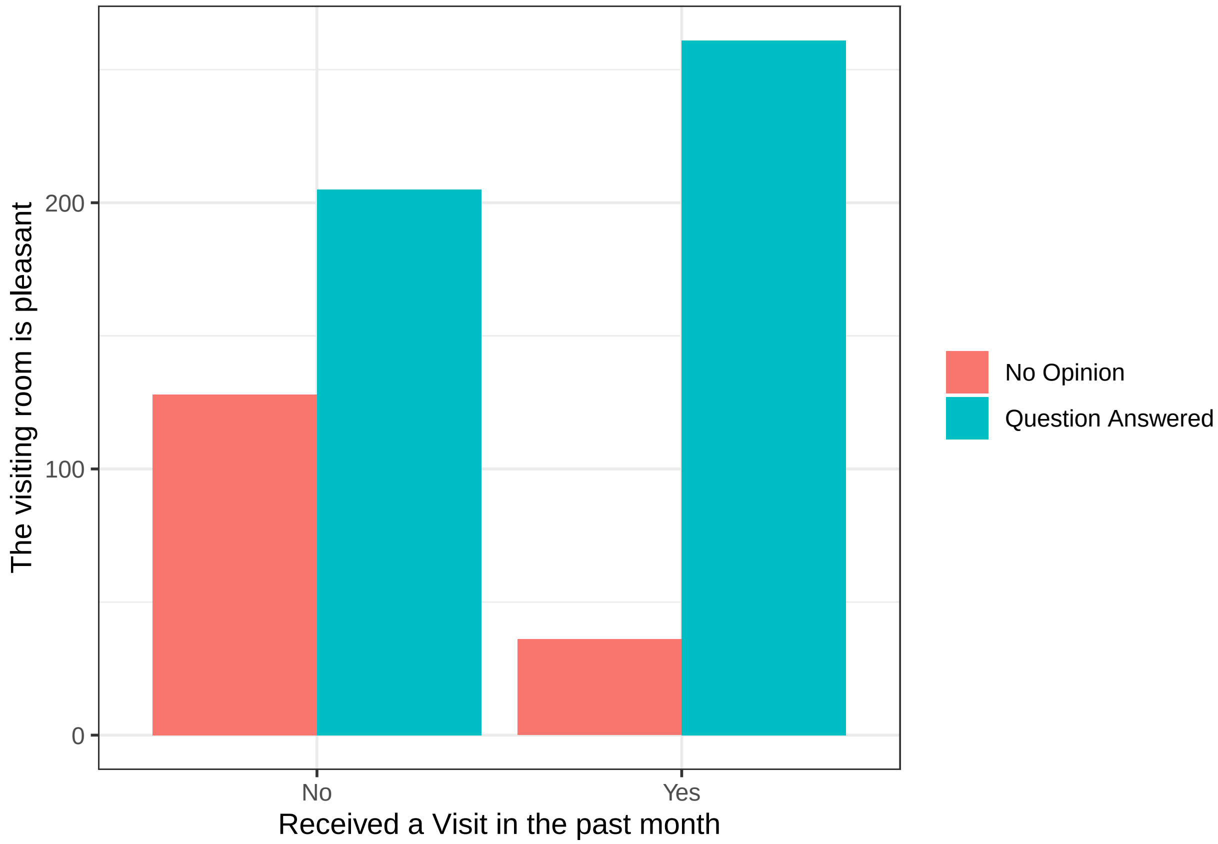

```
# Quality of Care
# Note: q111 and 112 also belong to this scale (on access and adequate care) but they do not have a no
vars <- paste0("q", seq(119,122,1))
temp <- pcq3[,vars]
names(temp) <- c("doctor", "nurse", "dentist", "psychologist")
# https://amices.org/ggmice/reference/plot_pattern.html
plot_pattern(
  temp,
  vrb = "all") +
  ggtitle("Missing Data distribution for items on the Quality of Care Scale")
```



The survey includes questions about individuals' service use in the past month. The barplots below show that the likelihood of 'no opinion' responses are indeed related to use of these services in the past month. Formal tests show that these relationships are highly significant.

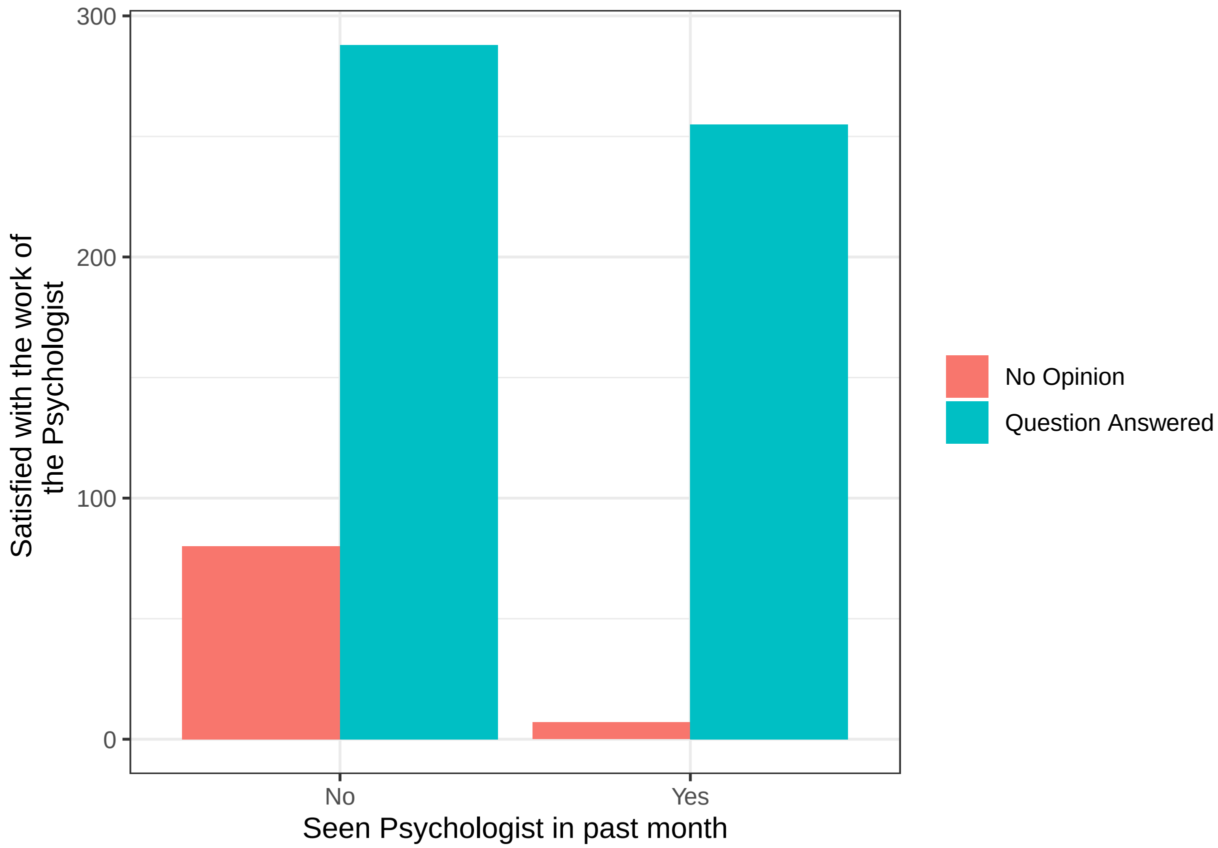
```
# Visits - visiting room is pleasant
pcq_full$q138_dummy <- ifelse(pcq_full$q138=="1",0,1)
pcq_full$q139_dummy <- ifelse(pcq_full$q139=="111", 1,0)

pcq_full$q139_dummy <- ifelse(pcq_full$q139_dummy=="1", "No Opinion", "Question Answered")
pcq_full$q138_dummy <- ifelse(pcq_full$q138_dummy=="1", "Yes", "No")
ggplot(pcq_full[which(!is.na(pcq_full$q138_dummy) & !is.na(pcq_full$q139_dummy)),], aes(x = as.factor(q
  xlab("Received a Visit in the past month") +
  ylab("The visiting room is pleasant") +
  theme_bw() +
  theme(legend.title=element_blank())
```



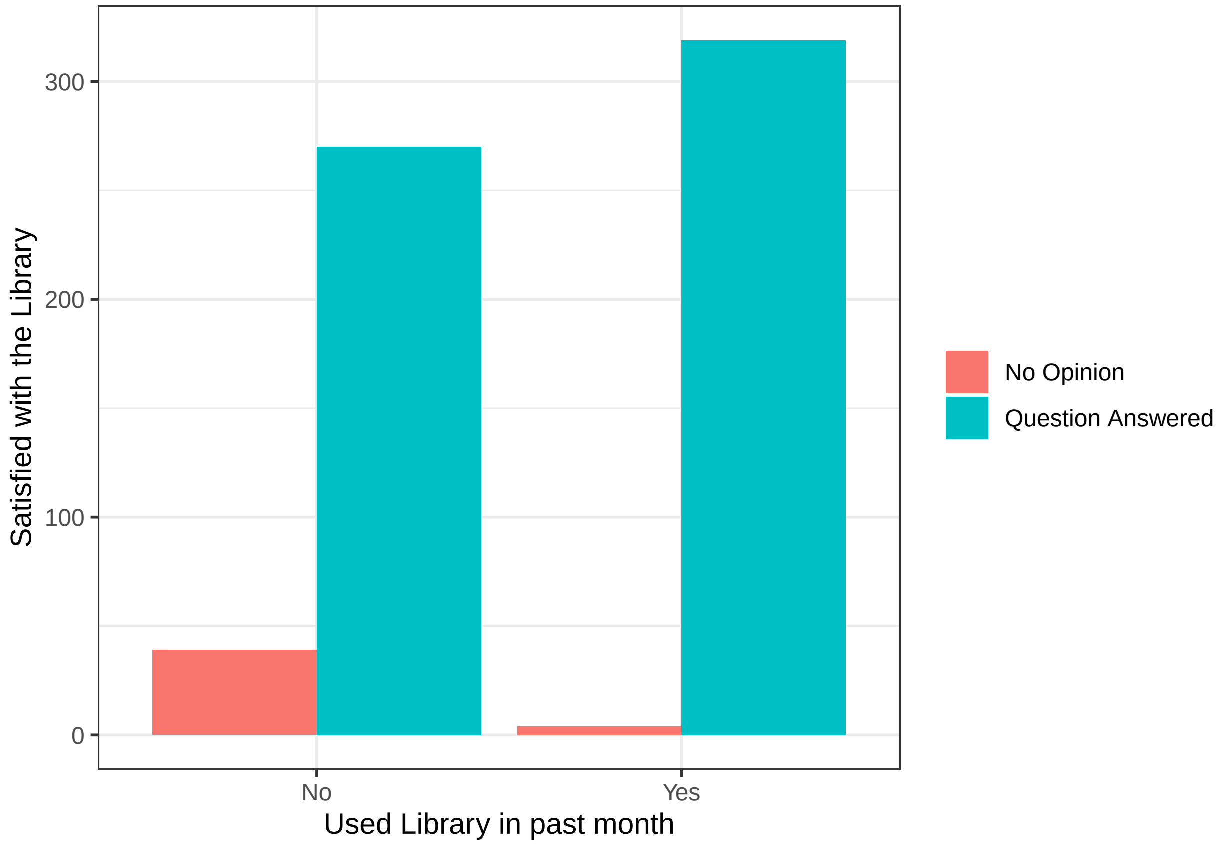

```
# Psychologist
pcq_full$q118_dummy <- ifelse(pcq_full$q118=="1", 0,1)
pcq_full$q122_dummy <- ifelse(pcq_full$q122=="111", 1,0)

pcq_full$q122_dummy <- ifelse(pcq_full$q122_dummy=="1", "No Opinion", "Question Answered")
pcq_full$q118_dummy <- ifelse(pcq_full$q118_dummy=="1", "Yes", "No")
ggplot(pcq_full[which(!is.na(pcq_full$q118_dummy) & !is.na(pcq_full$q122_dummy)),], aes(x = as.factor(q
  geom_bar(position="dodge") +
  xlab("Seen Psychologist in past month") +
  ylab("Satisfied with the work of \n the Psychologist")+
  theme_bw()+
  theme(legend.title=element_blank())
```




```
# Library
pcq_full$q57_dummy <- ifelse(pcq_full$q57=="1", 0,1)
pcq_full$q64_dummy <- ifelse(pcq_full$q64=="111", 1,0)

pcq_full$q64_dummy <- ifelse(pcq_full$q64_dummy=="1", "No Opinion", "Question Answered")
pcq_full$q57_dummy <- ifelse(pcq_full$q57_dummy=="1", "Yes", "No")
ggplot(pcq_full[which(!is.na(pcq_full$q57_dummy) & !is.na(pcq_full$q64_dummy)),], aes(x = as.factor(q57_
  geom_bar(position="dodge") +
  xlab("Used Library in past month") +
  ylab("Satisfied with the Library")+
  theme_bw()+
  theme(legend.title=element_blank())
```



3.1.2 Solutions to missingness

In most cases where someone answered “no opinion”, a variable is not defined because it has no meaning. Just like a question about “marital happiness” has no meaning for people who are not married, “I am satisfied with the library” has no meaning for people who have never visited the library. Most data imputation methods are designed for situations in which a real value is missing. In cases where the library was never used, there is no “real” value to be imputed because the data does not exist.

This creates challenges for our factor analysis.We therefore generate several data sets on which we will do our factor analysis separately

1. Dataset 1: All items except “I am satisfied with how often I can see my children here”, because this item by definition cannot be answered by a substantial subset of individuals. It retains only complete cases. This dataset is automatically the smallest because it only includes individuals who answered all available questions. (N=280, V=59)
2. Dataset 2: All items that belong to scales without ‘no opinion’ answers. Less than one percent of the data on these scales is missing. We impute this missing data using a random forest model. (N = 631, V=36)
3. Datasets 3-5: Three datasets for the quality of care, visits, and activity scales separately. It retains only complete cases within each scale. (N=516, V=6 / N=394, V=10 / N=487, V=7)

As the author of the psych library in R (William Revelle) writes, “Extracting interpretable factors means that the number of factors reflects the investigators creativity more than the data.” We will work from the perspective that any differences between factor analyses (model 1 vs model 2, and model 1 vs models 3-5) are informative with respect to margins for model adjustment.

4 Factor Analysis

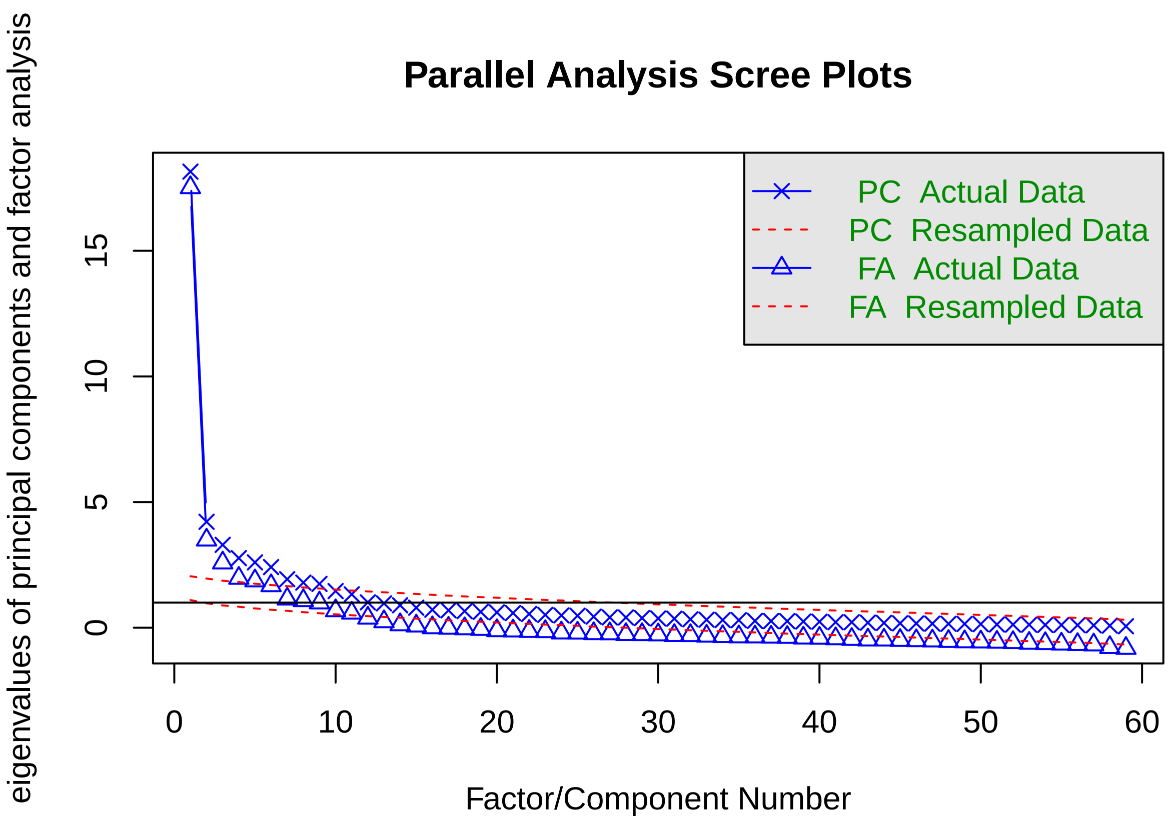
4.1 Identify number of factors in the data

After checking that the data are suitable for factor analysis by calculating the Kaiser–Meyer–Olkin statistic and Bartlett's test of Sphericity, we obtained an estimate of the number of factors in the data through a parallel analysis procedure.

There are multiple ways to determine the appropriate number of factors in exploratory factor analysis. One such way is to do a "parallel" analysis that compares the scree plot of factors of the observed data with that of a random data matrix of the same size as the original (Revelle R). We conduct this analysis using polychoric correlations, a technique for estimating the correlation between two hypothesised normally distributed continuous latent variables, from two observed ordinal variables.

Below, we display the scree plot for dataset 1. In the below table, we also provide the number of factors identified in the parallel analysis.

```
# Include plot for illustrative purposes
parallel1cc <- fa.parallel(pcq2_cc, fm="pa", fa="both", sim=FALSE) # nfactors = 11
```




```
## Parallel analysis suggests that the number of factors =  11  and the number of components =  9
```

```
tab <- data.frame(model = c(1,2,3,4,5),
                  observations = c(276,631,516,412,487),
                  variables = c(60,36,6,9,7),
```

```r
                  factors = c(11,8,2,3,2))
knitr::kable(tab)
```

model	observations	variables	factors
1	276	60	11
2	631	36	8
3	516	6	2
4	412	9	3
5	487	7	2

4.2 Exploratory factor analysis

After identifying the number of factors in the data, we run an exploratory factor analysis for each dataset.

Exploratory factor analysis is based on the common factor model, where each observed variable is a linear function of one or more common underlying latent variables ('factors') and one unique factor (i.e., error- or item-specific information). It partitions item variance into two components: (1) Common variance, which is accounted for by underlying latent factors, and (2) unique variance, which is a combination of item-specic reliable variance and random error. Exploratory factor analysis is a data-driven (rather than model-driven) approach to identifying the underlying latent variables.

The researchers who designed the PCQ had theoretical hypotheses about the factors underlying the data. We start by ignoring these presumptions, as an EFA is best positioned to tell us how closely the data aligns with the theory.

An exploratory factor analysis requires the researcher to specify the number of factors in the data. We conduct the explorartory factor analys with the number of factors as identified in the parallel analysis. In the table, the first column indicates the factor that the researchers the researchers hypothesized the item would belong to. Columns F1, F2, and F3-5 contain the factors that the items loaded onto in the exploratory factor analysis using each of the different datasets. Column F3-5 contains the results of factor analyses with datasets 3-5. The accompanying L-columns contain the loadings on each factor.

In both datasets, the staff relationships and procedural justice items load onto one scale. In the factor analysis with dataset 1, several of the 'availability of activities' items load onto the reintegration factor. In fact, several items double-load onto both the reintegration and the satisfaction with activities scales. The item "The daily program is interesting enough" is the only item that has a higher loading on the satisfaction with activities scale. In the dataset with complete cases for the no opinion scales, however, we see that the items load onto the two factors as originally expected.

The two items about how individuals feel split off onto a separate factor both when using the full dataset and when using complete cases on visits only. Two items about satisfaction with frequency of contact load onto the visits scale, but with low factor loadings. They split off when using complete cases on visits only.

Note that for the smaller datasets in columns F3-5/L3-5, the parallel analysis tends to identify more factors. For example, in the visits scale, the items about the frequency and length of the visiting hours load onto a separate factor. Similarly, items about the library, religious services, work and education split off from the other items. It is notable that these are the items on which we know there is more missingness because not everyone participates in these activities. We run this analysis again with just 1 factor per hypothesized scale to examine the loadings. Results now more closely resemble those of the full dataset. The factor loadings generally appear acceptable.

```r
knitr::kable(keep)
```

Item	Theory	F1	L1	F2	L2	F3-5	L3-5	F3-5 (2)	L3-5 (2)
1	Incarcerated people on this unit treat each other respectfully	prisoners	0.839	prisoners	0.839	NA	NA	NA	NA
1	Incarcerated people on this unit are quickly accepted into the group	prisoners	0.733	prisoners	0.719	NA	NA	NA	NA
1	Incarcerated people on this unit are considerate of each other	prisoners	0.833	prisoners	0.828	NA	NA	NA	NA
1	Incarcerated people on this unit get along with each other	prisoners	0.808	prisoners	0.794	NA	NA	NA	NA
1	Incarcerated people on this unit help and support each other	prisoners	0.829	prisoners	0.770	NA	NA	NA	NA
2	Staff on this unit help me if I have problems	procedure	0.742	procedure	0.747	NA	NA	NA	NA
2	Staff on this unit are kind to me	procedure	0.824	procedure	0.831	NA	NA	NA	NA
2	Staff on this unit are there to talk to if I feel worried or sad	procedure	0.838	procedure	0.820	NA	NA	NA	NA
2	Staff on this unit motivate and encourage me to participate in activities	procedure	0.739	procedure	0.691	NA	NA	NA	NA
3	Staff on this unit treat me fairly	procedure	0.891	procedure	0.879	NA	NA	NA	NA
3	Staff on this unit explain their decisions to me	procedure	0.857	procedure	0.798	NA	NA	NA	NA
3	Staff on this unit treat me with respect	procedure	0.826	procedure	0.871	NA	NA	NA	NA
3	Staff on this unit give me a chance to express my views before they make decisions	procedure	0.783	procedure	0.792	NA	NA	NA	NA
4	I feel safe in this institution	safety	0.281	safety	0.350	NA	NA	NA	NA
4	I sometimes feel threatened by incarcerated people	safety	0.797	safety	0.780	NA	NA	NA	NA
4	There are places in this building where I feel unsafe	safety	0.701	safety	0.736	NA	NA	NA	NA
4	I am afraid of some incarcerated people	safety	0.882	safety	0.870	NA	NA	NA	NA
4	I am afraid of some staff	safety	0.626	safety	0.506	NA	NA	NA	NA
7	My sleep is often restless	sleep	0.744	sleep	0.799	NA	NA	NA	NA
7	My sleep is often disturbed	sleep	0.815	sleep	0.834	NA	NA	NA	NA
7	Due to poor conditions in this institution and/or my cell, I can't sleep well	sleep	0.598	sleep	0.579	NA	NA	NA	NA
9	I am satisfied with the range of products in the commissary	shop	0.712	shop	0.761	NA	NA	NA	NA
9	I am satisfied with the prices in the commissary	shop	0.845	shop	0.848	NA	NA	NA	NA
9	I am satisfied with the quality of the products in the commissary	shop	0.811	shop	0.873	NA	NA	NA	NA
11	The daily program is interesting enough	actsat	0.391	actav	0.700	NA	NA	NA	NA
11	I learn useful skills here	reint	0.545	actav	0.858	NA	NA	NA	NA
11	I have enough to do here	reint	0.485	actav	0.730	NA	NA	NA	NA
11	The activities here help me to develop myself	reint	0.534	actav	0.838	NA	NA	NA	NA
12	On this unit, I can prepare well for my return into society	reint	0.744	reint	0.659	NA	NA	NA	NA
12	On this unit staff encourage me to make plans for after release	reint	0.642	reint	0.726	NA	NA	NA	NA
12	On this unit I can get extra support to prepare for my return to society	reint	0.753	reint	0.873	NA	NA	NA	NA
12	On this unit I can learn things that help me to stay away from crime after release	reint	0.812	reint	0.847	NA	NA	NA	NA

Item	Theory	F1	L1	F2	L2	F3-5	L3-5	F3-5 (2)	L3-5 (2)
13	On this unit there is much I can decide for myself	autonomy	0.846	autonomy	0.851	NA	NA	NA	NA
13	On this unit I can decide for myself on matters that are important to me	autonomy	0.908	autonomy	0.884	NA	NA	NA	NA
13	On this unit I am encouraged to arrange matters myself	autonomy	0.767	autonomy	0.726	NA	NA	NA	NA
13	On this unit I have enough freedom of movement	autonomy	0.363	autonomy	0.353	NA	NA	NA	NA
5	The visiting room is pleasant	visits	0.616	NA	NA	visits	0.645	visits	0.643
5	My visitor and I can have enough physical contact during visits	visits	0.805	NA	NA	visits	0.875	visits	0.802
5	The visiting hours are long enough	visits	0.695	NA	NA	visits.2	0.476	visits	0.763
5	I have sufficient privacy during visiting hours	visits	0.785	NA	NA	visits	0.647	visits	0.791
5	The staff treat my visitors nicely	visits	0.482	NA	NA	visits	0.466	visits	0.640
5	The visiting hours are frequent enough	visits	0.577	NA	NA	visits.2	0.987	visits	0.708
5	I enjoy receiving visits	visits.1	0.896	NA	NA	visits.1	0.970	NA	NA
5	After receiving a visitor, I feel good	visits.1	0.781	NA	NA	visits.1	0.805	NA	NA
6	I am satisfied with how often… I can see my family, friends or partner here	visits	0.340	NA	NA	contact	0.719	visits	0.452
6	I am satisfied with how often… I can see my child(ren) here	NA	NA	NA	NA	NA	NA	NA	NA
6	I am satisfied with how often… I can see my laywer here	visits	0.353	NA	NA	contact	0.686	visits	0.414
8	I can get medical care here if I want to	care	0.657	NA	NA	care	0.744	care	0.705
8	Health problems are being taken care of adequately here	care	0.700	NA	NA	care	0.936	care	0.767
8	I am satisfied with the work of the nurse	care	0.521	NA	NA	care.1	0.714	care	0.722
8	I am satisfied with the work of the doctor	care	0.632	NA	NA	care	0.541	care	0.768
8	I am satisfied with the work of the dentist	care	0.467	NA	NA	care.1	0.504	care	0.574
8	I am satisfied with the work of the psychologist	care	0.443	NA	NA	care.1	0.785	care	0.552
10	I am satisfied with the recreation activities	actsat	0.863	NA	NA	actsat	0.987	actsat	0.826
10	I am satisfied with the sports	actsat	0.869	NA	NA	actsat	0.877	actsat	0.856
10	I am satisfied with the library	actsat	0.575	NA	NA	actsat.1	0.378	actsat	0.703
10	I am satisfied with my work in this institution	actsat	0.562	NA	NA	actsat.1	0.793	actsat	0.676
10	I am satisfied with the education/courses	actsat	0.554	NA	NA	actsat.1	0.792	actsat	0.676
10	I am satisfied with the outdoor activity	actsat	0.722	NA	NA	actsat	0.661	actsat	0.783
10	I am satisfied with the religious services	actsat	0.548	NA	NA	actsat.1	0.618	actsat	0.629

```
# print(xtable(keep, digits=c(0,0,0,0,2,0,2,0,2), caption="Results of Exploratory Factor Analysis"), in
```

5 Confirmatory factor analysis

Both EFA and CFA are based on the common factor model, so they are mathematically related procedures. We will use CFA to examine whether the structure identified in the EFA works in a new sample. For this, we use data from a second round of data collection (conducted a few months after the first). This second round of data collection includes individuals who take the survey for a second time. We only use data on individuals who participated for the first time, which is about half of the total number of participants in wave 2.

5.1 Data preparation

Like before, we use several subsets of the data for analysis.

1. Dataset 1: All items except "I am satisfied with how often I can see my children here", because this item by definition cannot be answered by a substantial subset of individuals. It retains only complete cases. This dataset is automatically the smallest because it only includes individuals who answered all available questions. (N=122, V=59)
2. Dataset 2: All items that belong to scales without 'no opinion' answers. Less than one percent of the data on these scales is missing. We impute this missing data using a random forest model. (N = 320, V=36)
3. Datasets 3-5: Three datasets for the quality of care, visits, and activity scales separately. It retains only complete cases within each scale. (N=229, V=6 / N=175, V=10 / N=226, V=7)

5.2 Confirmatory Factor Analysis

In the confirmatory factor analysis, we specify a model. This model makes several adjustments from the originally envisioned scales, based on the results from the Exploratory Factor Analysis. We specify both a one-factor model and a model with the number of factors identified in the parallel analysis for the care, visits, and activities scales. We are interested in building an analysis structure that is broadly applicable across samples. We are concerned that the multiple factors represent differences in satisfaction with activities in the sample of interest, or of samples at this specific institution, and thus may not generalize well to other populations.

We drop the two items on how individuals feel about visits because these did not load onto the visits factor in the 1 factor solution. There is general evidence that these items measure something different from general satisfaction with the visits facilities.

We conduct the CFA using a diagonally weighted least squares estimator because the data are ordinal. We allow factors to covary. We cannot conduct a CFA for the full dataset because we have fewer observations (complete cases) than model parameters. We run the analysis for the other datasets only.

We construct a summary table with the results from the CFA analysis. We focus on model fit indices here. We report:

1. Comparative Fit Index (CFI), which is a relative fit index. It assesses the ratio of the deviation of the user model from the worst fitting model against the deviation of the saturated model from the worst fitting model. Conceptually, if the deviation of the researcher-specified model is the same as the deviation of the best fitting model, then the ratio should be one. Values greater than 0.90, conservatively 0.95 indicate good fit.
2. The Root Mean Square Error of Approximation (RMSEA), is an absolute fit index. It compares the researcher-specified model to the observed data. Values below .05 indicate close model fit, values between .05 and .10 indicate a reasonable model fit, and values greater than .1 indicate a poor model fit. We report both the RMSEA point estimate and the confidence interval for the RMSEA.

Results indicate a reasonable model fit for the model using data on scales without no opinion options. All 1-factor models fit poorly, whereas the models based on the number of factors as identified in the parallel analysis reach good (care, activities) and reasonable (visits) model fit.

```r
# Build summary table
tab <- data.frame(dataset = c("Scales without no opinion options",
                        "Care 1",
                        "Care 2",
                        "Visits 1",
                        "Visits 2",
                        "Activities 1",
                        "Activities 2"),
          factors = c(7,1,2,1,3,1,2),
          TLI = c(fitMeasures(m2_cfa)["tli"],
                  fitMeasures(m3a_cfa)["tli"],
                  fitMeasures(m3b_cfa)["tli"],
                  fitMeasures(m4a_cfa)["tli"],
                  fitMeasures(m4b_cfa)["tli"],
                  fitMeasures(m5a_cfa)["tli"],
                  fitMeasures(m5b_cfa)["tli"]),
          CFI = c(fitMeasures(m2_cfa)["cfi"],
                  fitMeasures(m3a_cfa)["cfi"],
                  fitMeasures(m3b_cfa)["cfi"],
                  fitMeasures(m4a_cfa)["cfi"],
                  fitMeasures(m4b_cfa)["cfi"],
                  fitMeasures(m5a_cfa)["cfi"],
                  fitMeasures(m5b_cfa)["cfi"]),
          RMSEA = c(fitMeasures(m2_cfa)["rmsea"],
                          fitMeasures(m3a_cfa)["rmsea"],
                          fitMeasures(m3b_cfa)["rmsea"],
                          fitMeasures(m4a_cfa)["rmsea"],
                          fitMeasures(m4b_cfa)["rmsea"],
                          fitMeasures(m5a_cfa)["rmsea"],
                          fitMeasures(m5b_cfa)["rmsea"]),
          RMSEA_Lower = c(fitMeasures(m2_cfa)["rmsea.ci.lower"],
                          fitMeasures(m3a_cfa)["rmsea.ci.lower"],
                          fitMeasures(m3b_cfa)["rmsea.ci.lower"],
                          fitMeasures(m4a_cfa)["rmsea.ci.lower"],
                          fitMeasures(m4b_cfa)["rmsea.ci.lower"],
                          fitMeasures(m5a_cfa)["rmsea.ci.lower"],
                          fitMeasures(m5b_cfa)["rmsea.ci.lower"]),
          RMSEA_Upper = c(fitMeasures(m2_cfa)["rmsea.ci.upper"],
                          fitMeasures(m3a_cfa)["rmsea.ci.upper"],
                          fitMeasures(m3b_cfa)["rmsea.ci.upper"],
                          fitMeasures(m4a_cfa)["rmsea.ci.upper"],
                          fitMeasures(m4b_cfa)["rmsea.ci.upper"],
                          fitMeasures(m5a_cfa)["rmsea.ci.upper"],
                          fitMeasures(m5b_cfa)["rmsea.ci.upper"]))
knitr::kable(tab)
```

dataset	factors	TLI	CFI	RMSEA	RMSEA_Lower	RMSEA_Upper
Scales without no opinion options	7	0.992	0.993	0.070	0.065	0.075
Care 1	1	0.971	0.983	0.215	0.179	0.254
Care 2	2	1.000	1.000	0.000	0.000	0.076
Visits 1	1	0.979	0.985	0.152	0.123	0.182
Visits 2	3	0.994	0.996	0.079	0.042	0.116

dataset	factors	TLI	CFI	RMSEA	RMSEA_Lower	RMSEA_Upper
Activities 1	1	0.993	0.995	0.153	0.123	0.184
Activities 2	2	1.000	1.000	0.024	0.000	0.072

6 Conclusion