REPORT – LAUSANNE – OCTOBER 2025

Unisanté – Centre universitaire de médecine générale et santé publique

Swiss PREMs Questionnaires Psychometric Validation

Julien Junod, André Frei, Chiara Storari and Isabelle Peytremann-Bridevaux

unisanté
Centre universitaire de médecine générale et santé publique • Lausanne

UNIL | Université de Lausanne

Table of contents

1. Executive Summary

No major issues were identified with the Swiss PREMs tool during the pilot study. However, the following points were noted and discussed:

- Filter questions may be difficult to identify and following questions are not always skipped correctly in the paper version.
- It is recommended to keep the choice of response method open, as requiring respondents to use QR-codes only reduces the response rate by half.
- No signs of questionnaire fatigue were observed. Therefore, it was decided to use slightly shortened versions of a single comprehensive tool, rather than offering both short and long versions of each questionnaire.
- The new questionnaire reflects a shift from measuring patient satisfaction to measuring patient experience, which has proven more useful for quality managers. PREMs scores should be reported using Top Box summary measures, while Bottom Box scores can provide additional insight into problematic responses. All desired thematic areas could be addressed using the existing questionnaire items.
- Funnel plots for national comparisons can be generated using aggregate PREMs scores, similar to how hospital mean satisfaction scores were used in previous routine measurements.
- The expert group proposed not to adjust results following discussions with the ANQ and Unisanté.

2. Introduction

The goal of this third mandate is to develop a new *Swiss PREMs* tool serving a dual purpose:

1. To replace the previous *PatZu*[a] questionnaire used in routine measurements of patient satisfaction for national comparisons.
2. To standardize and consolidate various patient satisfaction measurements currently conducted by hospitals for their internal quality management.

The instrument must therefore generate data that is:

- Detailed enough to support hospital quality management by identifying areas for improvement.
- Structured in a way that allows for aggregation and summarization, enabling meaningful comparisons of hospitals' Patient-Reported Experience Measures (PREMs) at the national level.

The objective of the validation procedure described in this report is to confirm the adequacy of the *Swiss PREMs* instrument in fulfilling these two goals. The first wave of routine measurements using the new Swiss

[a] "*PatZu* stands for *Patientenzufriedenheit* (*patient satisfaction* in German). This abbreviation will be used throughout the report.

PREMs questionnaire is scheduled for autumn 2025 in the acute care domain[b], followed by the psychiatry and rehabilitation domains in early 2026.

The original CPES-IC questionnaire, on which the new *Swiss PREMs* is based, has already undergone a comprehensive validation process. Accordingly, this mandate does not aim to reassess the questionnaire's fundamental content or structure. Rather, it focuses on evaluating the instrument's suitability for use in the Swiss context: specifically, its translation into the country's three national languages, the reformulation of certain selected questions to reflect local realities and sensitivities, and its extension to cover two additional domains of care (psychiatry and rehabilitation). This adaptation involves adding a small number of domain-specific questions and omitting those that are less relevant.

Overall, all original PREMs items, global satisfaction and filter questions were retained without modification. Apart from the domain-specific questions, the only section of the questionnaire that underwent a significant change was the final block of socio-demographic items, entitled "About You." This section was replaced with questions from the previous PatZu questionnaire and expanded to include a new item on education, as well as a control question asking respondents how easy it was to complete the questionnaire.

Validating the questionnaire's suitability for the Swiss context means ensuring that, as a comparative tool, it accurately measures hospital performance without being distorted by artifacts introduced through different modes of administration. Designed to serve as a unified measurement instrument, the questionnaire will exist in 18 distinct versions, resulting from the combination of the following factors:

- **Two formats**: paper and online, to reach the broadest possible population.
- **Three languages**: German, French, and Italian, to cover the entire country.
- **Three domains of care**: acute (somatic) care, psychiatry, and rehabilitation.

It is therefore of utmost importance that all versions produce comparable[c] results.

2.1. Context

In September 2021, the ANQ commissioned Unisanté to evaluate existing instruments for measuring patient experience. This initiative stemmed from the recognition that ANQ's current short-form instrument (comprising six questions) required adaptation to meet the evolving needs of Swiss hospitals. The primary objective of this evaluation was to propose a more comprehensive instrument, featuring an expanded set of questions, which would better align with the quality standards and informational needs of Swiss healthcare institutions.

[b] "Acute care" refers to "acute somatic care" throughout this report and represents one of the three domains of care assessed by the ANQ PREMs measurements.

[c] Not necessarily *similar* but allowing *fair comparisons*.

The ANQ outlined a structured procedure for identifying potential instrument variants, divided into three mandates:

1. Develop two to three potential variants by adapting existing patient experience instruments.
2. Implement the variant selected by the ANQ.
3. Validate the translated and adapted versions of the selected instrument and define a set of summary PREMs statistics for use in national hospital comparisons.

2.1.1. Questionnaire Selection

A procedure was established to identify a set of criteria any adapted instrument would need to meet to be considered for future use in this context. Twenty-five instruments were initially evaluated, of which six were considered for further analysis, including those of five Swiss hospitals: Centre Hospitalier Universitaire Vaudois (CHUV), Hôpitaux Universitaires de Genève (HUG), Inselspital Bern, Universitätsspital Zürich (USZ), Münsterlinger Patientenfragebogen (MüPF) as well as and the Canadian Patient Experiences Survey — Inpatient Care (CPES-IC).

None of these six instruments fully met the requirements set by the ANQ and the experts from the Commission for Quality and Patient Satisfaction (QA PatZu), nor were they immediately ready for implementation. Following the initial report outlining the procedure and results (Storari, Cathieni et al. 2022), the ANQ and QA PatZu thoroughly discussed the selected instruments, considering both methodological validity and their adaptability to the Swiss context and various domains of care. They also evaluated the feasibility of using these instruments without copyright restrictions.

Eventually, it was decided to adapt the CPES-IC instrument to the Swiss context and to the three domains of adult patient's care already covered by ANQ's PatZu routine measurements.

2.1.2. Questionnaire Translation and Qualitative Validation

During the second phase, the CEPS-IC instrument was translated into German, French, and Italian. These translations were adapted by the ANQ and its QA PatZu and field tested during qualitative interviews by Unisanté to suit the three different domains of care: Acute Care as the CPES-IC was originally designed for, as well as slightly modified versions tailored for rehabilitation and psychiatry.

A detailed description of the full procedure and adaptations made to create the newly developed instrument is provided in a second report (Storari, Junod et al. 2024).

2.1.3. Questionnaire Validation

The present report is devoted to the third phase of the project. It describes the psychometric validation procedure of the nine different adaptations of the original CPES-IC questionnaire (i.e., three domains of care and three languages).

Key Research Questions

The ANQ and its QA PatZu raised several points of attention regarding the reliability and suitability of the Swiss PREMs questionnaire for measuring patient experiences, both nationally and at the level of individual hospitals. These points of attention centered on the following areas.

2.1.4. Representativeness and Data Quality

Is the sample usable (i.e., free from bias and representative of the target population) or does it capture only a subset of that population?

For financial and logistical reasons, respondents are encouraged to use the online version rather than the paper format. Does the mode of administration affect who chooses to participate and how respondents complete the questionnaire?

Data validity: is the questionnaire clearly understood across all three language versions, and do we measure the intended variables?

2.1.5. Questionnaire Structure

The revised questionnaire is approximately seven times longer than the previous questionnaire used in PatZu measurements and no longer follows a linear format: respondents are directed along different pathways based on their responses to filter questions.

- Does the increased length discourage respondents from completing the questionnaire, and should it therefore be shortened?
- Do respondents face challenges when required to skip questions?

2.1.6. National Comparisons

The new tool's complexity requires the data to be condensed before it can be used for straightforward hospital comparisons at a national level.

- How can questions be organized and summarized to produce a manageable set of key statistics?
- What constitutes fair comparisons between hospitals, given the known disparities in patient mix? Should indicators be adjusted accordingly, or should raw values be used?

2.1.7. Graphical Representations

The needs of hospitals and decision makers continue to evolve. Additionally, the structure and richness of available data have changed.

- What type of visualization best capitalizes on the more detailed data now available?
- How can graphical comparisons of hospitals' PREMs performance support meaningful interpretation?

The purpose of the aforementioned questions is to ensure that all respondents are placed in similarly favorable conditions for completing the questionnaire, regardless of the setting. This issue was already addressed in the second phase of the mandate at a qualitative level, through interviews with a sample of 27 test patients. The interviews explored how well participants understood the questionnaire and identified sections requiring adaptation for greater clarity. Based on these findings, wording was simplified, and certain questions were reordered to support this objective.

2.2. Report Structure and Layout

This report's overall structure is organized around the similarities and differences among the three domains of care.

Among all care domains covered in this report, acute care receives the greatest attention and most detailed analyses, as it forms the foundation of the Swiss PREMs questionnaire. In contrast, chapters on rehabilitation and psychiatry serve a more confirmatory role. They illustrate how findings from acute care can be extended to these domains, highlighting areas of divergence and identifying where adaptations are necessary.

Consequently, results and comparisons relevant to all domains of care are presented together, while domain-specific issues are addressed separately. Detailed results from the extensive analyses conducted in acute care are available in the appendix.

Note that the results from the three pilot studies are not intended to compare hospitals, but rather to assess the validity of the tool that will later be used for such comparisons in routine national measurements. The findings presented in Chapter 4 are therefore methodological in nature. The content is organized as follows:

- Chapter 3 describes the methodology employed in the pilot studies across three domains of care.
- Section 4.1 covers validation of results.
- Section 0 proposes a short-form version of the instrument based on the validation findings.
- Section 4.3 presents key indicators to be used for national comparisons.
- Section 4.4 explains why the measurements should remain unadjusted, despite differences in hospital contexts. A more detailed discussion is presented in Section 5.3 of the appendix.
- Section 4.5 offers recommendations for visually presenting patient experience results in future national reports by the ANQ.

3. Methodology

This mandate concerns the quantitative step: the psychometric validation of the updated translations of the CPES-IC questionnaire.

Thanks to their similarity of structure and content, the results of all three pilot studies can be analyzed in a similar way and compared. Descriptive summary statistics classically found in questionnaire validation will be computed on each sample to look for possible accidents in the data collection process, signs of misunderstood questions or problems with the questionnaire structure caused by the various means of administrating the questionnaire.

3.1. Pilot Samples

The psychometric validation of the new Swiss PREMs questionnaire was conducted through three parallel studies involving adult patients, each targeting a specific domain of care: psychiatry, rehabilitation, and acute care. Since the original CPES-IC questionnaire, on which Swiss PREMs is based, was designed for patients in acute care, the acute care version was the first to undergo field testing in spring 2024. It is also the most significant both in terms of size and relevance. This was soon followed by the adapted versions for rehabilitation and psychiatry, which were tested in autumn 2024 (precise data collection periods are detailed in Table 2).

3.1.1. Survey Design

Regarding survey sampling, the pilot samples were obtained using a stratified cluster sampling approach, with the primary sampling units (PSUs) being hospitals. As participation in the pilot study (unlike in routine measurements) was voluntary, hospital involvement depended on their availability. Consequently, this stage of sampling was non-random. However, care was taken to ensure the sample was as diverse as possible in terms of linguistic region, size, and type of hospital—three variables used as strata.

The patients were the secondary sampling units (SSUs). At this stage, sampling was random, as all eligible patients in participating hospitals were invited to take part, with inclusion depending solely on their willingness to participate.

The eligibility criteria remained consistent with those used in previous PatZu routine measurements:

- Patients aged 18 years or older at the time of discharge.
- Hospital stay of at least 24 hours.

The exclusion criteria included:

- Patients who died during their hospital stay.
- Patients residing outside Switzerland (apart from Liechtenstein).

Patients who were discharged and readmitted during the measurement period received only one questionnaire.

3.1.2. Determining Sample Size

Determining an appropriate sample size was less a matter of purely theoretical calculation and more a balancing act between multiple practical constraints. Data collection depended primarily on the willingness of hospitals to participate in the pilot study, in addition to their ongoing routine data collection efforts.

Nevertheless, drawing on experience from previous measurements, and considering the expected patient volume during the data collection period and an observed response rate of approximately 40%, decisions were made regarding which and how many hospitals needed to be contacted. The aim was to secure regional subsamples that were both large and diverse enough to enable basic local-level analyses, while also ensuring a total national sample size of at least 1,000 respondents in smaller domains of care. A higher target was deemed unrealistic and overly burdensome for Swiss hospitals.

The final list of participating institutions was compiled by the ANQ coordination team, who directly contacted a diverse selection of hospitals with the goal of maximizing variety and meeting the projected sample size, as detailed in Table 2. Further details on the hospital demographics can be found at the beginning of Section 5.1.1 (acute care) and 5.2.1 (psychiatry and rehabilitation) in the appendix.

3.2. Measuring Data Quality

Standard descriptive statistics were employed to identify potential errors or issues in the data. At the logistics level, the data collection process was susceptible to problems such as the complete loss of questionnaires from a hospital or miscoded responses. At the psychometric level, the questionnaire layout could affect respondent's willingness to participate, respondents' ability to correctly interpret its structure (e.g., which questions to skip), while translation issues could hinder comprehension of certain items in specific languages.

To detect such issues, descriptive statistics were used to examine the proportion of missing or alternative responses, as well as the frequency of extreme values in the four-point PREMs item scales (e.g., is there a question that stands out? Do we observe similar response patterns across different versions of the questionnaire?). These univariate checks were complemented by descriptive principal component analyses (PCA) to provide a multivariate perspective. Stratified analyses were also performed on subgroups differentiated by language and mode of questionnaire administration.

3.2.1. Sample Representativity

Five socio-demographic variables were used to characterize the respondents participating in the pilot studies: gender, age, self-reported health, insurance status and education. Potential biases were identified by comparing their distributions with those observed in prior routine measurements and a reference data population.

In each pilot study, a voluntary subset of participating hospitals provided a complete list of all patients eligible for inclusion. This anonymized list could not be linked to the actual sample (meaning individual participation could not be inferred), but it was used to calculate reference population values for three socio-demographic variables (age, gender and insurance status) at hospital level, enabling the detection of possible biases.

3.2.2. Impact of Data Collection Methods

A key logistical issue was to further investigate the acceptance and impact of alternative methods for administering the questionnaire. In routine PatZu measurements, data were primarily collected via paper questionnaires, even though respondents were offered an online option, which gained acceptance only gradually.

Due to the cost and logistical burden associated with paper forms, the ANQ and the QA PatZu decided to explore the feasibility of transitioning to an exclusively electronic format by testing the outcomes when respondents were required to use the online version. To this end, prior to distribution, the target population was randomly stratified into two equally sized groups:

- The **control group** could choose between completing a paper questionnaire and returning it by mail, or accessing an online form instead, the dual-mode approach already in use in prior PatZu routine measurements.
- The **test group**, by contrast, was only offered the online version.

After hospital discharge, all participants received the questionnaire at home, but with different materials: the test group received a cover letter with a QR code linking to the online form, whereas the control group received a cover letter with both a QR code and an enclosed paper questionnaire. No questionnaires were handed out in hospitals, unlike in previous PatZu psychiatry surveys (see Section 0 for a further details and implications).

The overall impact of both electronic and paper formats on respondents' comprehension of the questionnaire, as well as its translation, was also systematically examined, as well as a potential influence of socio-demographic factors.

3.3. Shortening Indicators

An expert group identified a set of *core* topics deemed essential for comparing hospitals at the national level, along with *optional* questions to provide additional insights when available. The categorization of questions into core and optional groups was informed by qualitative requirements from the ANQ and the QA PatZu, alongside considerations of the instrument's psychometric properties, a methodology previously validated in similar studies (Epstein, Santo et al. 2015, Hadibhai, Lacroix et al. 2018).

Ultimately, the ANQ and Unisanté chose a cautious approach, retaining most questions from the original CPES-IC questionnaire. The resulting unified tool is therefore intended to support both national hospital comparisons and internal quality management.

3.3.1. Reasons for Shortening

In addition to logistical and economic considerations, this validation procedure allows us to measure two potential negative effects of excessive questionnaire length:

- **Questionnaire fear**: Are we losing a significant number of respondents at the outset? By comparing response rates between the previous PatZu measurements and the pilot study, can we detect a substantial drop?
- **Questionnaire fatigue**: Are respondents dropping out midway through the questionnaire? Do we observe increasing missing data rates from the beginning to the end of the questionnaire?

Affirmative results for both would indicate that the length of the Swiss PREMs questionnaire contributes to data loss and adversely affects measurement reliability.

3.3.2. Shortening Strategy

Efforts to shorten the questionnaire represent a balancing act between two opposing objectives:

- Ensuring the questionnaire is **sufficiently detailed** and comprehensive to meet quality management requirements, the very reason the ANQ decided to replace its brief patient satisfaction survey with a more elaborate patient experience measurement tool.
- Keeping the questionnaire **concise enough** to avoid discouraging respondents from completing it or imposing an excessive logistical and analytical burden. This was the primary motivation behind the shortening effort.

Importantly, the questions serve very different functions within the questionnaire: in acute care, there are 4 filter questions, 4 global satisfaction items, 6 socio-demographic questions, 32 core PREMs items using a four-point scale, 2 binary PREMs questions, and one final control item intended for use during the pilot phase only. As a result, these questions cannot be treated uniformly in the evaluation process.

Since there is no established standard for the ideal length of a questionnaire, and the selected questions needed to fulfill a dual purpose (supporting both quality management requirements at hospital level and serve in national-level comparisons), a combination of qualitative and quantitative methods was adopted to address the variety of criteria involved.

First, descriptive statistics were used to identify PREMs items potentially affected by the following issues: lack of relevance, redundancy, limited discriminative power or incoherence (see Section 3.4.3 for detailed definitions).

PREMs items displaying such issues were compiled into a list and submitted to the QA PatZu expert committee for further review and deliberation regarding the benefits and drawbacks of retaining them in the questionnaire. The remaining items may be evaluated individually on a case-by-case basis.

The decision to retain or remove a question often carries broader implications, as each of the 28 to 31 PREMs items is a potential candidate for inclusion in a composite PREMs score, as outlined in Section 3.4 below. Consequently, any effort to shorten the questionnaire must take these global scores into account, since they

require a minimum number of items to form a valid construct. Such an evaluation may also reveal further redundancies and complementary relationships among questions.

3.4. Constructing PREMs Score

The comprehensive set of PREMs questions effectively addresses quality management concerns at the hospital level. However, its breadth and detail can become overwhelming when used to compare dozens of hospitals at the national level. Therefore, some kind of reduction and simplification are necessary before proceeding with national comparisons. These figures may also prove useful at the hospital level, particularly when a brief summary of performance is needed.

3.4.1. Themes

In alignment with the requirements set by the ANQ and the QA PatZu, a set of broad key themes of interest was defined to ensure that any reduction in content would still capture meaningful performance metrics:

- **Admission**: Entry procedures.
- **Information**: Communication with nurses, medical staff, and other healthcare personnel.
- **Involvement**: Participation in decision-making.
- **Organization**: Processes, structure, and waiting times.
- **Discharge**: Discharge planning, coordination, and aftercare.

Additionally, the following secondary themes should ideally be included as well:

- **Environment**: Information on the hospital environment
- **Respect**: Patient–provider interaction
- **Medication**: Information on medication use, side effects, and pain management

The final instruments must also meet the following functional criteria:

- **Relevance**: Contribute meaningfully to hospital quality improvement.
- **Specificity**: Enable long-term monitoring of patient perspectives.
- **Simplicity**: Present results in a clear and patient-friendly format.
- **Differentiation**: Provide comparative insights for ANQ partners and stake holders (e.g., payers).

The CEPS-IC and its adapted versions evaluated in this report all meet these content and measurement criteria.

3.4.2. Questions

Having determined which themes to address, the next step involves assigning each question to an appropriate thematic score. This process unfolds through several iterative stages, combining trial-and-error with both qualitative expertise and quantitative evaluation:

- First phase (QA PatZu, ANQ, Unisanté): Initial intuitive assignment by experts from three institutions, followed by inter-judge agreement to establish a preliminary structure.
- Second phase (Unisanté only): Use of the initial assignments to group related questions, assessing coherence through quantitative methods.
- Third phase (Unisanté only): Adjustment of assignments by reallocating questions to themes deemed more suitable.
- Fourth phase (Unisanté only): Calculation of metrics for the revised assignments and qualitative evaluation of the newly defined themes.
- Next phases (Unisanté only): Through trial and error, an assignment structure was developed that meets the quantitative criteria at both the item and thematic levels, as described in Sections 0 and 5.2.2 of the appendix.
- …
- Final phase (QA PatZu, ANQ, Unisanté): Review of the proposed assignments, discussion of validity, and refinement suggestions…
- …until consensus is reached.

3.4.3. Score Development Principles

Beyond expert input and guidance from the field, the following quantitative measures were employed to assess the relevance of each assignment:

- **Internal coherence** of the question sets forming a theme was evaluated using Cronbach's alpha, to determine how effectively each item contributed to the overall construct.
- **Similarity** among questions (expected to be higher within themes and lower across themes) was assessed through correlation analysis and principal component analysis.
- **Discriminative power**, the ability of a question to distinguish between respondents with low versus high scores on the PREMs scale within a given theme, was analyzed using item response theory (IRT) information functions. Moderate floor or ceiling effects were considered desirable for enhancing discrimination.
- **Relevance** of PREMs items was further validated by their correlations with four global satisfaction questions and by examining the frequency of alternative responses such as "does not apply."

Once a set of questions has been chosen to build a score, they must be combined to produce a single number. We follow the strategy adopted by the CIHI in analyzing their CPES-IC questionnaire, by focusing on extreme values. This approach is a radical change from the former PatZu routine measurements, since it does not rely on a (central) mean but turns its attention to the edges of the scale.

3.4.4.Score Formula

The CPES-IC departs from five-item Likert-scales containing a central, neutral item, as they were used in the PatZu measurements. The new scales only have four items, lacking the neutral one, and hence forcing respondents to choose a lower or higher grade.

The questionable assumption that scale items are evenly spaced is no longer required once we compute the proportion of respondents selecting the lowest or highest response option, rather than the mean of responses. Therefore, we assess floor and ceiling effects. Not as nuisances, but as the very metrics of interest.

Each theme produces two scores: one for the highest answers, one for the lowest ones, called a *top box* and a *bottom box* respectively. Technically, this is a three-step process:

- Compute indicator variables for each question targeting those respondents having answered the lowest or highest item. For a respondent i answering a question $q_{\cdot j}$, the box values would be $t_{ij} = 1$ if $q_{ij} = 4$ and 0 otherwise, while the bottom box indicator is defined as $b_{ij} = 1$if $q_{ij} \leq 2$.

Question $q_{\cdot j}$	Top Box $t_{\cdot j}$	Bottom Box $b_{\cdot j}$
4 (Best)	1	0
3 (Better)	0	0
2 (Worse)	0	1
1 (Worst)	0	1

- For each respondent i, compute the score $s_i = \frac{1}{J}\sum_{j=1}^{J} t_{ij}$ as a (horizontal) average of indicator variables yielding the proportion of top boxes $t_{i1}, \dots, t_{iJ}$ (resp. bottom boxes $b_{i1}, \dots, b_{iJ}$) inside a given theme based on questions $q_{\cdot 1}, \dots, q_{\cdot J}$:

Respondent	…	$t_{\cdot 1}$	$t_{\cdot 2}$	$t_{\cdot 3}$	Score s	…
1	…	1	0	0	33.3%	…
2	…	0	1	1	66.7%	…
…	…	…	…	…	…	…
i	…	…	t_{ij}	…	s_i	…
…	…	…	…	…	…	…

- Compute the (vertical) average of all participants' scores $S = \sum_{i=1}^{I} s_i$, hence an average proportion over the entire population, or restricted to a given hospital, etc.

Respondent	…	t_1	t_2	t_3	Score s	…
1	…	1	0	0	33.3%	…
2	…	0	1	1	66.7%	…
…	…	…	…	…	…	…
Total	…	…	…	…	50.0%	…

- Missing values in q_{ij} are handled by excluding them from the calculation and computing the average based on the remaining responses[d]. This approach assumes that top and bottom box responses are relatively homogeneous distributed within each theme (see Figure 45 in acute care and Figure 85 in psychiatry and rehabilitation) with minimal internal variation. This method follows the practice of the CIHI and differs from previous methodology used for PatZu measurements in child psychiatry, where averages were calculated only if at least half of the responses were available.

3.5. Adjusting PREMs Scores

When evaluating hospitals' performance in quality management at the national level, it is expected that patient experience measures be compared on a fair basis. It is assumed that a hospital's case mix may influence patient experience measures, and that the scores should reflect the hospital's actual performance, not the complexity of its patient population. It is also assumed that some hospitals may handle more challenging cases, while others benefit from easier ones. To account for these differences, it has been proposed to adjust PREMs scores according to patient demographics, enabling a more standardized comparison across hospitals.

Context

This adjustment procedure has been consistently applied in previous patient satisfaction reports (with the exception of child psychiatry, due to smaller sample sizes). However, questions remain as to whether such adjustments still serve their purpose in national comparisons. Before determining whether adjustment remains appropriate, it is important to assess how and when adjustments affect individual hospital scores, and most importantly, whether the data used to perform these adjustments are sufficiently relevant to ensure genuinely fair comparisons.

Data

The pilot sample in psychiatry will be used throughout this exploratory data analysis, as it includes information about the mode of patient admission (involuntary or elective), which is known to directly influence patient experience.

[d] If only a single value is available, it is used as the average. If two values are available, the average is based solely on those with missing responses ignored. When all questions are answered, the average is computed using the full set of questions defining a theme.

Hospitals of different types are expected to serve different patient populations, who, in turn, are likely to evaluate their hospital experiences differently. The most suitable proxy for defining a hospital's *type* is its category according to the BfS taxonomy. Meanwhile, the only patient-related characteristics available for analysis are a set of socio-demographic variables.

3.5.1. How Results are Adjusted

Instead of directly computing raw averages of PREMs scores across a hospital's patients, an adjusted hospital's score is derived using a regression analysis (Savidan-Niederer, Peytremann Bridevaux et al. 2022). This model incorporates the following available socio-demographic factors along with a hospital indicator variable; age, sex, …. A subsequent post-analysis step then calculates the marginal mean for each hospital.

3.5.2. Information Available for Adjusting PREMs Scores

In the context of the Swiss PREMs questionnaires, information about a hospital's case-mix is obtained through the socio-demographic factors reported by respondents. The final "About you" section includes questions about a patient's age and gender. Socio-economic data may be gathered via indicators such as insurance coverage (common or half-private/private) and education level (to be discarded from future questionnaires). Self-reported health status addresses the medical dimension.

The key question regarding adjustment is whether this set of variables is appropriate for that purpose. Specifically, whether their inclusion is:

1) **Relevant**: Do they influence patient experience?

2) **Effective**: Do they help mitigate potential bias?

While other health-related variables like comorbidities have proven valuable for adjusting scores in comparative analyses, it is less clear whether the limited set of variables available in the Swiss PREMs questionnaire can achieve the same level of effectiveness.

3.6. Representing Measurement Results

Adopting the Swiss PREMs questionnaire to replace the former PatZu one entails not only a change in the measurement tool but also a rethinking of how to use, summarize, and present the measurement results. The scope (forty questions instead of six) and format (four-item scales rather than five-point Likert scales) of the collected data require new approaches to process and communicate the findings effectively.

3.6.1. Extreme Values

The Canadian Institute for Health Information (CIHI) adopted an extreme-value-based approach to represent findings from its CPES-IC questionnaire (CIHI 2022). This contrasts sharply with ANQ's former PatZu measurements, which relied on averages computed from a five-point Likert scale that included a neutral midpoint. In this context, respondents overwhelmingly selecting either the highest or lowest option were

viewed as problematic, a phenomenon known as *floor* and *ceiling* effects. These effects suggest that the scale used may not adequately distinguish between satisfied and dissatisfied patients.

The CPES-IC's four scale-item scales on which the PREMs questions are based forces respondents to choose either positive or negative sides (see Table 1), concentrating informational value at the extremes: how many respondents selected the best or worst possible ratings. The two middle options carry less interpretive weight, as confirmed by item response theory (see the end of Section 0 in the appendix).

Table 1 Example of scale differences between the former PatZu measurements (5-item Likert scale with neutral item) and the new Swiss PREMs questionnaire (4-item scale without neutral item) based on CPES-IC when measuring frequency.

PatZu	Swiss PREMs
Never	Never
Rarely	Sometimes
Occasionally	---
Very often	Usually
Always	Always

Notably, even items that exhibit a pronounced ceiling effect can still be informative. For instance, if 95% of respondents typically select the highest rating, any sudden deviation from this pattern may serve as a warning signal at the hospital level.

A similar method was already applied in the PatZu reports: the proportion of problematic responses, defined as the percentage of respondents choosing either of the two negative options, not just the lowest, was used to identify areas with potential for improvement.

Given the consistently low number of respondents selecting the low ratings, and to maintain continuity with the PatZu reports, we chose to apply the same approach in the Swiss PREMs questionnaire by defining the bottom box as the proportion of respondents who selected either of the two negative ratings (lowest and lower values). The top box is defined in the same way as CIHI does, by using only the highest rating.

4. Results

4.1. Measurements of Data Quality

4.1.1. Summary

No major issues were identified in the three pilot samples, such as a faulty data collection or biases resulting from different data collection methods. However, three areas show potential for improvement or warrant further scrutiny:

- The experiment administering the questionnaire via QR-code did not meet expectations. Whether the format (paper vs. online) consistently exerts an adverse effect on response rates will be further examined during the first nationwide wave of measurements in 2025.
- The layout of the paper questionnaire's needs refinement to help respondents better understand filter questions and which subsequent items to skip.
- Response rates in psychiatry are markedly lower than in other domains of care. Nevertheless, the ANQ and the QA PatZu ultimately endorsed a uniform methodology, i.e. sending questionnaires to patients' homes after discharge, as already practiced in rehabilitation and acute care. This contrasts to the former PatZu approach, in which hospital staff handed out questionnaires directly to patients before discharge, a procedure that produced substantially higher response rates. The resulting decline in returns is expected to be offset by a significantly reduced risk of selection bias (a smaller but more representative sample).

Sample Sizes

In total, the three samples included 8,053 patients treated in 118 hospitals across Switzerland (see Table 2), representing a broad range of BfS[e] hospital categories. Eligibility criteria required patients to be Swiss residents who had been discharged during the relevant sampling period. The duration of each data collection phase was adjusted based on patient volume in the respective specialties, ranging from two weeks to three months.

[e] Federal Statistical Office, Krankenhaustypologie, Neuchâtel, 2006.

Table 2 The three samples of the Swiss PREMS Pilot studies.

	Psychiatry	Rehabilitation	Acute Care
Questionnaires used	1 128	1 650	5 275
Hospitals participating	37	34	47
Sampling period start	01.09.2024	01.09.2024	15.04.2024
Sampling period end	30.11.2024	31.10.2024	01.05.2024 (DE) and 15.05.2024 (FR+IT)

A small portion of questionnaires is consistently returned completely blank, either submitted online without any answers selected (despite a layout designed to prevent this) or sent back entirely untouched. The frequency of such occurrences was similar across samples and aligns with patterns observed in the PatZu measurements. Table 3 illustrates these findings, along with the distribution of language versions within the sample.

Table 3 Number of usable questionnaires in the Swiss PREMs pilot samples, along with the distribution of language versions for each.

	Psychiatry		Rehabilitation		Acute Care	
Questionnaires received	1 173		1 704		5 422	
Empty Questionnaires	45	3.8%	54	3.2%	147	2.7%
Questionnaires Used	1 128	96.2%	1 650	96.8%	5 275	97.3%
In German	916	81.2%	1 157	70.1%	3 264	61.9%
In French	137	12.1%	410	24.9%	1 545	29.3%
In Italian	75	6.7%	83	5.0%	466	8.8%

4.1.2.Response Rates

Table 4 lists the overall response rates achieved in the three pilot studies. In acute care, hospitals' individual rates range from 20.0% to 63.7%. At first glance, this suggests a notable decline (approximately one-quarter fewer responses) compared to previous rates observed in the PatZu Measures, which reported an overall response rate of 42.7% in 2023 (Table 5).

Table 4 Overall response rates across the three pilot samples, along with differences influenced by the mode of respondent contact.

	Psychiatry	Rehabilitation	Acute Care
Questionnaires sent	7 780	6 091	18 745
Questionnaires used	1 128	1 650	5 275
Overall response rate	14.5%	27.1%	28.1%

However, response rates are highly sensitive to data collection procedures, as discussed further in Section 4.1.9.

Table 5 Overall response rates from the last waves of PatZu measurements across the three domains of care (2023–2024), along with the proportion of respondents who used the online form. Note that the administration of questionnaires in psychiatry differed markedly from other domains of care: instead of being mailed to patients' homes after discharge, they were handed directly to respondents by hospital staff. Moreover, due to an unconscious selection during this distribution process, the number of questionnaires "sent" is lower by about a third than the actual number of eligible patients, resulting in a substantially higher response rate.

	Psychiatry	Rehabilitation	Acute Care
Last year of PatZu Measures	2024	2024	2023
Questionnaires sent	13 433	15 542	75 808
Questionnaires used	8 448	6 727	32 378
Overall response rate	62.9%	43.3%	42.7%
Share of electronic answers	3.6%	5.5%	9.7%

4.1.3.Sample Description

Participants can be described based on a set of socio-demographic questions, which they were asked to complete at the end of the questionnaire, following the PREMs and satisfaction sections.

Age

Respondents in the psychiatry sample are generally younger than those in the rehabilitation sample (Figure 1).

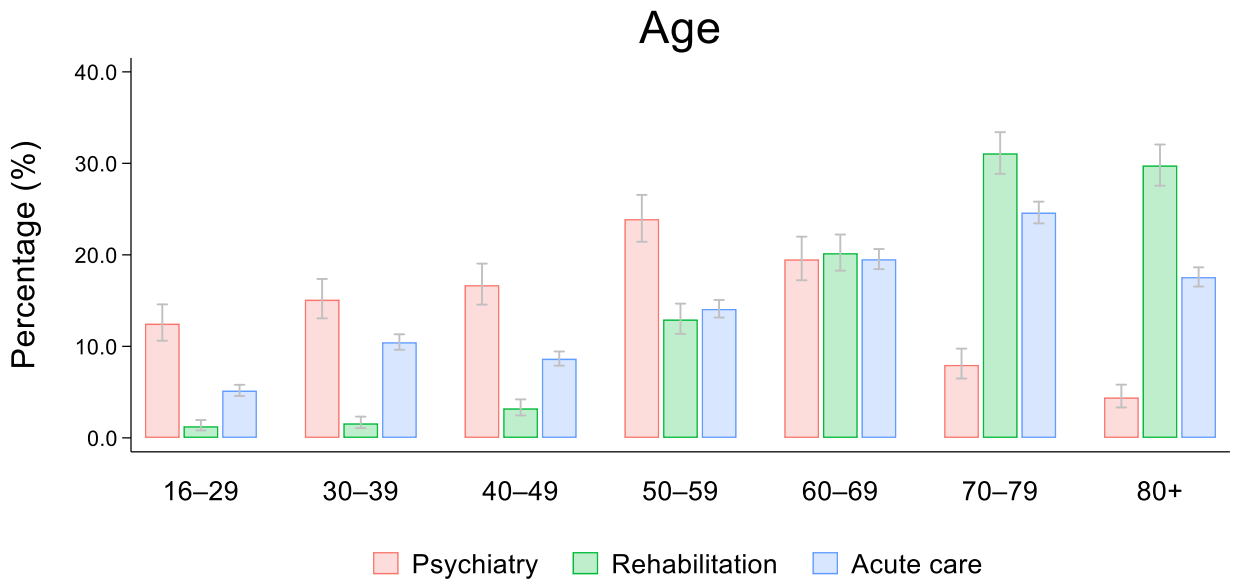


Figure 1 Differences in respondent's age across pilot samples.

Gender

The gender distribution is balanced and nearly identical in the rehabilitation and acute care samples. However, the psychiatry sample includes a slightly higher proportion of women (about 5% more) as shown in Figure 2. The third category, "diverse", was selected infrequently, with only 2 to 6 respondents in each sample.

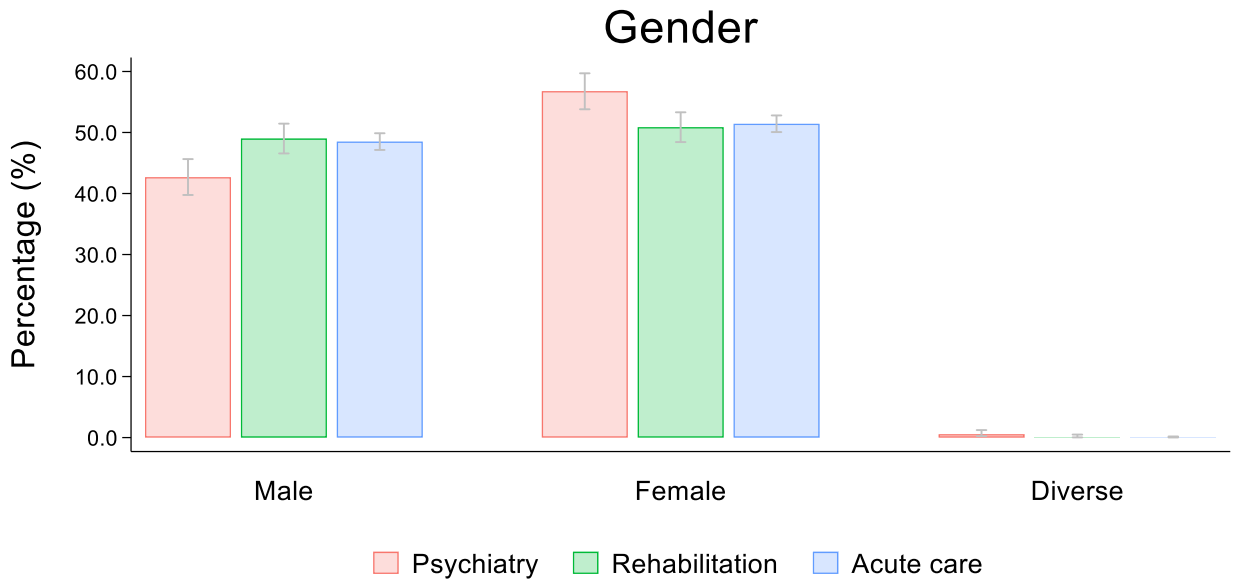


Figure 2 Differences in gender distribution across pilot samples.

Self-reported Health

The distribution of self-reported health ratings (Figure 3) is consistent across all domains of care, with most respondents rating their health in the medium category “good”. Patients in psychiatry tend to report poorer health compared to those in acute care. Rehabilitation falls in between.

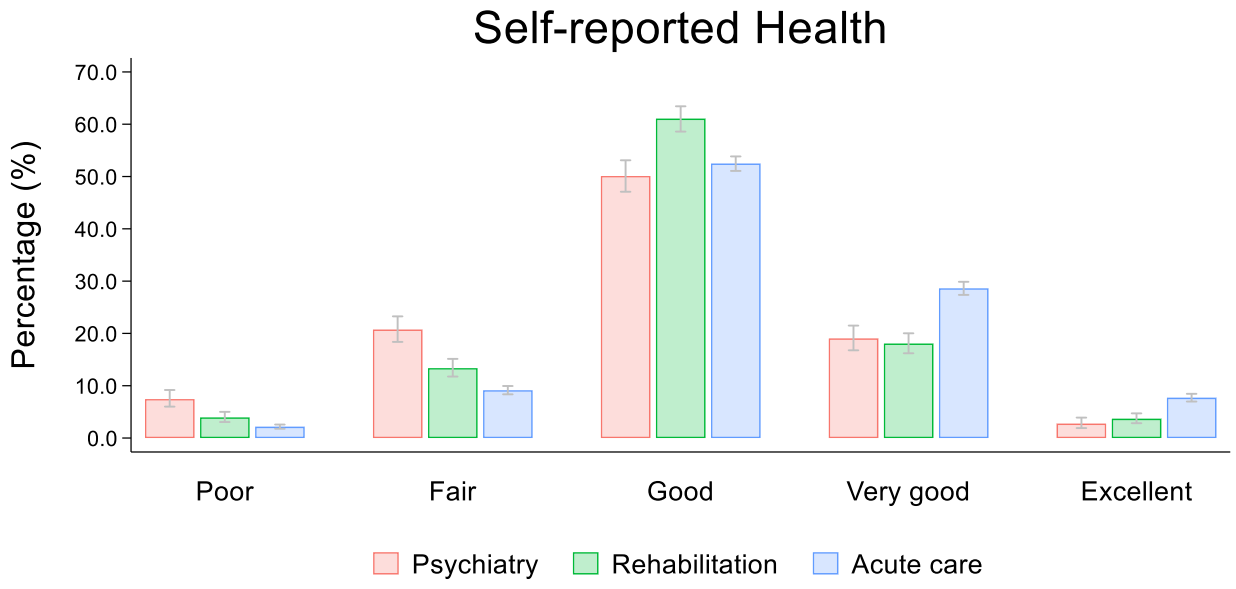


Figure 3 Differences in self-reported health across pilot samples.

Insurance Status

The proportion of patients with semi-private or private insurance is nearly identical (one quarter) in the acute care and rehabilitation samples. The psychiatry sample differs slightly, with about 5% fewer patients holding higher-tier insurance coverage.

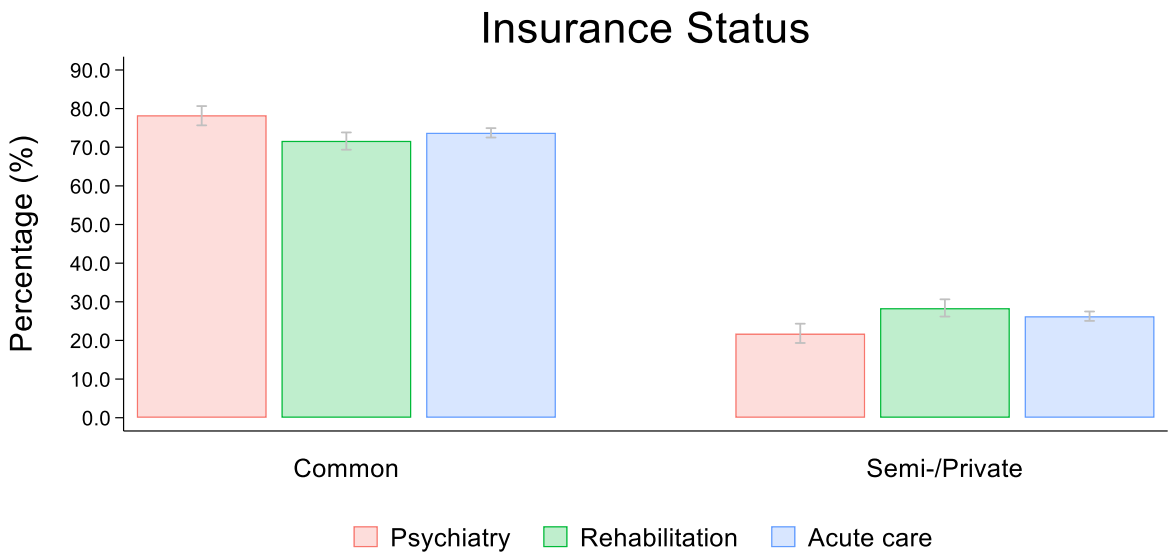


Figure 4 Differences in patient's insurance coverage across pilot samples.

Education

Given the sample size, differences in education are statistically significant at the 5% level: more patients with an intermediate level in rehabilitation, and more with a university degree in acute care and psychiatry. However, no clear pattern or meaningful differences can be inferred from Figure 5.

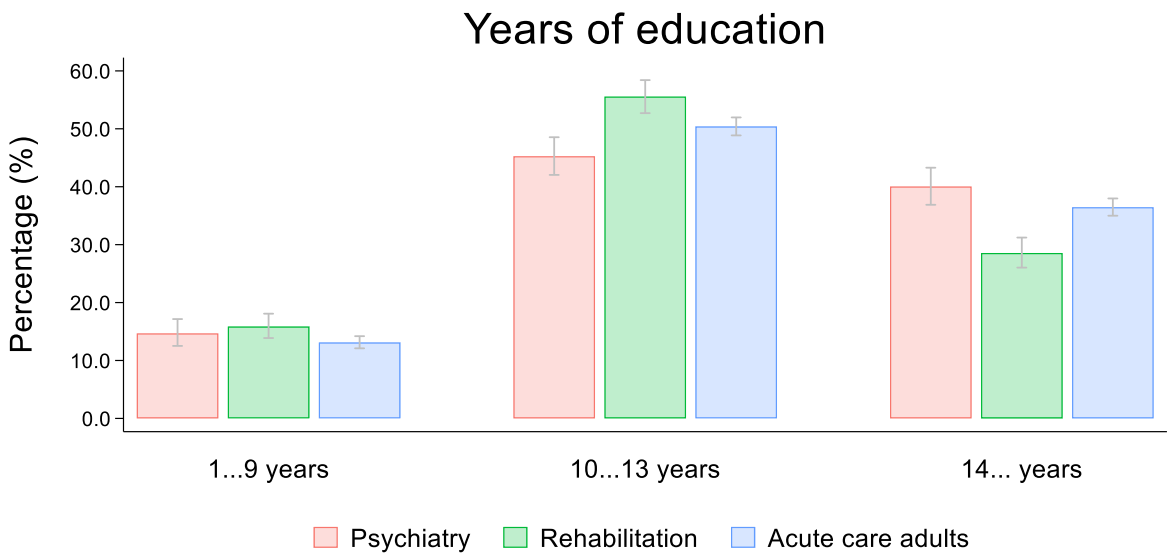
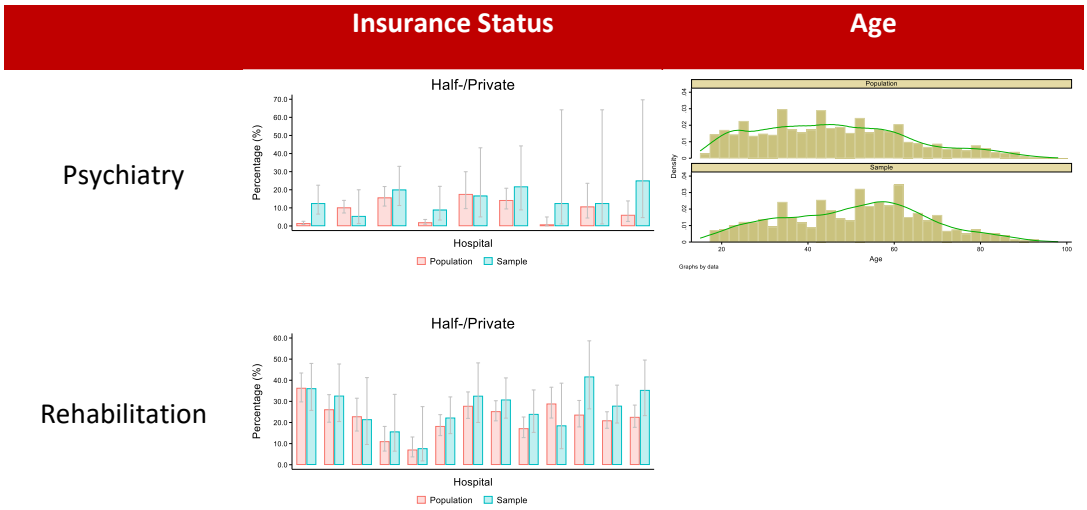


Figure 5 Differences in education level across the three pilot samples. Increasing durations correspond to education levels: compulsory school, vocation/middle/high school and university respectively.

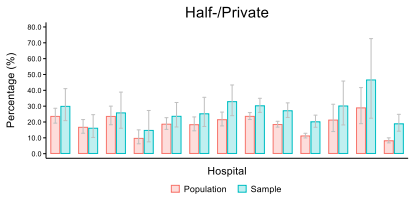
4.1.4.Representativity

Among the three socio-demographic factors available for comparison between the reference population (see Section 3.2.1) and the pilot samples, the only notable differences were an underrepresentation of younger respondents in the psychiatry domain, and a non-significant but consistent overrepresentation of individuals with semi-private or private insurance, observed across nearly all participating hospitals, as shown in Figure 6. Differences in gender were never statistically significant.

Figure 6 Relevant differences between pilot samples and reference populations in the three pilot studies. Empty cells indicate that no statistically significant differences were observed.



Acute Care



4.1.5.Missing Values

Questions

Rates of missing values (i.e., questions left unanswered without selecting an alternative response) are presented in Figure 7. Most items exhibit missing rates below 5%, which is consistent with findings from earlier PatZu measurements. The elevated rates observed in the center of the figure are largely attributable to misunderstandings involving skip logic, a topic explored in greater detail in Section 4.1.8.

There are no notable differences across the questionnaires, with the exception that psychiatric patients show higher missing rates in items related to therapists, as well as in Question 22 concerning symptom monitoring in rehabilitation.

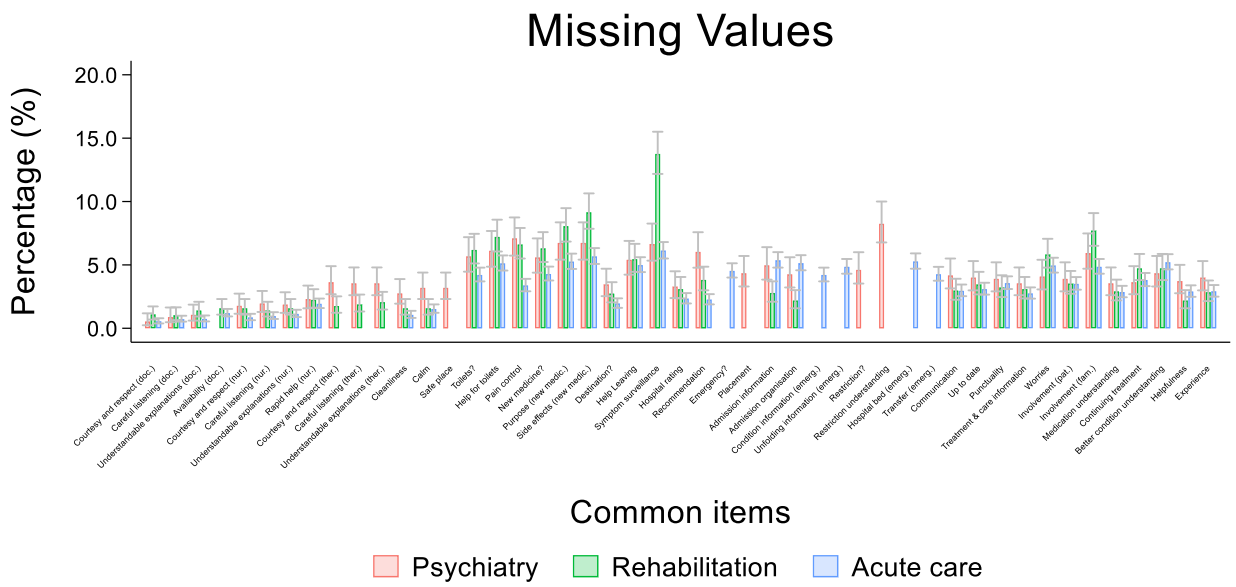


Figure 7 Proportions of missing answers to corresponding questions across pilot samples.

Socio-demographic Factors

The proportion of respondents who did not complete the personal questions in the final section remains below 10%, aligning closely with findings from the PatZu studies. Missing data rates are also consistent across the three domains of care. However, the newly introduced question on education exhibits a threefold

increase in missing responses (combining both actual omissions and explicit refusals) which would significantly reduce the sample size available for potential adjustments, as discussed in Section 3.5.

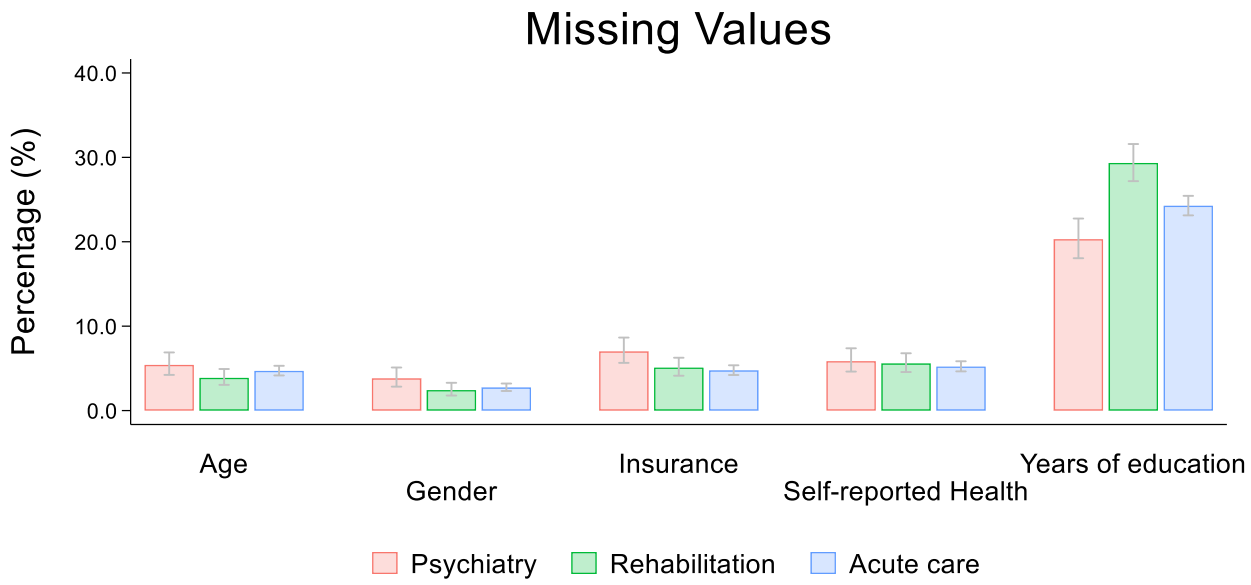


Figure 8 Proportions of missing values for socio-demographic variables in the pilot studies conducted across the three domains of care.

4.1.6. Alternative Answers

Alternative responses arise from refusals, indications that a question does not apply, or skip patterns triggered by filter questions. Figure 9 reflects structural differences in the questionnaires across the three domains of care with fewer common items than those shown in Figure 8, which displays missing values.

The highest rates of alternative answers stem from skipped questions: for instance, the four items related to the emergency department in acute care. Similarly, alternate responses may indicate that a question is irrelevant. Such cases are found in questions about new medications, pain control, and most notably, assistance with toileting. The exceptionally high proportion of respondents indicating that this question does not apply is further examined in Section 4.2.3, which addresses questionnaire shortening in psychiatry and rehabilitation.

Finaly, the elevated rate of alternative responses in the involvement question, shown on the right side of Figure 9, reflects patients' preference not to involve their family or relatives.

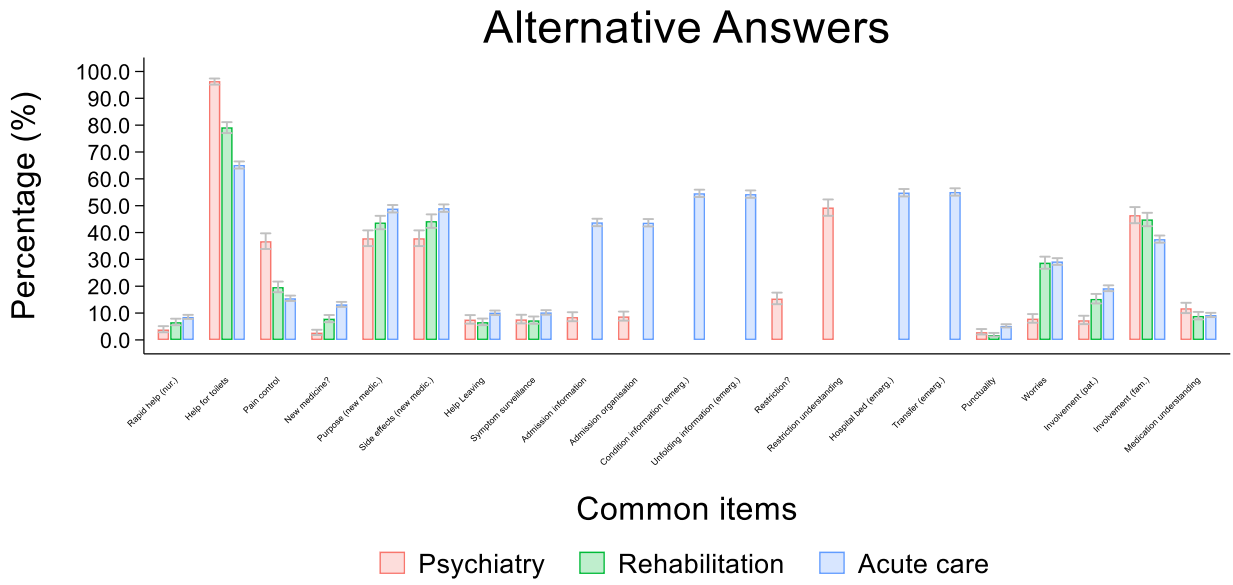


Figure 9 Proportion of respondents selecting an alternative answer (several choices may be possible) to comparable questions across pilot studies.

4.1.7.Extreme Values

Floor and ceiling effects were regarded as a nuisance in the previous PatZu studies. These saturation effects occur when a majority of respondents select the lowest, respectively highest response. When calculating averages on 5-point Likert scales, a disproportionate clustering of respondents at the extremes may indicate a poorly designed scale: one that is too narrow and lacks sufficient discriminatory power among respondents.

However, the context of the Swiss PREMs is different, with the newly introduced 4-point scale (see Table 1), necessitating a redefinition of indicators and summary statistics, as outlined in Section 3.4.3.

Ceiling Effect

Figure 10 illustrates respondents' tendency to select the highest value in PREMs questions. Although variations across domains of care are known to occur, the proportions remain within a comparable range, confirming that the three questionnaires yield consistent results across all domains.

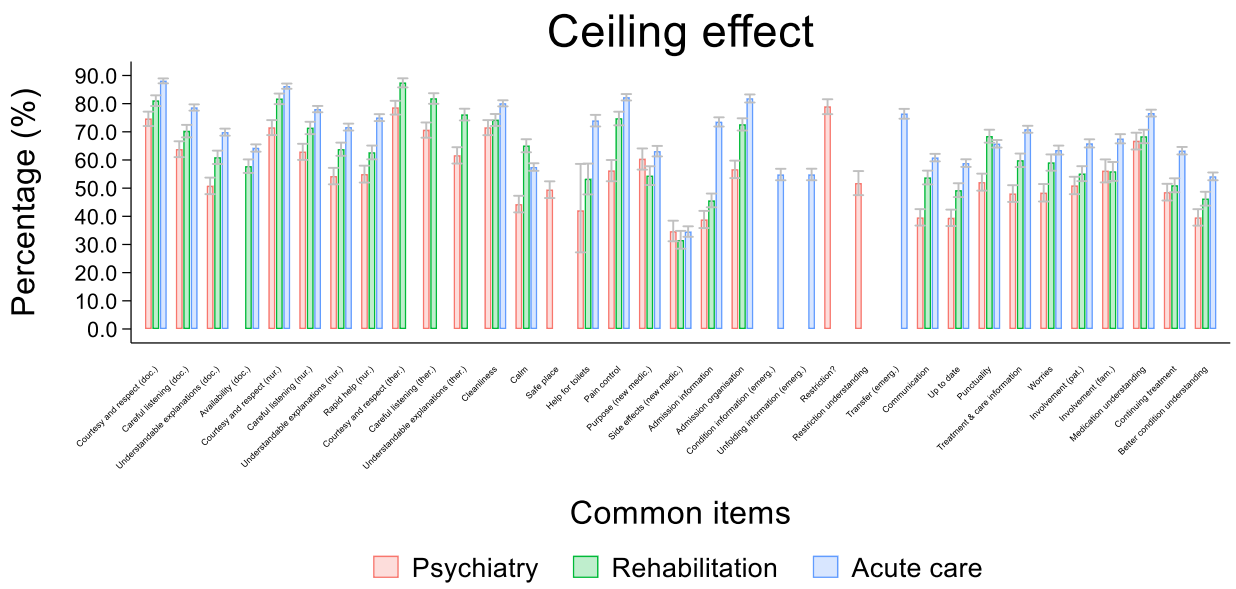


Figure 10 Proportion of respondents selecting the highest response to comparable PREMs items across pilot studies.

Floor Effect

Figure 11 illustrates respondents' tendency to select the two lowest response options in PREMs questions. However, the number of respondents contributing to the floor effect is very small, and these figures are less stable compared to the more populated ceiling effect. Nevertheless, the patterns observed in Figure 11 remain consistent across all domains of care.

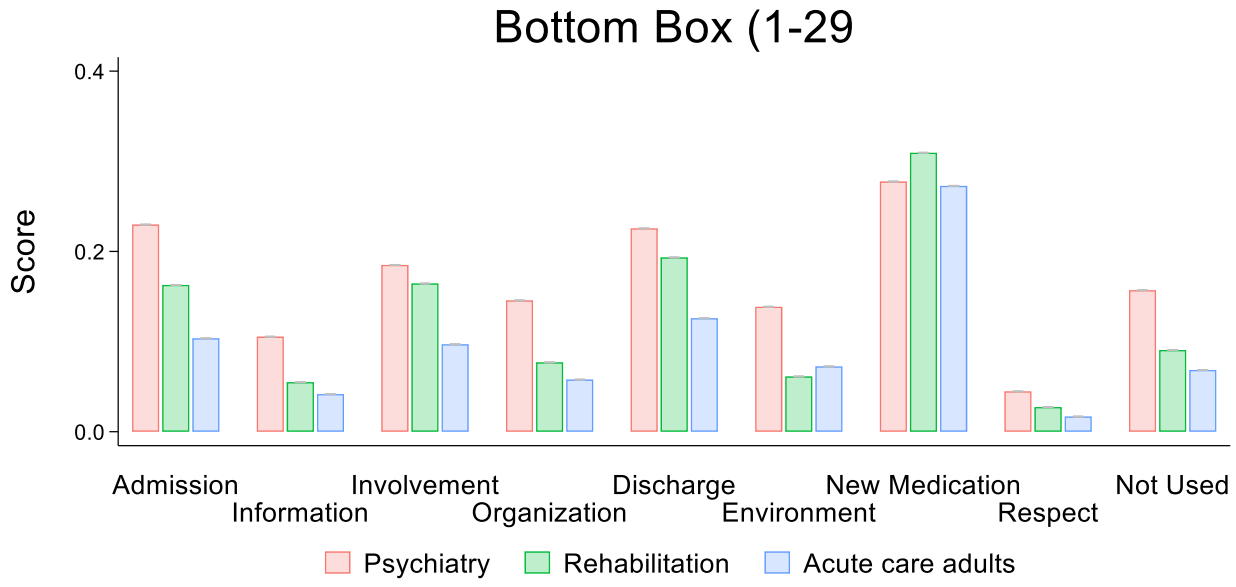


Figure 11 Proportion of respondents selecting the two lowest response options on comparable PREMs items from pilot samples across domains of care.

4.1.8. Skip Questions

The error rate caused by filter questions—such as answering when one should skip, or conversely, leaving a question blank when a response is expected—remains below 5% in most cases, as shown in Figure 12.

The discharge filter question consistently accounts for the highest error rates, with patients returning home often failing to answer the two subsequent questions. In the rehabilitation domain, this error rate even exceeds 10%. The second most frequent issue involves the new medication filter, where respondents frequently neglect to answer questions about the purpose and side effect of newly prescribed drugs.

In the rehabilitation sample, no respondents overlooked the filter question and mistakenly answered the subsequent set of questions. However, such instances did occur in the admission sections of acute care and psychiatry.

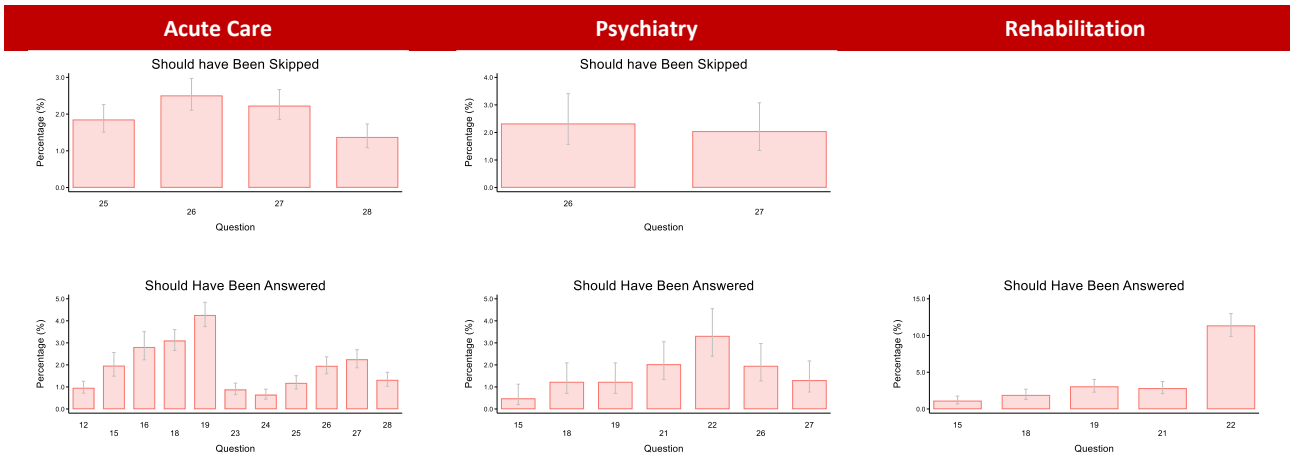


Figure 12 Misunderstood filter and skip questions across the three domains of care in the pilot studies. The top row shows the proportion of respondents who should have skipped the question but answered it, while the bottom row shows those who should have answered it but left it blank.

4.1.9.Influence of Data Collection

Across all three pilot studies, respondents' self-reported health and questionnaire acceptability (i.e., ease of completion) appear systematically correlated with patient experience, as shown in Table 6. The prevalence of missing values and alternative responses is more often affected by the questionnaire format (see discussion below).

Language and age also seem to exert an influence on the proportion of extreme values, particularly within the acute care domain. Note however, that the acute care sample is five times larger than those from other care domains, which may account for the higher number of statistically significant associations observed.

Table 6 Proportion of items influenced by different data collection modes or socio-demographic factors across pilot studies. Associations between factors or administrative variables and quality measures were considered

significant at the 5% level (green indicates no association, red a consistent association, and yellow an association present in half of the cases).

	Psychiatry				Rehabilitation				Acute Care			
Influence	**Mis.**	**Alt.**	**Ceil.**	**Floor**	**Mis.**	**Alt.**	**Ceil.**	**Floor**	**Mis.**	**Alt.**	**Ceil.**	**Floor**
Contact	46.3%	4.9%	3.2%	9.7%	36.8%	13.2%	55.2%	48.3%	60.0%	37.5%	13.8%	17.2%
Format	51.2%	9.8%	6.5%	3.2%	47.4%	13.2%	51.7%	34.5%	62.5%	37.5%	13.8%	13.8%
Language	12.2%	22.0%	35.5%	67.7%	5.3%	21.1%	69.0%	17.2%	20.0%	47.5%	89.7%	82.8%
Acceptability	61.0%	19.5%	93.5%	96.8%	26.3%	21.1%	100%	96.6%	50.0%	32.5%	100%	100%
Gender	9.8%	7.3%	38.7%	54.8%	26.3%	15.8%	24.1%	27.6%	7.5%	32.5%	44.8%	48.3%
Age	29.3%	22.0%	61.3%	45.2%	63.2%	10.5%	27.6%	37.9%	80.0%	45.0%	86.2%	79.3%
Health	19.5%	9.8%	87.1%	90.3%	18.4%	7.9%	96.6%	93.1%	42.5%	37.5%	100.0 %	96.6%
Insurance	9.8%	2.4%	67.7%	71.0%	2.6%	5.3%	10.3%	6.9%	5.0%	2.5%	31.0%	17.2%
Education	14.6%	9.8%	6.5%	3.2%	2.6%	5.3%	20.7%	10.3%	50.0%	42.5%	20.7%	17.2%

QR-Code Only versus Paper + QR-Code

The sample was randomly stratified based on two different methods of administering the questionnaire: in one group, respondents were required to complete an online form (with a cover letter containing only a QR code); in the other, respondents received both a QR code and a paper version, allowing them to choose between the online and print formats. Table 7 illustrates how sensitive the response rates are to the mode of questionnaire administration.

Table 7 Influence of contact mode on response rate in the three pilot samples (2024–2025).

	Psychiatry	Rehabilitation	Acute Care
Year of data collection	2025	2025	2024
Overall response rate	14.5%	27.1%	28.1%
QR-code only	10.7%	16.5%	19.7%
Print + QR-code	18.3%	37.9%	36.6%

Across all three samples, the response rate decreased by half when participants were forced to use the QR-Code. This suggests that giving respondents the freedom to choose helps maintain response rates at levels comparable to those observed in previous PatZu campaigns (see Table 5), despite the current questionnaire being considerably longer.

This trend also holds in the psychiatry sample, although response rates in this domain are significantly lower than in the other two, a well-documented challenge in a field where questionnaire administration is inherently more complex.

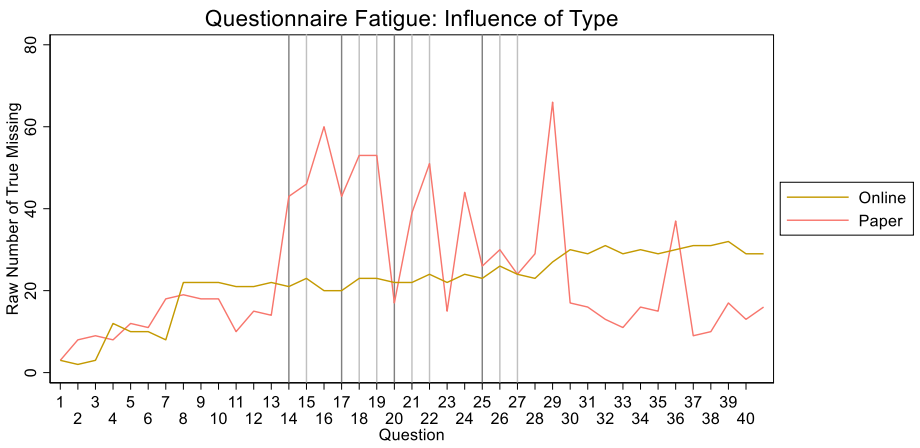
Online Forms versus Paper Questionnaires

While the preceding section examines how various approaches to addressing respondents may impact data quality, this section focuses on the influence of questionnaire format (paper versus online) on respondents' behavior. Although this distinction is closely related to the previous one, the formats were intermixed in the results of the Paper + QR-Code subsample. Here, we explore how the format itself may affect answers.

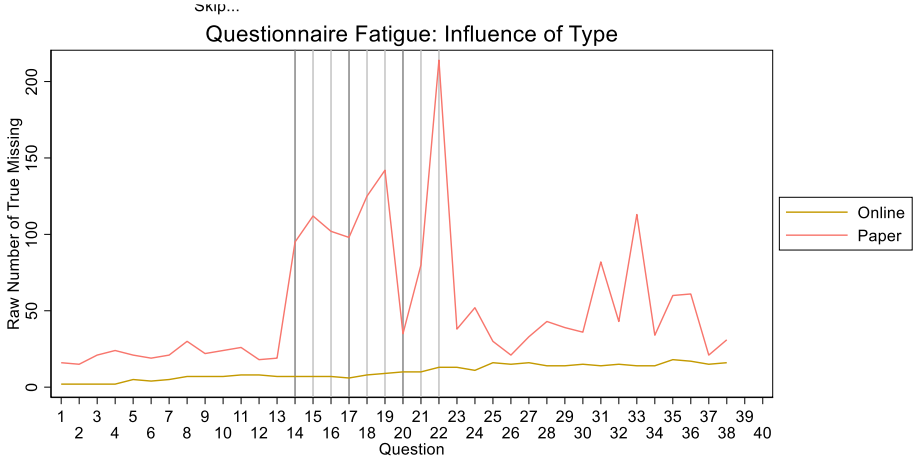
Population characteristics differ between respondents who used the paper questionnaire and those who completed the online form. However, the most notable distinction is a layout-related issue. The online form guided respondents through the structure, minimizing misunderstandings, whereas the paper format required respondents to interpret and follow instructions on their own. Figure 13 illustrates how such issues consistently emerged in paper questionnaires across all care domains. Online respondents exhibited higher rates of alternative responses (i.e., "question skipped"), while those using the paper format were more likely to overlook filter questions and unintentionally leave items unanswered, resulting in increased missing data.

Figure 13 Impact of questionnaire format on respondents' comprehension of skippable questions. Vertical lines denote questions that may be skipped, while "truly" missing values also encompass those not marked as skippable following the filter question.

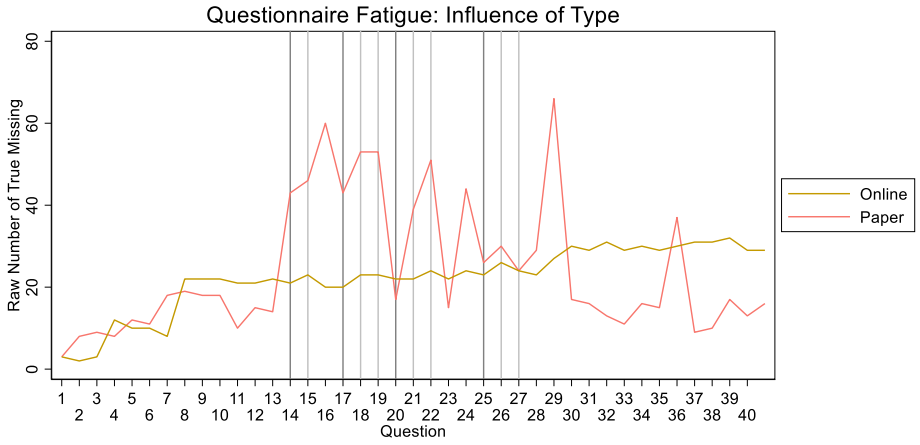
Psychiatry



Rehabilitation



Acute Care

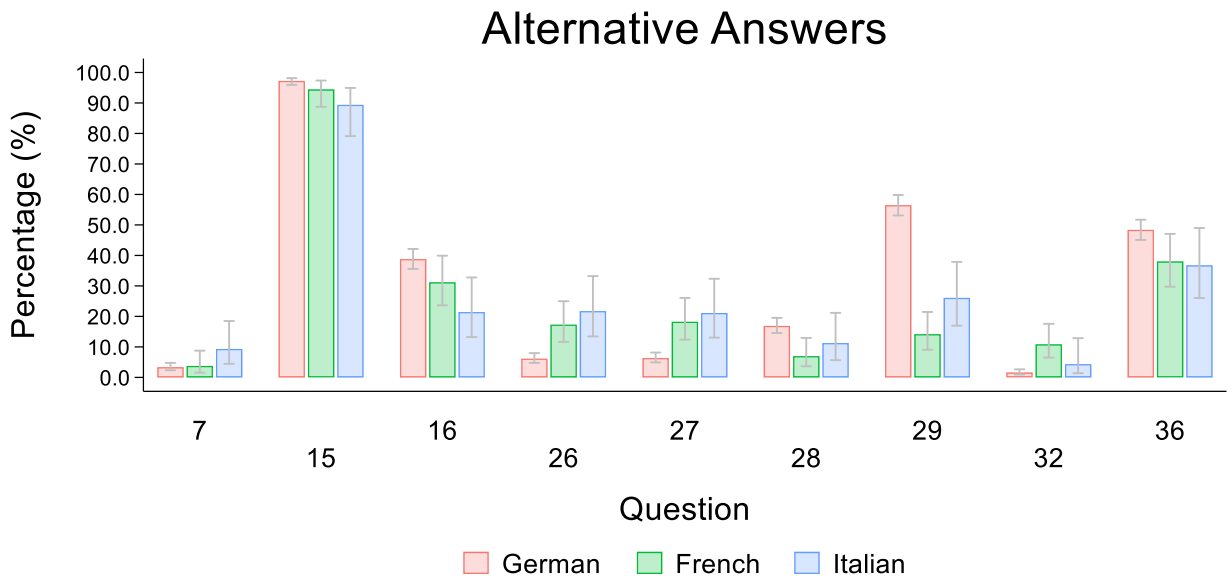


Translations

Difference measured through different versions of the questionnaire can be due to an inadequate translation, cultural differences between linguistic regions, or simply fluctuations resulting from smaller sample sizes in the French- and Italian-speaking parts.

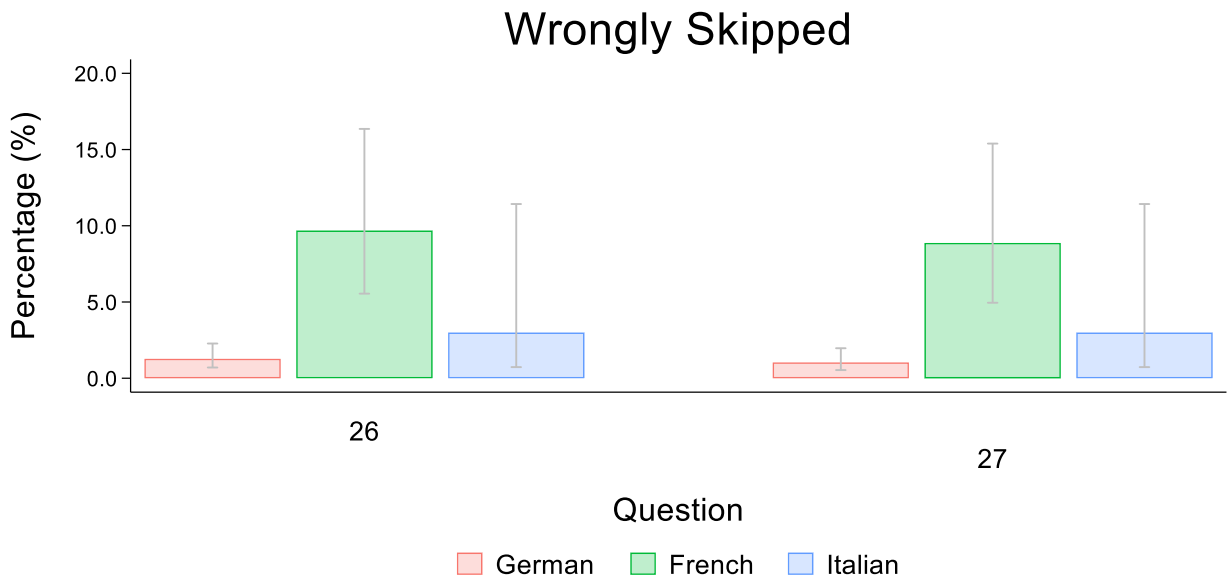
Noticeable differences appear essentially on question 29 in the psychiatry questionnaire (whether the clinic staff provided clear explanations for the reasons behind the restriction of freedom of movement), where much higher proportion of alternative answers (“does not apply”) was found in the German questionnaires (see Figure 14).

Figure 14 Impact of questionnaire language on the frequency of alternative responses in the psychiatry pilot sample.



On two questions on admission, the filter question was overseen more of often in the French pilot sample in psychiatry, see Figure 15.

Figure 15 Impact of questionnaire language on the frequency of skipped questions in the psychiatry pilot sample.



Beside these two points, no major differences between the three linguistic version were observed that would indicate there is a problem with the translations.

4.2. Shortened Questionnaires

4.2.1. Summary

The absence of evidence for questionnaire fatigue, along the usefulness of PREMs indicators in informing quality management, supported a conservative approach to shortening efforts. Approximately ten questions were removed from the acute care questionnaire. However, due to the inclusion of additional domain-specific questions in psychiatry and rehabilitation, the reduction was more modest, as shown in Table 8.

Table 8 Length of Swiss PREMs questionnaires measured in number of questions before and after shortening, in all domains of care. Note that binary questions about admission, whether involuntary (psychiatry), following an accident (rehabilitation) or through emergency (acute care), may be categorized differently depending on their intended use.

	Psychiatry		Rehabilitation		Acute Care	
	Original	Final	Original	Final	Original	Final
PREMs questions	33	31	31	29	32	28
Filter questions	4	3	3	2	4	3
Global satisfaction questions	4	2	4	2	4	2
Socio-demographic factors	6	4	6	5	6	4
Total	**47**	**40**	**44**	**38**	**46**	**37**

The acute care questionnaire was the first to undergo reduction efforts. The feasibility of applying the same procedure to psychiatry and rehabilitation was then examined. However, due to structural differences from the acute care version, some adaptations were necessary. Table 46 in the appendix shows the final structure of all three questionnaires after shortening.

4.2.2. Acute Care

The acute care questionnaire used in the pilot study originally comprised 40 questions, excluding six additional socio-demographic items at the end (see Section 0 in the appendices). Following an analysis of the pilot data and discussions with the QA PatZu, the final version was reduced to 33 questions, accompanied by four socio-demographic factors describing the respondent, totalizing 37 questions (nine fewer than in its original format).

Questions in the questionnaires vary in both nature and function. As a result, they require differentiated consideration and treatment. To address this, the ANQ and Unisanté organized them into distinct blocks based on their potential for shortening. These groupings were then submitted to the QA PatZu to determine which items should be retained or removed. The following sections describe the decisions made for each block.

Questionnaire Fear and Fatigue

The low impact on response rates compared to previous PatZu questionnaires has already been discussed in Section 0. Questionnaire fatigue appears to be low either, as indicated by the online curve in Figure 11. The steady rise in the online curve allows us to estimate an average loss of two respondents per additional question. For example, with a 40-question questionnaire, we would expect to lose roughly 80 respondents, or approximately 3.5% of the 2 281 participants in the subsample that completed the online form.

As discussed in Section 4.1.8 which focuses on the (mis)understanding of filter questions, most missing values stem from the structure and layout of the paper versions rather than from respondent dropout.

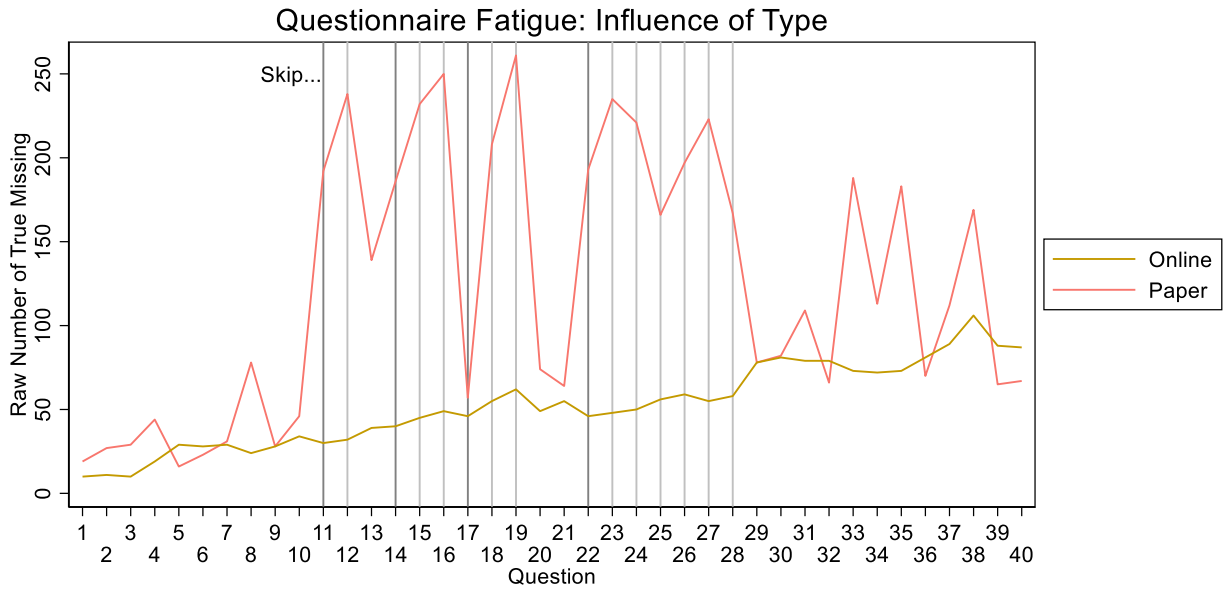


Figure 16 Comparison of missing values increase between respondents using the paper or online versions.

Admission Section

The original CPES-IC questionnaire from 2019 included a seven-question section focused on patient admission. A filter question further divided this section into two subsections, targeting patients admitted via either the emergency or elective route.

The issue with this structure is that it separates the sample into two mutually exclusive groups, resulting in no single respondent answering all the questions. This poses a challenge for multivariate analyses, involving variables that are not compatible across groups.

In 2024, a reduced version of the CPES-IC was introduced, consisting of just two general questions about the patient's admission experience, regardless of their care pathway, see Table 9.

Adopting this shorter CPES-IC format would lead to a substantial reduction, by up to four questions, and a welcome simplification of the questionnaire's structure. Respondents would no longer need to skip questions, and data analysis would not require separate models or imputations, all without any meaningful loss of information. However, filter question 22, which distinguishes between emergency and elective admissions, was considered essential and retained alongside the two revised admission questions.

The Swiss PREMs on acute care questionnaire could therefore be shortened by five questions in the admission section alone.

Table 9 Longer and shorter versions of the CPES-IC section on admission.

Original CPES-IC 2019		CPES-IC 20M 2024	
Nr.	Question	Nr.	Question
22 (skip question)	*When you arrived at the hospital, did you go to the emergency department?*		
23	Before coming to the hospital, did you have enough information about what was going to happen during the admission process?		
24	Was your admission into the hospital organized?		
25	When you were in the emergency department, did you get enough information about your condition and treatment?	20	During your admission, was the information you received about where to go in the hospital easy to understand?
26	Were you given enough information about what was going to happen during your admission to the hospital?	21	It's not uncommon to have to wait while being admitted to the hospital. While waiting to be admitted, were you kept informed?
27	After you knew that you needed to be admitted to a hospital bed, did you have to wait too long before getting there?		
28	Was your transfer from the emergency department into a hospital bed organized?		

Filter Questions

Most filter questions require skipping several items (such as in the admission section) while a few affect only the question that follows immediately, see Table 10. Due to issues caused by misunderstood filter questions in the paper questionnaire (see discussion on paper versus online layout in Section 4.1.9), the QA PatZu approved the suggestion to replace filter questions with corresponding alternative response options, where simple configurations allowed. This resulted in an additional reduction within the acute care domain.

Table 10 Simple filter question in the acute care questionnaire.

Original CPES-IC 2019		Swiss PREMs Acute Care	
Nr.	Question	Nr.	Alternative Answer
11	During this hospital stay, did you need help from nursing staff or other hospital staff in getting to the bathroom or in using a bedpan?		-
12	How often did you get help within the expected time getting to the bathroom or in using a bedpan?		+ I did not need any help

Global Satisfaction

The four global satisfaction questions (numbers 20, 21, 39, and 40 in the two sections titled "Overall rating"*)* served as gold-standard references for comparing the PREMs questions against patient satisfaction. To address concerns about potential redundancy, only two of the four questions were kept, based on both qualitative and quantitative criteria. Notably, question pairs located along the diagonals of Table 11 may not be redundant, which informed the decision to retain questions 21 (net promoter score) and 40 (patient's overall experience). This revision reduced the questionnaire by two additional items. See also Figure 43 in the appendix for the correlation structure of satisfaction questions according to principal component analysis (PCA).

Table 11 Redundancies between the four global satisfaction questions in the acute care sample. Questions appearing on same line or column may be redundant.

Quantitative Dimensions	Qualitative Dimensions	
Proximity on PCA Plane	Hospital Rating	Personal Experience
Pair 1	21. Recommend this hospital	39. Did your hospital stay help?
Pair 2	20. Rate this hospital	40. Rate your experience

Socio-demographic Questions

Questions describing the respondents were identical to those already used in the PatZu routine measurements, along with two additional items that were ultimately discarded.

The first of these, regarding education level, aimed to gather data with equity considerations in mind. However, the disappointingly—though perhaps unsurprisingly—high rate of missing responses and participants' reluctance to answer made the variable impractical for use in multivariate analyses. Table 12

shows how the proportion of observations not available to analysis (missing or alternative) consistently exceeds 20% in all domains of care.

Table 12 Distribution of self-reported education levels in all domains of care, including missing values and alternative responses.

	Psychiatry		Rehabilitation		Acute Care	
Years of Education	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
1...9 years	132	11.7%	185	11.2%	524	9.9%
10...13 years	407	36.1%	648	39.3%	2 014	38.3%
14... years	360	31.9%	333	20.2%	1 457	27.6%
Missing	100	8.9%	246	14.9%	581	11.0%
Refusal	129	11.4%	238	14.4%	699	13.2%

The final question, which asked respondents how easy they found completing the questionnaire, functioned as a quality control item designed exclusively for the pilot phase and will be omitted from the final version. A majority of respondents consistently chose the most favorable response in all domains of care, as shown in Table 13.

Table 13 Distribution of self-reported acceptability levels in all domains of care, including missing values and alternative responses. Answers to the question “It was easy for me to answer these questions. To what extent does this statement apply to you?”.

	Psychiatry		Rehabilitation		Acute Care	
Years of Education	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Does not apply	16	1.4%	30	1.8%	67	1.3%
Tends not to apply	46	4.1%	51	3.1%	105	2.0%
Tends to Apply	286	25.3%	367	22.2%	1 045	19.8%
Applies	734	65.1%	1 123	68.1%	3 830	72.6%
Missing	46	4.1%	79	4.8%	228	4.3%

This section contributed to the overall reduction of the questionnaire by two more questions.

PREMs Questions

With the exception of the admission-related questions, which, as explained above, were replaced by a simplified version, all PREM questions remained unchanged. The decision not to discard any additional question was based on the following considerations:

- The questionnaire had already been reduced by nine items without affecting the core PREM content.
- Multiple rounds of discussion took place, but no consensus emerged regarding the suitability of any particular question for removal.
- Qualitatively, each question provided information valuable for hospital-level quality management.
- Quantitatively, there was no indication that any question was irrelevant, redundant, or inconsistent with the rest of the questionnaire. Even question 33 (addressing support with pain, anxiety, and emotional distress, which was the most difficult to categorize) was considered meaningful and retained.

Psychometric analyses tended to confirm the robustness of the CPES-IC, which had previously undergone a similar validation process, as had the HCAHPS it is based on. Detail of results are listed in Section 0.

4.2.3.Psychiatry and Rehabilitation

The same strategy of grouping questions by their potential for shortening was applied to the two additional domains of care. The procedure used in acute care needed to be validated for both psychiatry and rehabilitation, and the impact of introducing domain-specific questions was also carefully assessed.

Admission Section

These two questionnaires differ from the acute care version in that all patients are admitted electively, eliminating the need to consider two paths at the admission stage. Consequently, the four questions related to the emergency department were already discarded during the questionnaire design phase.

The original two questions pertaining to elective admissions were retained and not replaced by the newer generic alternatives, as they were deemed more specific and relevant.

Third Staff Section

The QA PatZu expressed the desire to include all professional categories (not just medical and nursing staff but also specialized therapists) in the initial sections of the questionnaire.

There was concern, however, that a third section containing questions similar to the first two might cause confusion or inconvenience for respondents. Figure 17 suggests this is not the case, and that including all professional categories is a valuable enhancement to the questionnaire.

Indeed, when using correlations as an indicator of question similarity, redundancy was more pronounced across question types (second row in Figure 17) than across professional categories (first row in Figure 17). This suggests that eliminating specific types of questions may be more effective than removing an entire

section dedicated to a particular staff group. This interpretation is supported by item response theory, which shows less overlap in item information functions when comparing staff categories than when comparing different items. Hence, professional categories are not redundant, according to PCA and IRT.

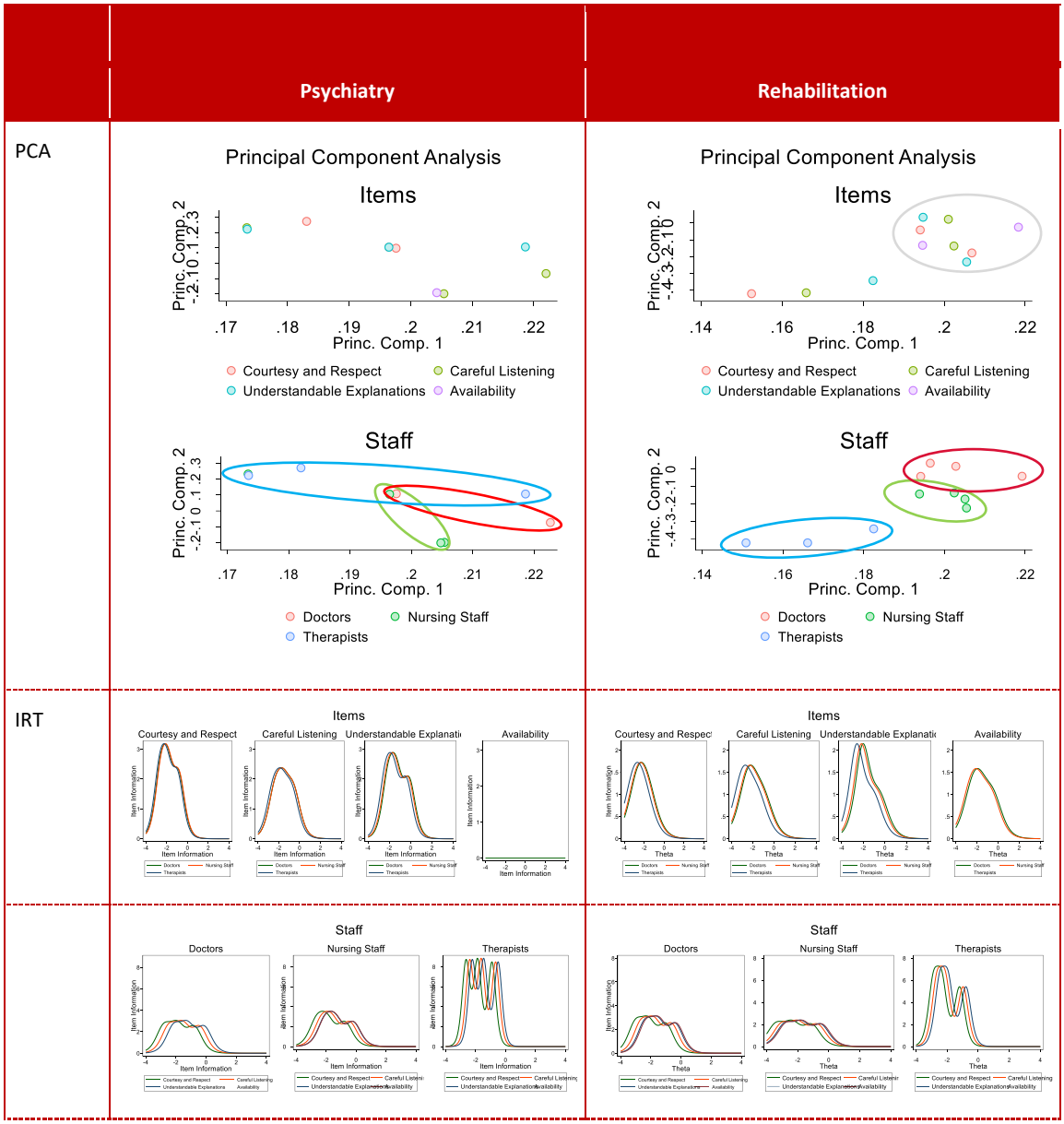


Figure 17 Similarity (based on principal component analysis in the top rows) and contribution (derived from item information functions in the bottom rows) of PREMs items across the three professional categories in the psychiatry and rehabilitation pilot studies.

Filter Questions

Due to the absence of emergency admissions in the psychiatry and rehabilitation domains, these two questionnaires include the two questions intended for elective patients in acute care, without requiring a filter question. However, the QA PatZu, recognizing the importance of admission context for stratified analyses, decided to retain the existing question on whether admission was voluntary in psychiatry, and introduced a new question to determine whether admission in rehabilitation was the result of an accident, see next section.

The remaining filter questions concern assistance with toileting (discussed below) and the two questions related to new medications. The latter were retained as they apply not to a single item but to a pair and adding alternative answers would have been cumbersome.

Domain-Specific Questions

Beside the additional therapist category, the psychiatry questionnaire contains a question about safe spaces (13) and two questions (28 and 29) about restrictions of freedom of movement. There was no evidence of redundancy, those questions providing useful information, as described in the appendix in Section 5.2.2, nor problems with missing values (see Section 4.1.5) or alternative answers (Section 4.1.6). It therefore has been decided to keep all domain-specific questions.

The rehabilitation questionnaire did not contain any domain-specific questions, but it was ultimately decided to add one regarding the type of admission, which had not been included in the pilot version. In the acute care questionnaire, patients are asked whether they were admitted through the emergency department, and in psychiatry, whether they were admitted against their will. Discussions between the ANQ and the QA PatZu highlighted that patients' attitudes towards their healthcare may differ depending on whether their admission followed an accident.

As a result, the rehabilitation questionnaire was supplemented with an "admission type" question, intended to support result stratification in the same way as in the two other care domains. Section 5.3 on numerical adjustments and Section 4.5 on graphical results presentation elaborate on this rationale.

Irrelevant Questions

Two items were considered irrelevant in the psychiatry and rehabilitation contexts: assistance with toileting and pain control. The very high proportion of respondents indicating that the former was not applicable (over 80% in rehabilitation and over 90% in psychiatry, as shown in Figure 9) strongly suggests that it can be safely omitted. This decision was further supported by qualitative reflections indicating that the questions lack relevance in these settings. Consequently, two questions (there was also a filter question) were removed, allowing space for additional, domain-specific items. Similar reasoning, combined with the aim of keeping the questionnaire concise, ultimately led to discard the question on pain control.

4.3. PREMS Scores

4.3.1. Summary

Even when limiting the selection to questions defined by a four-point scale for compatibility purposes, the three Swiss PREMs questionnaires include enough items to cover all eight thematic domains originally identified by the expert committee (see Section 3.4.1).

Consequently, the categorization presented in Table 15 does not follow the internal logic of the questionnaires, which primarily reflect a patient's journey through the healthcare system, but instead consolidates questions from various sections that pertain to one of these eight domains of concern.

This categorization can be uniformly applied across all domains of care, with only minor adjustments for a few domain-specific questions. As a result, comparable summary statistics can be calculated for all questionnaires.

4.3.2. Thematic Categorizations

Four staff members, belonging to ANQ and Unisanté, were asked to assign each PREMs question to one of the seven themes previously defined by the QA PatZu (the initial category “Confidence” was later divided into “Environment” and “Respect”, resulting in one additional theme in the final set). This assignment was based solely on the formulation of the themes and questions, without any knowledge of the pilot data. The resulting grouping, shown in Table 14, served as a starting point for evaluating the validity of the categorization using psychometric criteria. Since this procedure was caried out early on, the items belong to the original acute care questionnaire, prior to its shortening.

In addition to a few sub-optimal alpha coefficients, an issue is the imbalance in the number of items supporting each theme. Themes such as *Organization* and *Information* are densely populated, whereas *Confidence* is built on only two items, which may be insufficient.

Table 14 Preliminary construction of themes was based solely on qualitative considerations. The upper block is mandatory, while the lower block is optional. The items (questions) composing each scale (theme) are listed along with their corresponding Cronbach’s alpha statistics in the columns on the right.

Theme	Questions	Number of items in scale	Average interitem covariance	Scale reliability coefficient
Admission	23, 24, 25, 26	4	0.315	0.857
Information	2, 3, 6, 7, 32	5	0.157	0.830
Involvement	34, 35	2	0.361	0.751
Organization	4, 8, 12, 28, 29, 30, 31	7	0.159	0.823
Discharge	36, 37, 38	3	0.319	0.734
Confidence	9, 10	2	0.108	0.407

Medication	13, 15, 16	3	0.284	0.688
Dignity	1, 5, 33	3	0.109	0.650

The categorization ultimately agreed upon by the ANQ, the QA PatZu, and Unisanté is presented in Table 15. These choices were the result of successive trial-and-error efforts to assign items to thematic categories, taking into account both quantitative and qualitative aspects. Further details on the descriptive data analyses and the rationale behind these decisions can be found in Sections 0 (Acute Care) and 5.2.3 (Rehabilitation and Psychiatry). Note that the categorization is based solely on the pilot subsample. It will be reconsidered during the first nationwide campaign in the light of a comprehensive sample encompassing all participating hospitals and shortened questionnaires.

Table 15 Final categorization of items by theme across the three domains of care. Numbers correspond to questions from the pilot questionnaires.

Theme	Item	Acute Care	Psychiatry	Rehabilitation
Admission	Information on admission	23	26	25
	Admission organization	24	27	26
	Information on health cond. (emerg.)	25		
	Information on unfolding (emerg.)	26		
Information	Careful listening (doc.)	2	2	2
	Understandable explanations (doc.)	3	3	3
	Careful listening (nurs.)	6	5	6
	Understandable explanations (nurs.)	7	6	7
	Careful listening (ther.)		9	10
	Understandable explanations (ther.)		10	11
	Information on health condition	32	33	30
Involvement	Decisions (pat.)	34	35	32
	Decisions (fam.)	35	36	33
Organization	Rapid help (nurs.)	8	7	8
	Communication between staff	29	30	27
	Treatment information	30	31	28

	Punctuality	31	32	29
Discharge	Understanding medication	36	37	34
	Continuing treatment	37	38	35
	Understanding condition	38	39	36
Environment	Cleanliness	9	11	12
	Calm	10	12	13
	Safe place		13	
New Medication	Purpose	15	18	18
	Side effects	16	19	19
Respect	Courtesy and respect (doc.)	1	1	1
	Courtesy and respect (nurs.)	5	4	5
	Courtesy and respect (ther.)		8	9

The categorization remains consistent across all domains of care: each item is always assigned to the same theme in every questionnaire. Domain-specific questions targeting additional therapists were added alongside similar questions already addressed to doctors and nursing staff. The psychiatry-specific question regarding safe spaces was included under the *Environment* theme, while all questions related to freedom of movement were excluded.

4.3.3. Summary Statistics

There are two key differences between the patient satisfaction questions used in previous PatZu reports and the new Swiss PREMs scores:

- **PREMs scores are no longer based on single questions** but instead are derived from sets of two to seven items, as outlined in Table 15.
- **PREMs items can no longer be summarized using simple averages**, due to the absence of a neutral option in the four-point scale. This necessitates a focus on extreme responses and the use of *top* and *bottom* box metrics, as explained in Section 3.4.4.

Figure 18 illustrates how each theme can be represented by a pair of PREMs scores: one indicating the average proportion of respondents who selected the highest rating (4), and the other showing the proportion who chose the lowest ratings (1–2) across the questions defining that theme.

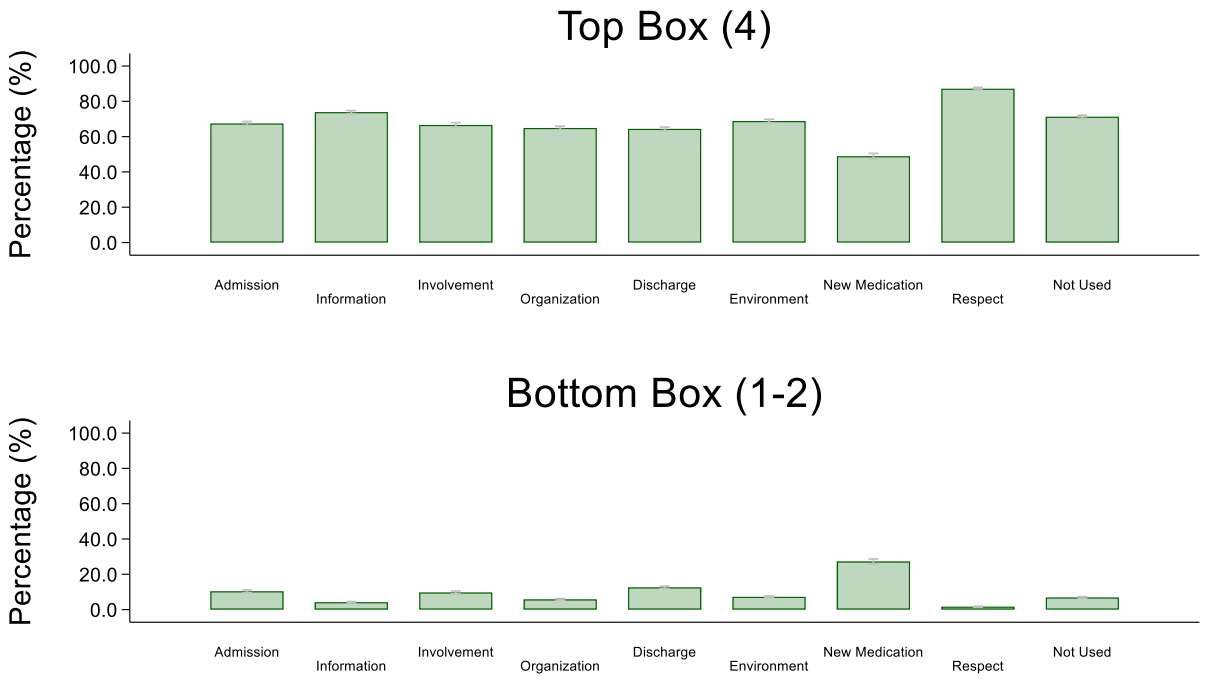


Figure 18 National averages of top and bottom box scores for the eight thematic categories were calculated using the entire Swiss PREMs pilot sample in acute care. The final category, “Not Used”, includes unrelated PREMs questions that could not be assigned to any theme and is not intended for use in hospital comparisons.

The rehabilitation sample (Figure 19) and psychiatry sample (Figure 20) produced similar results.

These summary statistics show consistent results across the different data collection methods. Further details on the potential influence of socio-demographic factors and questionnaire versions are provided at the end of Section 0 (acute care) and 5.2.3 (psychiatry and rehabilitation) in the appendix.

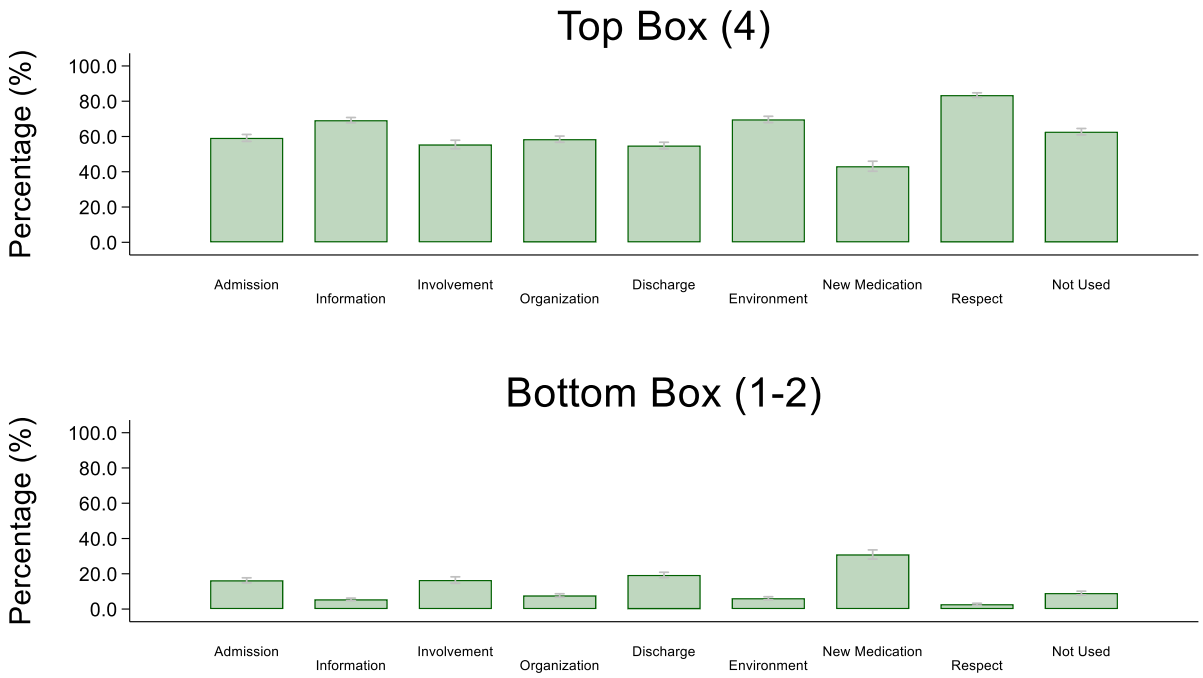


Figure 19 National averages of top and bottom box scores for the eight thematic categories were calculated using the entire Swiss PREMs pilot sample in rehabilitation. The final category, “Not Used”, includes unrelated PREMs questions that could not be assigned to any theme and is not intended for use in hospital comparisons.

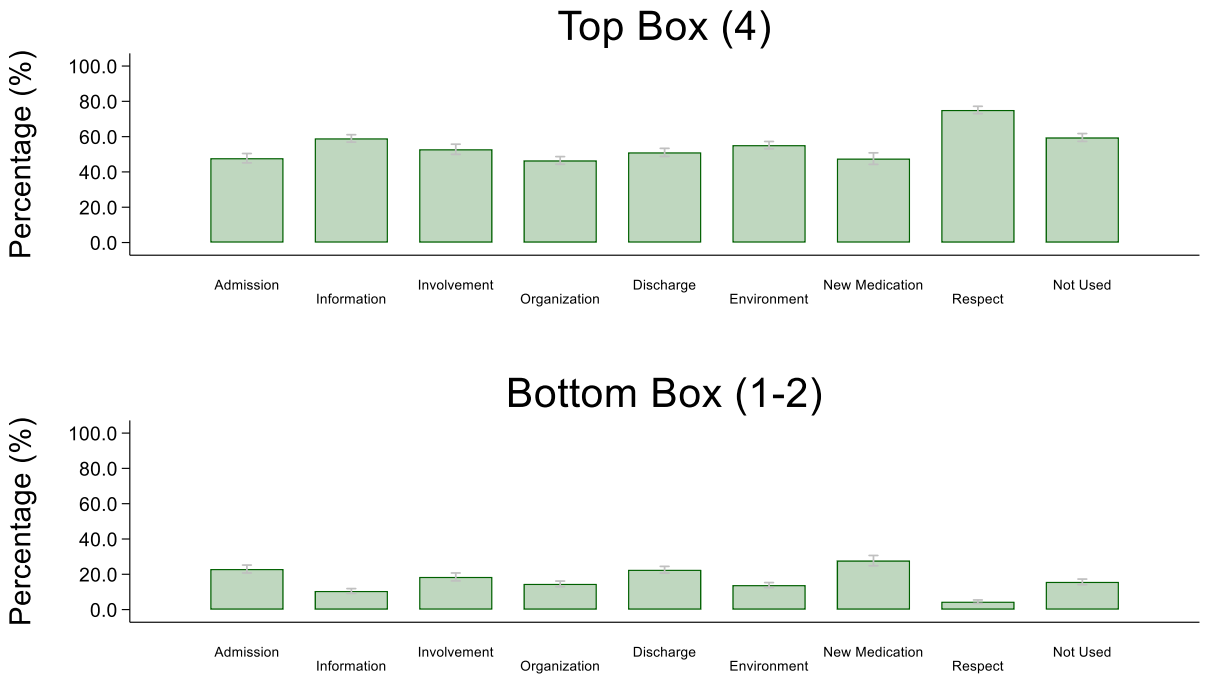


Figure 20 National averages of top and bottom box scores for the eight thematic categories were calculated using the entire Swiss PREMs pilot sample in psychiatry. The final category, "Not Used", includes unrelated PREMs questions that could not be assigned to any theme and is not intended for use in hospital comparisons.

Vertical and Horizontal Averages

As described in Section 3.4.4, thematic scores summarize the responses of multiple individuals to several questions and are therefore derived from two successive averaging steps. There are two different methods for computing such a score:

- **Horizontal approach**: First, calculate a score for each respondent by averaging their answers across the relevant questions. Then, compute the overall score by averaging these individual scores to obtain a hospital-level or national-level score.
- **Vertical approach**: Alternatively, begin by averaging responses across the (sub-)population of interest for each question within the theme. Then, compute the thematic score by averaging these question-level averages.

In the absence of missing values, both methods would yield identical results. However, missing data introduces discrepancies in the denominators used during the computation of intermediate averages, leading to slight differences in the final scores (typically less than 1%, as shown in Table 16).

The item-centered vertical approach (i.e. averaging column-wise questions first) has been adopted by CIHI when publishing results from CPES-IC questionnaires. In contrast, the respondent-centric horizontal approach (i.e., averaging row-wise across individuals first) has been in use in PatZu reports devoted to child and adolescent psychiatry since 2021. In this report, all scores were calculated using the horizontal,

respondent-centric approach. To ensure compatibility with the results shown in hospitals' dashboards, future reports will adopt the vertical approach.

Table 16 Differences in averaging bottom and top box PREMs scores, either vertically (column-wise first) or horizontally (row-wise first).

	Bottom Box				Top Box			
Theme	Vertical	Horizonal	Abs diff.	Rel diff.	Vertical	Horizontal	Abs. diff.	Rel. diff.
Admission	10.90%	10.45%	-0.45%	-4.1%	66.28%	67.15%	0.86%	1.3%
Information	4.20%	4.18%	-0.01%	-0.3%	73.85%	73.86%	0.01%	0.0%
Involvement	9.91%	9.52%	-0.39%	-3.9%	66.73%	67.14%	0.40%	0.6%
Organization	5.76%	5.89%	0.13%	2.3%	65.14%	64.77%	-0.37%	-0.6%
Discharge	12.47%	12.79%	0.32%	2.5%	64.72%	64.05%	-0.66%	-1.0%
Confidence	7.25%	7.25%	0.00%	0.0%	68.81%	68.81%	-0.01%	0.0%
Medication	27.33%	27.13%	-0.19%	-0.7%	48.86%	49.12%	0.27%	0.5%
Dignity	1.68%	1.70%	0.02%	1.2%	87.17%	87.14%	-0.03%	0.0%

Section 4.5 further examines how these metrics can be visualized for national comparisons and how existing plots can be adapted accordingly.

4.4. Adjusted PREMs Scores

4.4.1. Summary

While considering how to compute and display PREMs metrics for national comparisons, the ANQ and the QA PatZu, after several rounds of discussions, chose to follow Unisanté's recommendation to discontinue adjustment of thematic scores when comparing hospitals. This decision was based on several observations:

- Case-mixes as described by the available socio-demographic variables differ individually from hospital to hospital, but differences do not appear between groups of different hospital types.
- Adjusting scores can have a non-negligible impact on a hospital's results, particularly when sample sizes are small. Otherwise, the vast majority of hospital scores is not significantly affected by adjustment.

- Adjusting scores may affect the placement of a small minority of hospitals within or outside the confidence intervals of a funnel plot.
- However, the variables used for adjustment are very limited and may not be the most relevant, and certain confounding factors are absent. This creates a risk of substituting one bias for another, due to residual or unmeasured confounding. Introducing a new variable was not satisfactory either, as seen when asking about education. More relevant variables from the medical domain, such as comorbidities, will remain inaccessible to these questionnaires.

Therefore, a simpler and more transparent approach may be preferred: stratifying results using a variable with a well-established influence on patient experience.

A more detailed presentation of the analysis and conclusions drawn from the psychiatry pilot sample that informed this decision is provided in Section 5.3 of the appendix. Using raw, unadjusted scores offers several advantages:

- The adjustment procedure added an extra step to data preparation and analysis, which delayed the publication of PatZu reports. With raw scores, measurement results can now be released more promptly.
- Discrepancies between raw scores used at the hospital level for quality management and previously adjusted scores used in national comparisons will no longer occur, reducing confusion among readers.
- Eliminating the need for data preprocessing enables flexible computation and visualization on customized dashboards, aligning with the new ANQ publication concept.

4.5. New Propositions for Representing Measurement Results

This chapter outlines an ongoing reflection and dialogue between the ANQ, QA PatZu and Unisanté on the best ways to visually represent PREMs data in the context of hospital quality management and national benchmarking.

4.5.1. Top and Bottom Boxes

The simplest representation of PREMs measurements is provided by national averages (Figure 18 to Figure 20). Although structurally different, their form and function are reminiscent of the national averages for patient satisfaction and the proportion of problematic responses found in the earlier PatZu reports.

4.5.2. Funnel Plots

Each score pair (top and bottom) can be calculated for individual hospitals and displayed on funnel plots, in the same manner as patient satisfaction averages. The only difference lies in the scale: Swiss PREMS uses proportions of highest and lowest ratings (from 0% to 100%), while PatZu employs a continuous scale ranging from 1 (lowest) to 5 (highest).

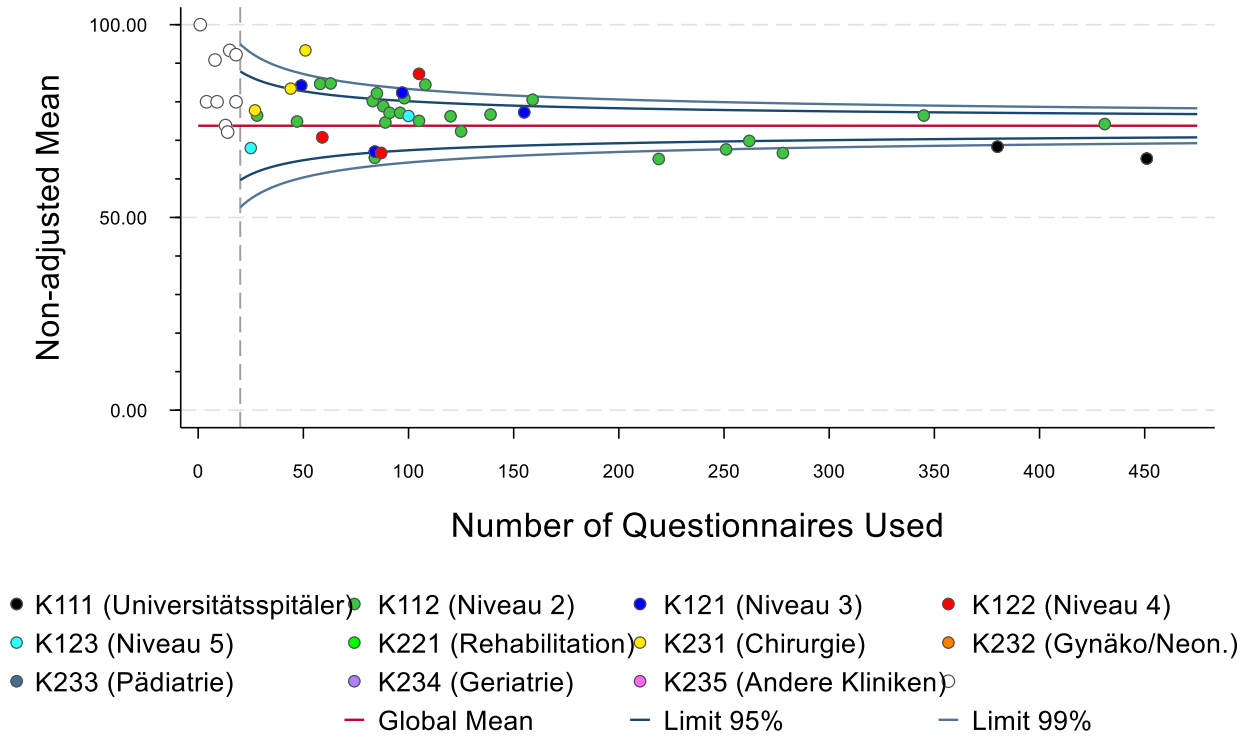


Figure 21 A national comparison (here the top boxes of the *information* theme in acute care) conducted using a funnel plot. Hospitals falling above or below the funnel boundaries display proportions of best responses that significantly deviate from the national average.

Figure 21 illustrates how seamless the transition from average satisfaction to PREMs boxes can be: the only change in the familiar funnel plot is the scale of the y-axis, while all other elements remain unchanged.

4.5.3. Significance Levels

The role of confidence bands in funnel plots is to suggest how the variability of hospital scores decreases as the number of returned questionnaires increases. However, they may also create the impression that the 95% confidence bands serve as a strict threshold for determining whether a hospital falls outside the normal range and thus performs better or worse than the national average.

A more nuanced approach involves using two funnels, one at 95% level and another at the 99% level, to distinguish between minor and more pronounced deviations. This method has the added benefit of retaining more hospitals within the average range, while more effectively identifying those diff significantly, as illustrated in Table and in Figure 21.

Table 17 Hospitals lying inside and outside the funnel at different levels of significance. Example taken from the top box of the *Information* theme in the acute care pilot study.

Confidence Intervals	Hospitals	
+99%	3	6.4%
+95%	10	21.3%
Not significant	27	57.5%
-95%	2	4.2%
-99%	5	10.6%

4.5.4. Stratified Results

An alternative to adjusted results is the use of stratified plots. These display multiple PREMs scores across more comparable subpopulations, or simply provide some insight into how different subpopulations behave, offering richer and more nuanced visualizations. When the stratifying variable is relevant, this approach allows for fairer comparisons, as illustrated in Figure 22.

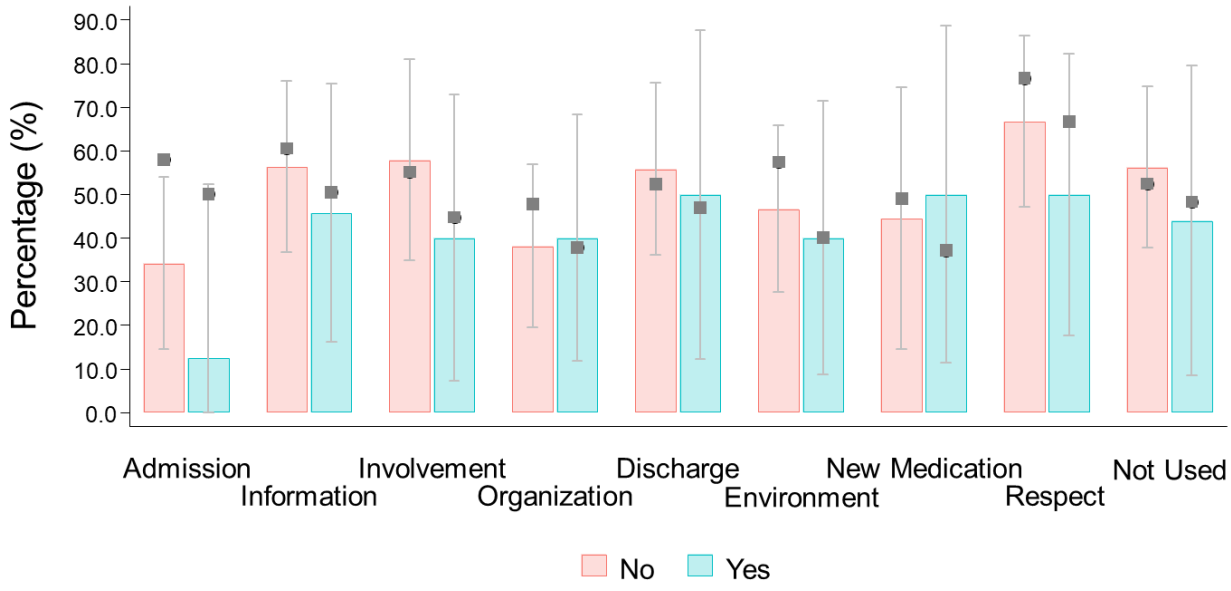


Figure 22 How does a hospital perform against the national average (grey dots)? Example of thematic scores (top boxes) of a given hospital in psychiatry, stratified by placement: forced admission in blue, elective patients in red.

4.5.5.Future Directions

At the time of writing, the ANQ, QA PatZu and Unisanté have not yet determined the final form that the visualization of national comparisons will take. The definitive procedures and choices will be published in the forthcoming *Auswertungskonzept* (analysis framework), which will accompany the new national reports starting in 2026. Nevertheless, the content and form of representation could already be agreed upon.

Data

To keep the set of statistics in national comparisons relatively concise, the focus was placed on the PREMs top boxes, which most closely align with the satisfaction measures readers of previous national reports are accustomed to. The bottom boxes will be available as complementary information through pop-ups.

The eight PREMs top boxes will be supplemented by two global satisfaction top boxes, resulting in a total of ten national PREMs measures. The scales of these two additional questions can be harmonized into top boxes: the Net Promoter Score is already coded on a four-item qualitative scale allowing the same procedure

as all PREMs questions. A top box can also be derived from the 11-item discrete scale of the global experience question by setting a threshold at 9 (an approach chosen by CIHI, see Table 1 on page 6 of CIHI (2022)).

Plots

The funnel plot remains the preferred graph for displaying the relative positions of hospital scores. Their number will increase from six to ten, and they will be based on two funnels, reflecting confidence levels at both 95% and 99%.

While the overall nature and content of these representations is clear, a few details remain to be discussed. Specifically, should confidence bands be displayed, should a point's color indicate whether a hospital's score differs significantly from the national average, or should both features be used? An example of such a funnel-less funnel plot is given in Figure 23.

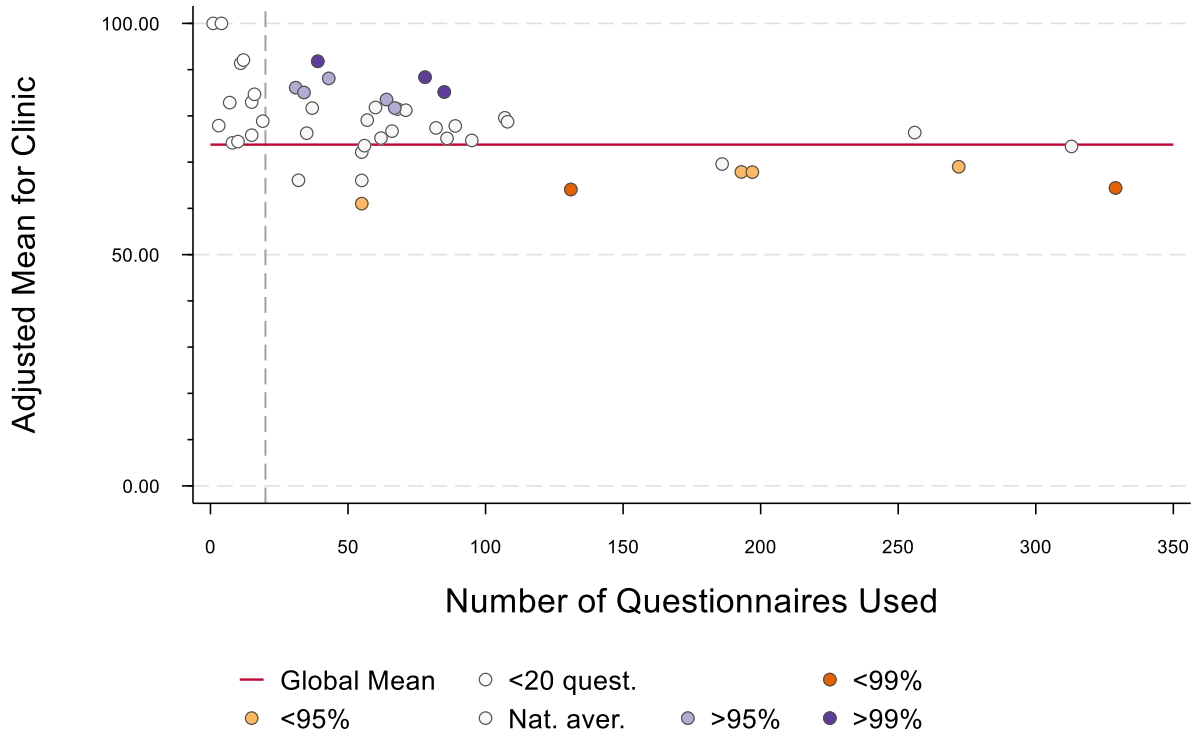


Figure 23 A funnel plot of the information top box in the acute care pilot study, from which the confidence intervals have been removed, replaced by point colorization.

5. Appendix

Chapter 4 highlighted the key findings from the three pilot studies. The following sections offer a more detailed examination of the results that underpin the report's interpretations and conclusions, following the adopted procedure: a thorough analysis of outcomes in the acute care domain, followed by validation in the psychiatry and rehabilitation domains.

5.1. Detail of Results in Acute Care

The following sections expand on those with the same titles in Chapter 4 by offering additional details specific to the acute care pilot sample.

5.1.1. Data Quality Measurements

Sample

The Swiss PREMs Pilot Study sample in acute care includes responses from 5,275 participants who completed the pilot questionnaire. Initially, 5,422 questionnaires were returned, either electronically or by mail. However, 147 of these were excluded from the sample because they were entirely blank (no responses were provided to any of the 40 main questions or the socio-demographic section) rendering them unusable, as shown in Table 3.

The data were collected from patients treated at 47 participating hospitals between the mid-April and mid-May 2024 (see Table 1). These hospitals were selected to represent all three linguistic regions and various hospital categories, with more than half classified as level 2 institutions (larger general care hospitals; see Table 18).

Most hospitals (38, or 80.1%) contributed a subsample size exceeding the threshold of 20 completed questionnaires, which is considered sufficient for conducting data analysis at hospital level. The distribution of language subsamples among hospitals (see Table 18) closely mirrors that of the Swiss population (see Table 3).

Table 18 Hospitals participating in the Swiss PREMs Pilot Studies in acute care, by linguistic region and hospital type.

	Acute Care	
Total Hospitals	47	
In German-speaking part	28	59.6%
In French-speaking part	14	29.8%
In Italian-speaking part	5	10.6%
K111 – University Hospital	2	4.3%

K112 – Medical Center Level 2	26	55.3%
K121 – Medical Center Level 3	4	8.5%
K122 – Primary Care Level 4	4	8.5%
K123 – Primary Care Level 5	4	8.5%
K231 - Surgery	5	10.6%
K235 – Specialized Clinic	2	4.3%

Missing Values

Figure 24 display the distribution of missing values across hospitals and questionnaire items. Questions consistently omitted across all hospitals (appearing as brighter vertical bars) follow misunderstood filter questions, as explained at the beginning of Section 4.2.2. The plot clearly illustrates that this is a general phenomenon observed throughout the entire sample. The yellow horizontal line highlights one hospital with a notably high dropout rate after question 13. However, this is attributed to its extremely small sample size.

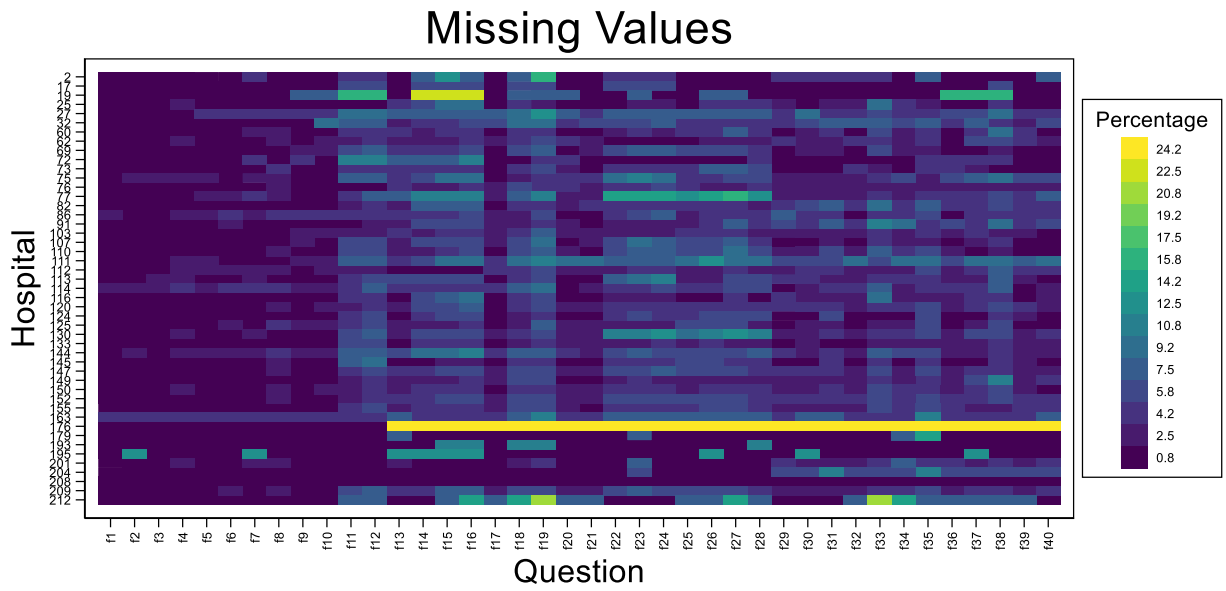


Figure 24 Heat plot of missing values in the acute care pilot study. Each row corresponds to a hospital, each column to a question.

Figure 25 illustrates similarity between questions in terms of missing values. The first principal axis can be interpreted as representing the number of missing values-lower on the right, higher on the left. The figure

shows that proximity within the questionnaire corresponds to similarity in the PCA plane: successive questions tend to be completed or left blank together, notably in the sections that are (incorrectly) bypassed.

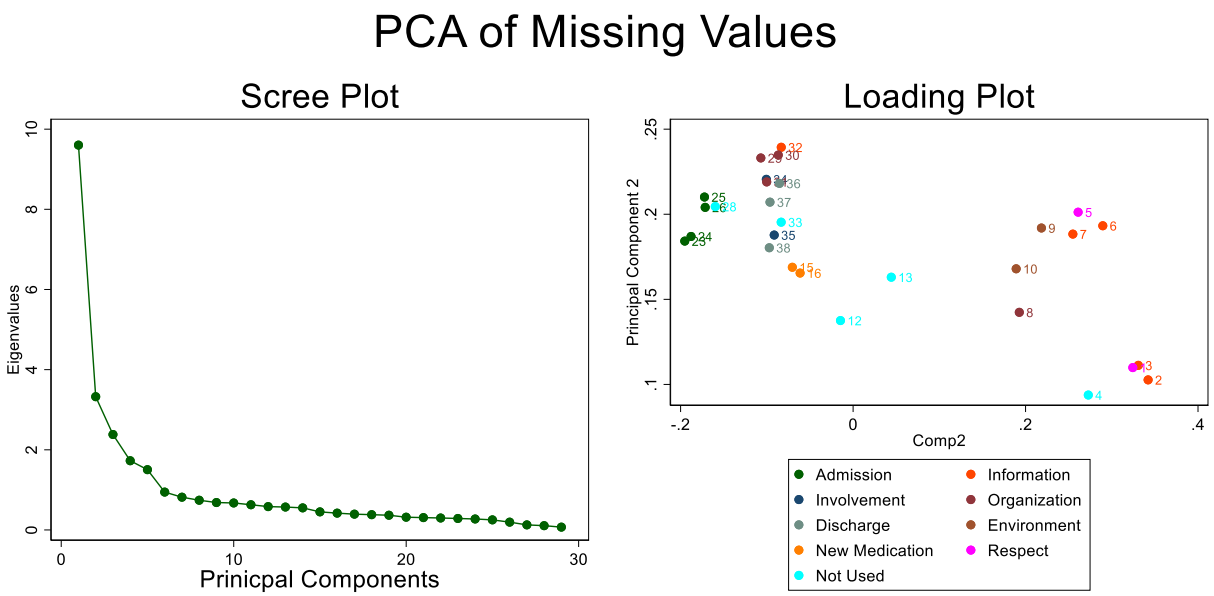


Figure 25 Principal component analysis (PCA) of indicator variables of missing values in the acute care pilot study.

Alternative Answers

Figure 26 shows how questions are consistently skipped across all hospitals in the pilot sample. Brighter vertical bars represent questions with alternative answers all display a uniform color, indicating similar rates. The nearly 100% rate of alternative responses found in the admission section (yellow horizontal lines for questions 22 to 28) suggests that these hospitals likely do not have an emergency department.

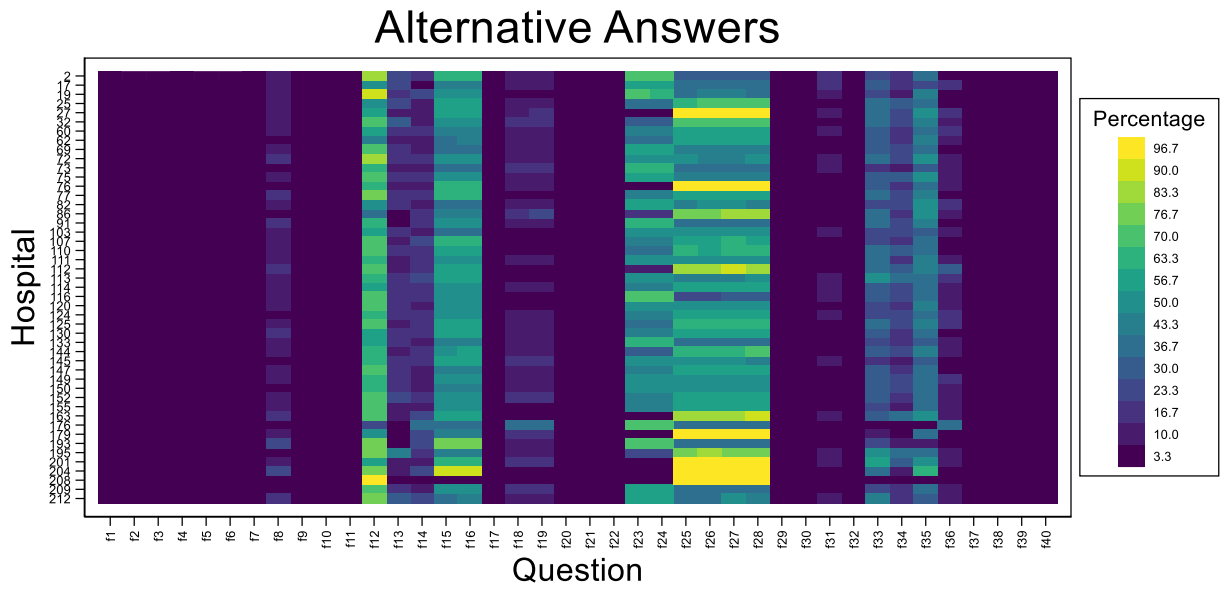


Figure 26 Heat plot of alternative answers in the acute care pilot study. Each row corresponds to a hospital, each column to a question.

The principal component analysis in Figure 27 reveals that the skip pattern for questions addressed to patients admitted through the emergency department form a distinct cluster. This is likely due to the corresponding filter question, which applies to half of the sample. Note that questions targeting elective patients are grouped with other skipped items.

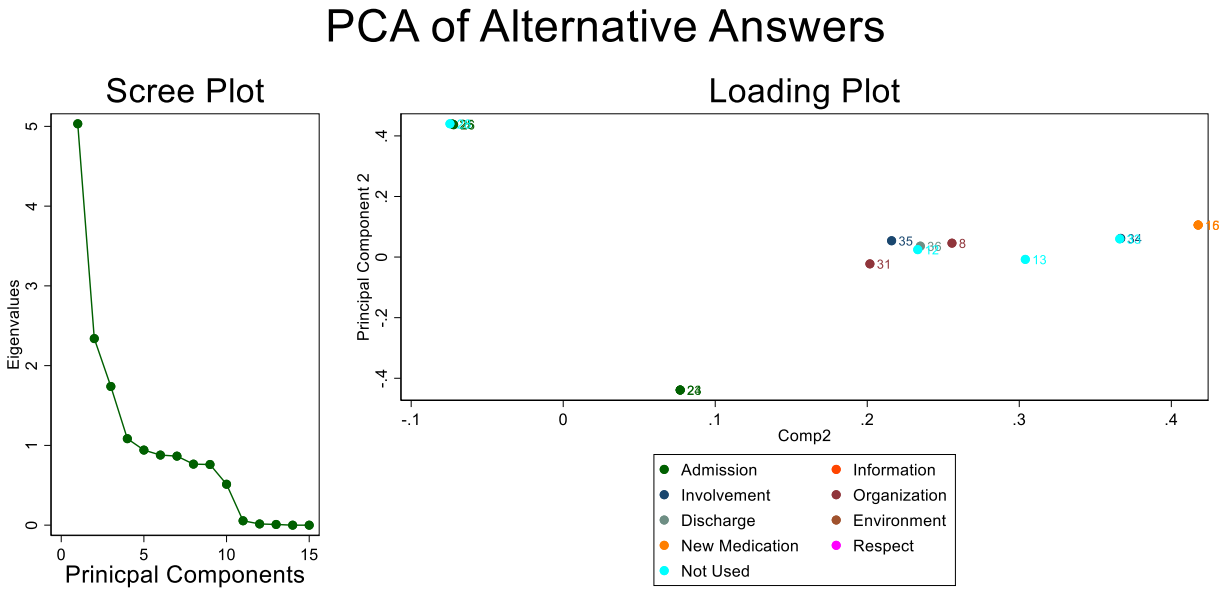


Figure 27 Principal component analysis (PCA) of indicator variables for alternative answers in the acute care pilot study.

Ceiling Effect

Figure 28 displays the distribution of ceiling effects, defined as the proportion of respondents selecting the highest response option for a PREMs question, across hospitals and questionnaire items. Vertical lines appear more prominently than horizontal ones, indicating that questions tend to be answered consistently across hospitals. The brighter lines at the bottom of the plot also contain darker gaps, reflecting greater variability caused by small sample sizes.

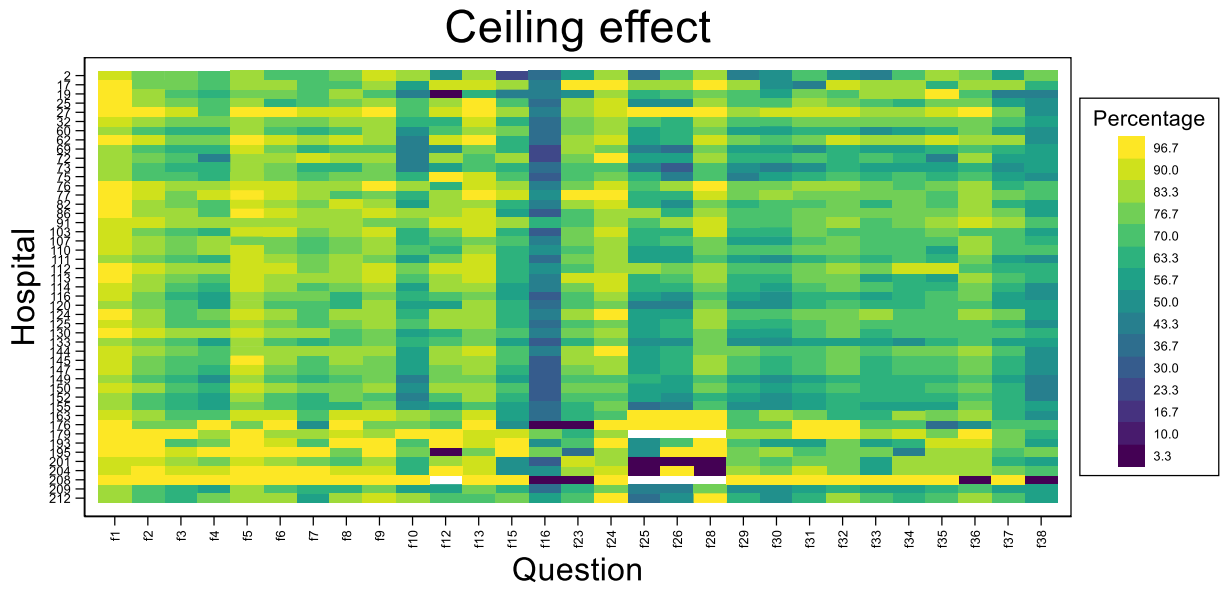


Figure 28 Heat plot of ceiling effects in the acute care pilot study. Each row corresponds to a hospital, each column to a PREMs question.

Figure 29 presents the principal component analysis of respondents' behavior regarding top-box responses. Most items across all themes cluster together, except for question 2 (doctors carefully listening) and question 9 (cleanliness).

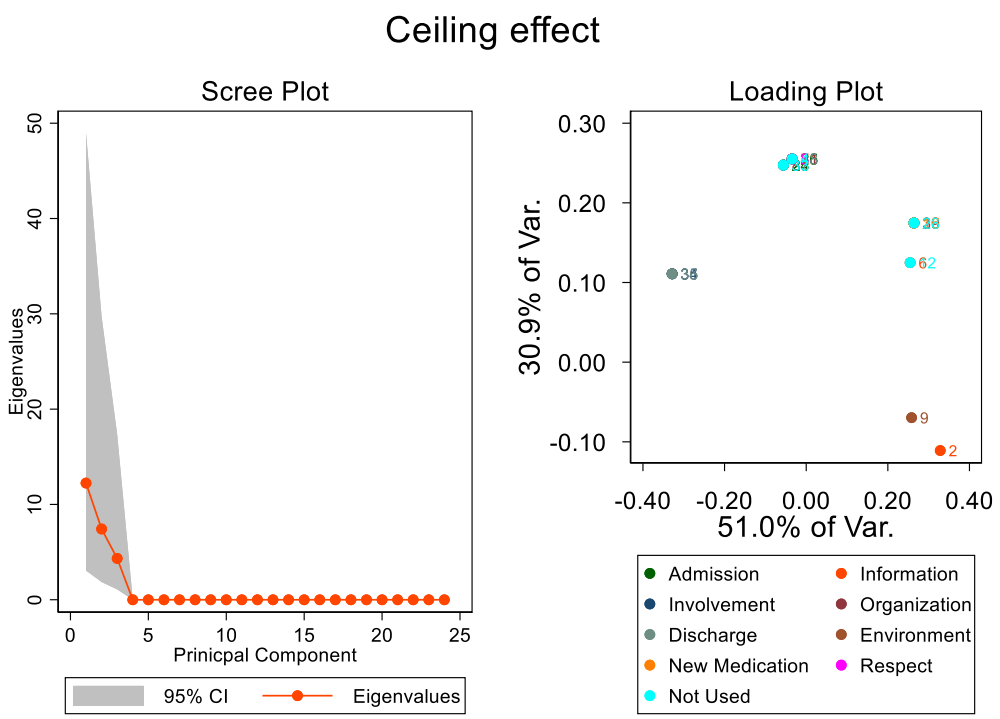


Figure 29 Principal component analysis (PCA) of indicator variables for ceiling effect in the acute care pilot study.

Floor Effect

The floor effect can be assessed by examining the proportion of respondents who select the two lowest response options on PREMs questions. As shown in Figure 28, the vertical bands in Figure 30 confirm that respondents tend to answer each question consistently across hospitals, with generally low rate, except in a few hospitals where the sample size is particularly small.

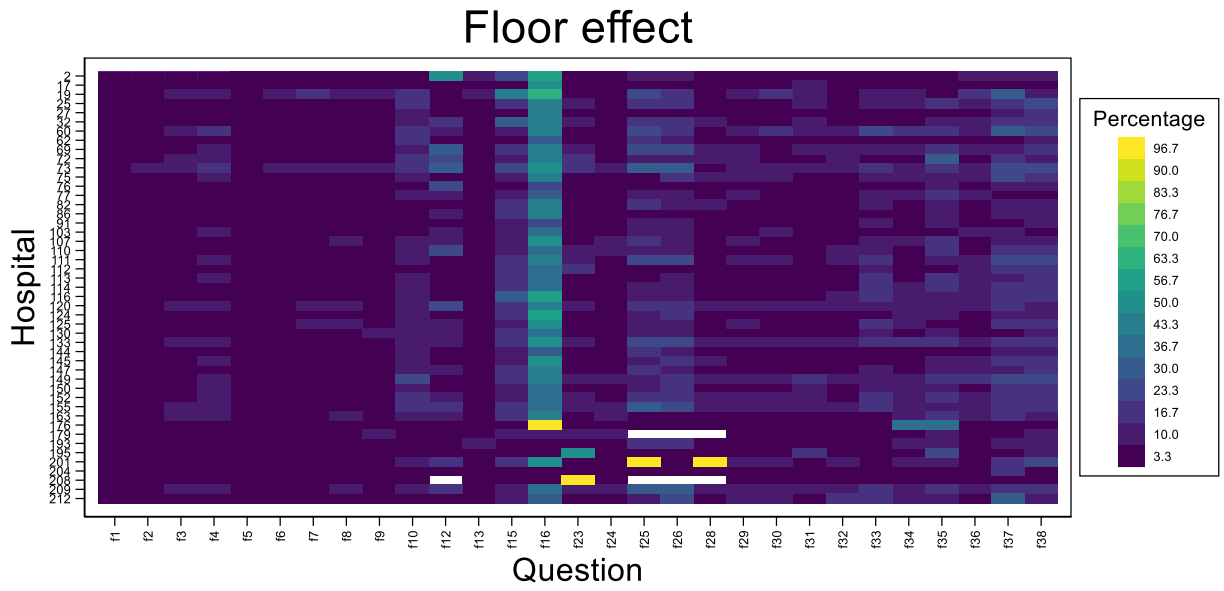


Figure 30 Heat plot of floor effects in the acute care pilot study. Each row corresponds to a hospital, each column to a PREMs question.

All PREMs questions cluster together in Figure 31, except for two attributed to the *Environment* theme: cleanliness (9) and calm (10). Their floor effect rates are not prominent in Figure 11, nor is that question 34 on patient involvement. However, the plot suggests that respondents who select lower answers in the Environment theme are not the same individuals who contribute to lower ratings in other themes.

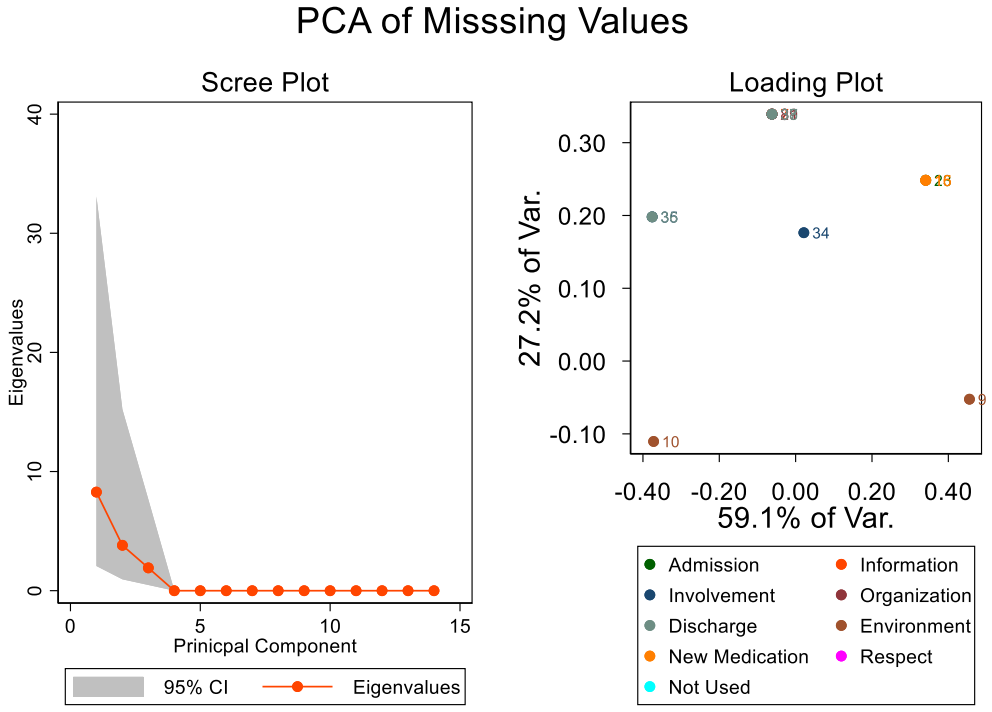


Figure 31 Principal component analysis (PCA) of indicator variables for floor effect in the acute care pilot study.

Influence of Data Collection

To assess the robustness of the acute care questionnaire across its various iterations and data collection methods, all data quality indicators were stratified according to these criteria.

QR-Code Only Versus Paper + QR-Code

Data collection in the Swiss PREMs Pilot Study followed an approach different from previous PatZu measurements. To assess how varying data collection methods influence response rates, the pilot sample was randomly divided into two groups. One group, labeled “QR-code only,” required participants to complete the questionnaire online. The other group, labeled “Print + QR-code,” allowed participants to choose between returning a paper questionnaire or responding online.

Response rates varied significantly between the two groups. Only 19.7% of those in the “QR-code only” group completed the questionnaire: roughly half the 36.6% response rate observed in the “Print + QR-code” group, which is comparable to the rates observed in previous PatZu studies. This pattern was consistent across all participating hospitals (see Figure 32).

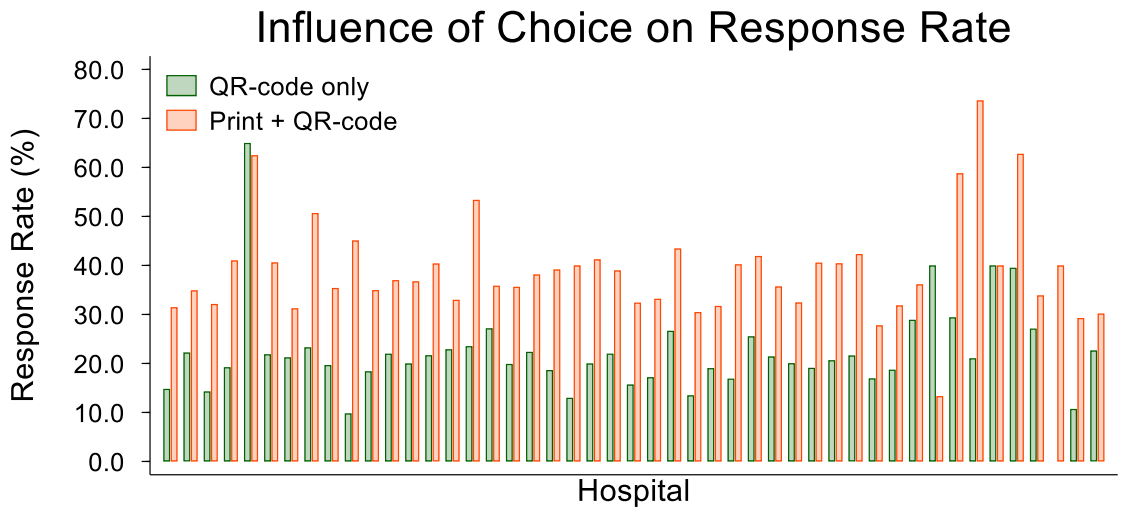


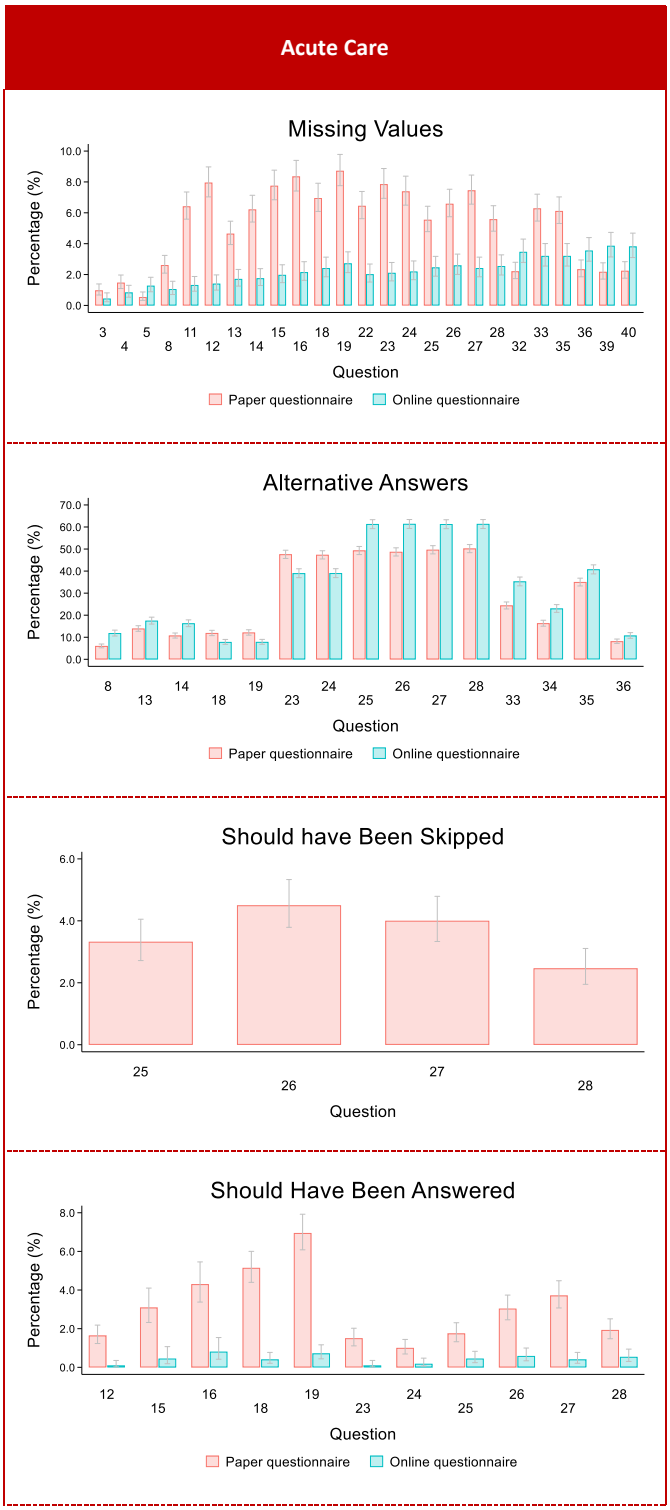
Figure 32 Comparison of response rates across hospitals: Requiring respondents to complete the questionnaire online reduces the response rate by approximately half, a trend consistently observed across all 47 participating hospitals. There are two exceptions: in the Klinik Seeschau, response rates were nearly identical for both data collection methods, while in the Hôpital du Pays-d'Enhaut, the number of returned questionnaires was too low to draw meaningful conclusions. In contrast, the Fondation Rive-Neuve used only paper questionnaires.

The reluctance to use the online form is further evidenced by the low uptake when participants were given a choice: only 12.7% (437 out of 3,431 respondents in the “Print + QR-code” group) chose the online version over the paper questionnaire.

Online Forms Versus Paper Questionnaires

Since respondents in the “Print + QR-code” could choose between the paper and online versions, this stratification combines both formats and cannot be used to distinguish differences between those completing the paper questionnaire and those using the online form. Figure 33 highlights where differences

emerge: the rate of missing values is significantly higher in the paper version, due to misunderstood skip patterns (the online version also prevents respondents from answering questions to skip, as can be seen in the third row of Figure 33). The subsample that completed the online form also consistently selected more often the alternative answers. However, this does not affect the overall experience, as floor and ceiling effects remain comparable.



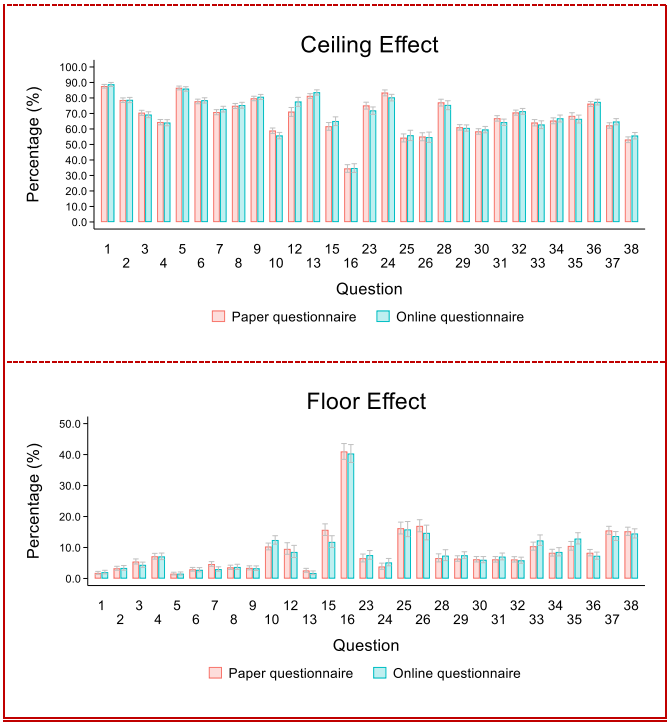
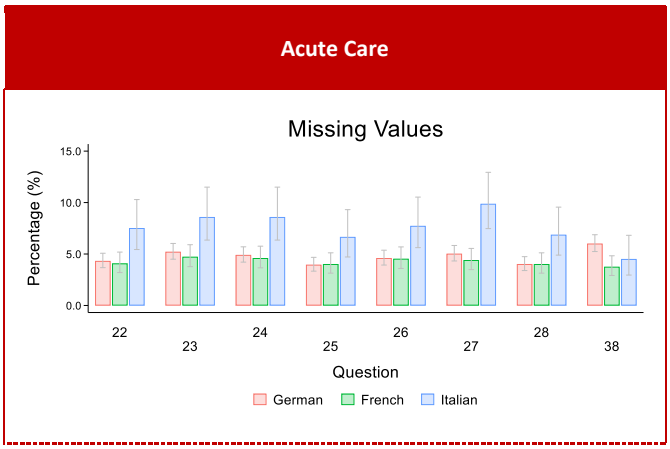


Figure 33 Influence of data collection on data quality indicators in the acute care pilot sample. Only questions showing significant differences are plotted.

Translations

The Italian-speaking subsample diverges from the other two linguistic regions in Figure 34, showing higher proportions of missing values, while the French-speaking subsample tends to rate patient experience lower. This may reflect not so much a bias introduced by translation, but rather greater variability due to smaller sample size and potential cultural differences. French-speaking respondents had already rated hospitals lower in the PatZu studies.



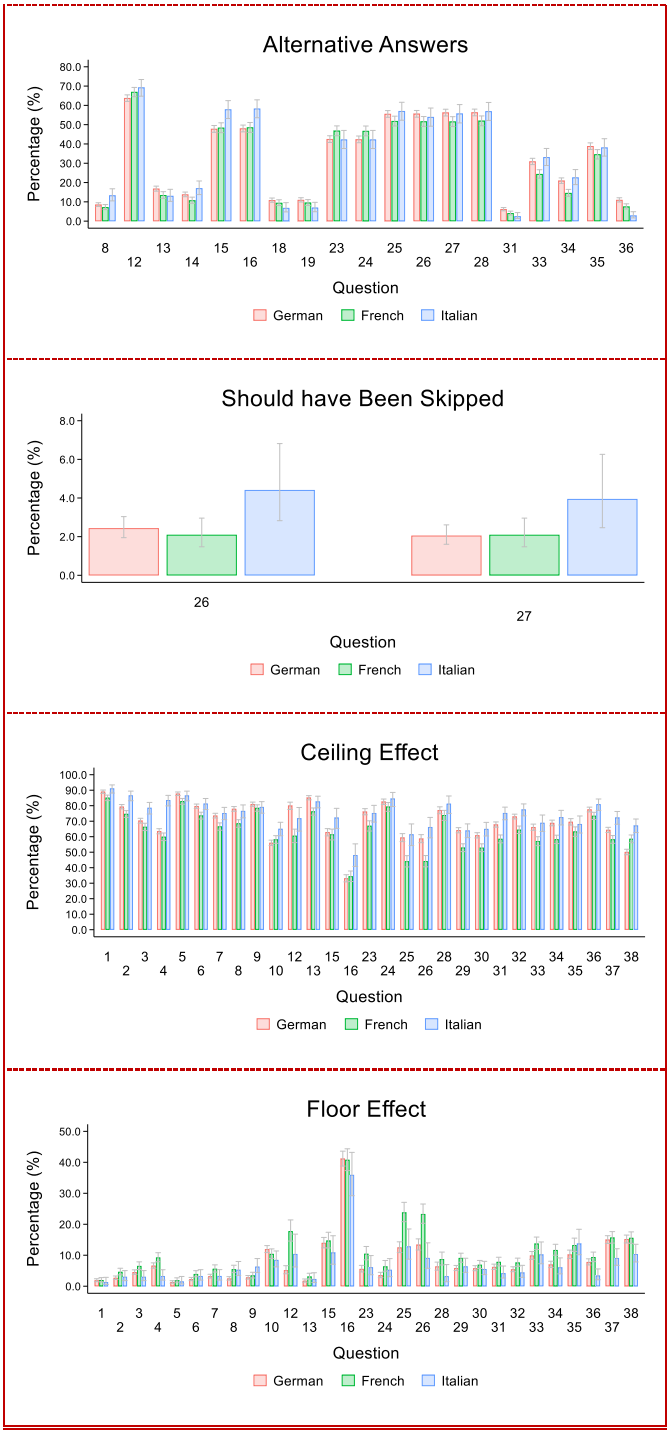


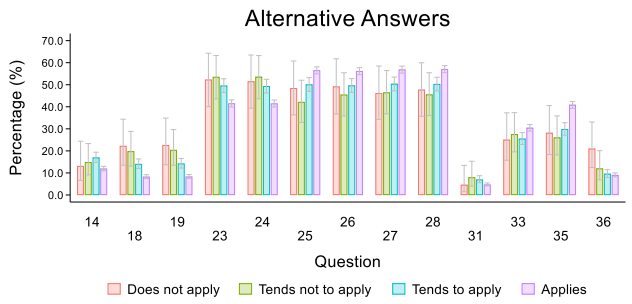
Figure 34 Influence of language on data quality indicators in the acute care pilot sample. Only questions showing significant differences are plotted.

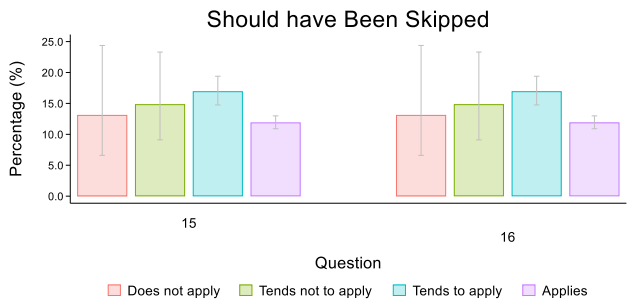
Self-Reported Acceptability

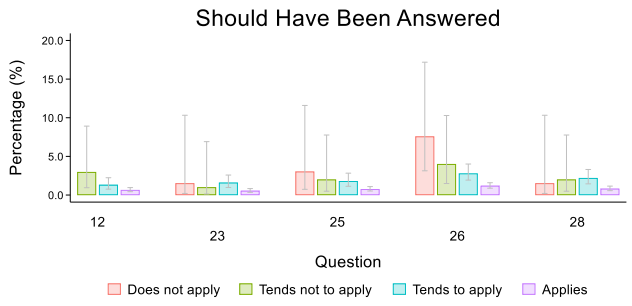
The pilot questionnaire included a final question added for validation purposes (to be removed in the final version), asking respondents to what extent the statement "It was easy for me to complete the questionnaire" applied to them. Figure 35 does not reveal significant differences across levels of acceptability, with variability largely driven by the small number of respondents giving lower ratings. Over 75% selected the highest value, while fewer than 4% chose one of the two lowest options.

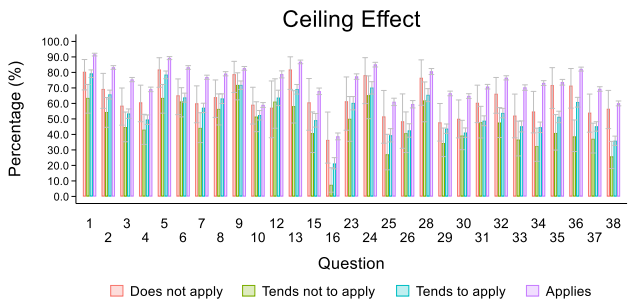
Acute Care


Missing Values
Percentage (%)
20.0
15.0
10.0
5.0
0.0
2 3 4 5 6 7 12 17 25 26 28 30 31 32 35 36 37 38 39 40
Question
Does not apply
Tends not to apply
Tends to apply
Applies


Alternative Answers
Percentage (%)
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
14 18 19 23 24 25 26 27 28 31 33 35 36
Question
Does not apply
Tends not to apply
Tends to apply
Applies


Should have Been Skipped
Percentage (%)
25.0
20.0
15.0
10.0
5.0
0.0
15
16
Question
Does not apply
Tends not to apply
Tends to apply
Applies


Should Have Been Answered
Percentage (%)
20.0
15.0
10.0
5.0
0.0
12 23 25 26 28
Question
Does not apply
Tends not to apply
Tends to apply
Applies


Ceiling Effect
Percentage (%)
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1 2 3 4 5 6 7 8 9 10 12 13 15 16 23 24 25 26 28 29 30 31 32 33 34 35 36 37 38
Question
Does not apply
Tends not to apply
Tends to apply
Applies

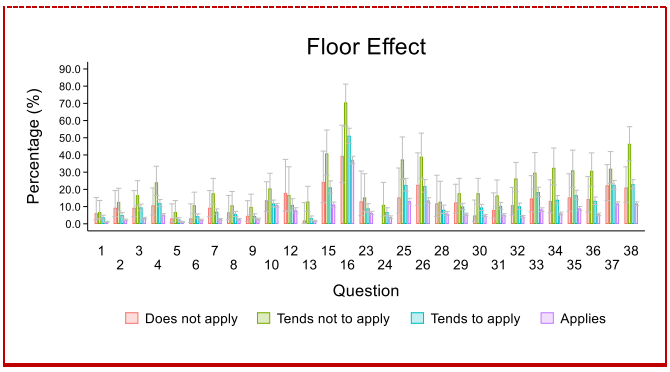


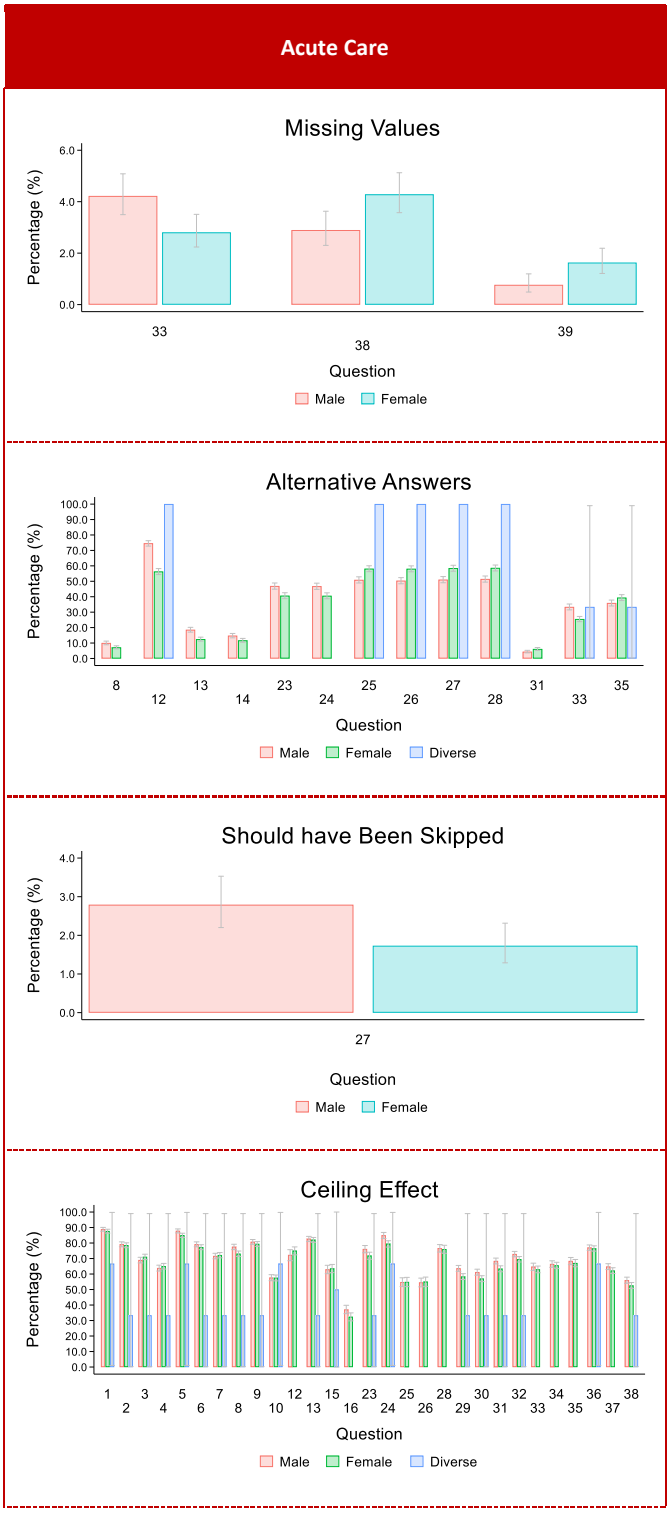
Figure 35 Influence of self-reported acceptability on data quality indicators in the acute care pilot sample. Only questions showing significant differences are plotted.

Influence of Socio-Economic Factors

This section examines which socio-demographic factors may influence data quality and identifies subgroups where differences are likely to emerge.

Gender

Gender-related differences are minimal, as shown in Figure 36. Most apparent divergences stem from the extremely small number of respondents identifying as a diverse gender (only three questionnaires), which prevents any meaningful interpretation.



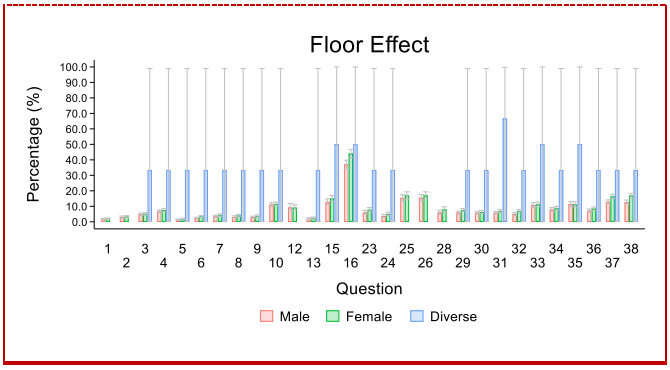
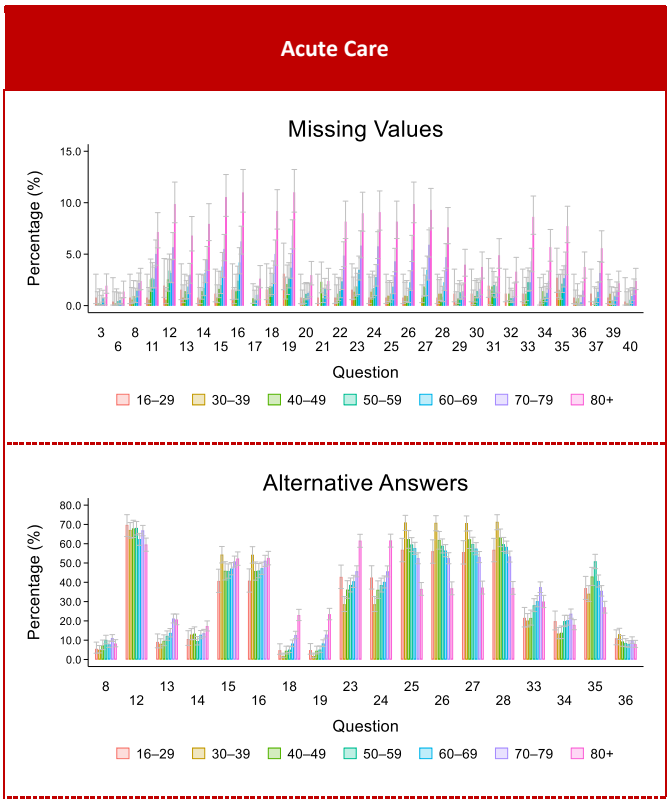


Figure 36 Influence of gender on data quality indicators in the acute care pilot sample. Only questions showing significant differences are plotted.

Age

Age-related trends are evident in Figure 37, particularly in the distribution of missing values, which are largely caused due to misunderstood skip patterns in the paper questionnaire, a format more commonly used by older respondents. Consistent patterns also show that both the youngest and oldest participants tend to rate patient experience lower.



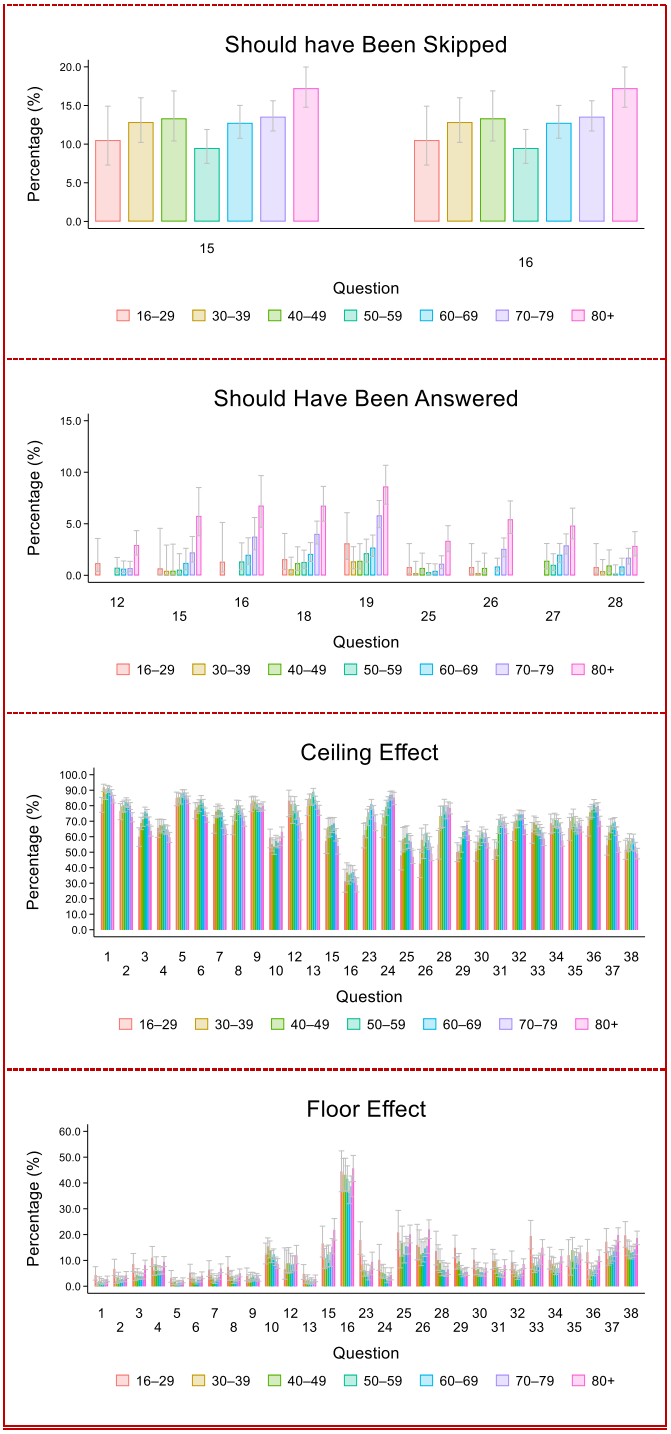


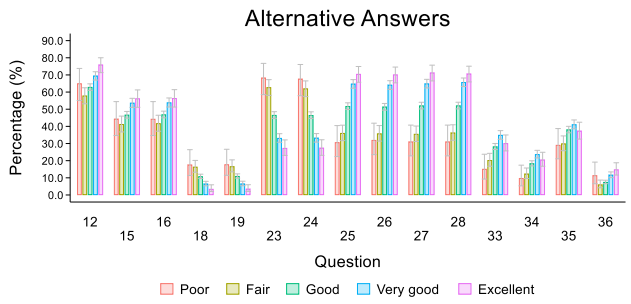
Figure 37 Influence of age on data quality indicators in the acute care pilot sample. Only questions showing significant differences are plotted.

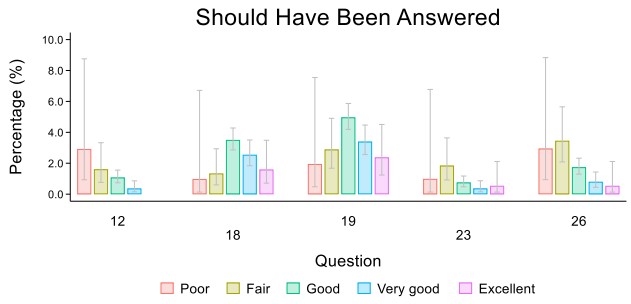
Self-Reported Health

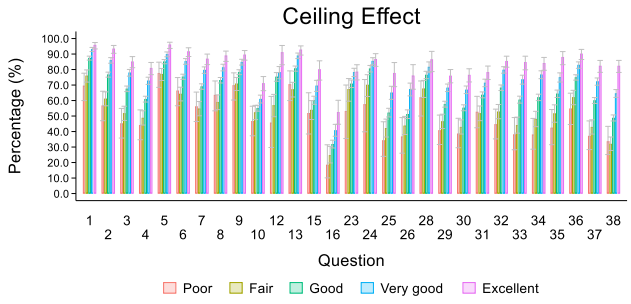
The influence of self-reported health is clearly visible in Figure 38. Patients in better health tend to complete the questionnaire more thoroughly, and gradients emerge in health-related alternative responses. For example, a reduced need for new medication in questions 15 and 16. Patient experience ratings also improve with better health status.

Acute Care


Missing Values
Percentage (%)
15.0
10.0
5.0
0.0
2 3 4 5 6 7 12 20 25 26 29 30 31 32 33 36 40
Question
Poor Fair Good Very good Excellent


Alternative Answers
Percentage (%)
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
12 15 16 18 19 23 24 25 26 27 28 33 34 35 36
Question
Poor Fair Good Very good Excellent


Should Have Been Answered
Percentage (%)
10.0
8.0
6.0
4.0
2.0
0.0
12 18 19 23 26
Question
Poor Fair Good Very good Excellent


Ceiling Effect
Percentage (%)
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1 2 3 4 5 6 7 8 9 10 12 13 15 16 23 24 25 26 28 29 30 31 32 33 34 35 36 37 38
Question
Poor Fair Good Very good Excellent

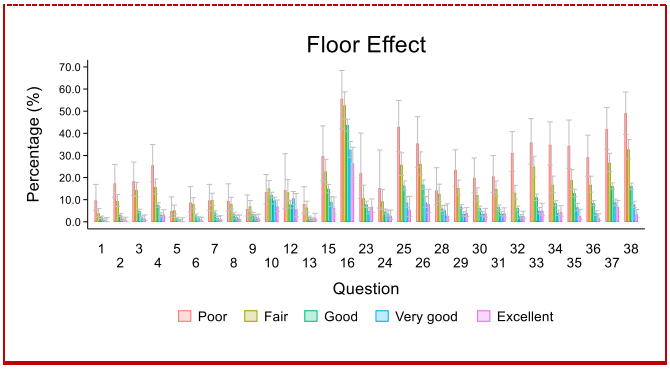
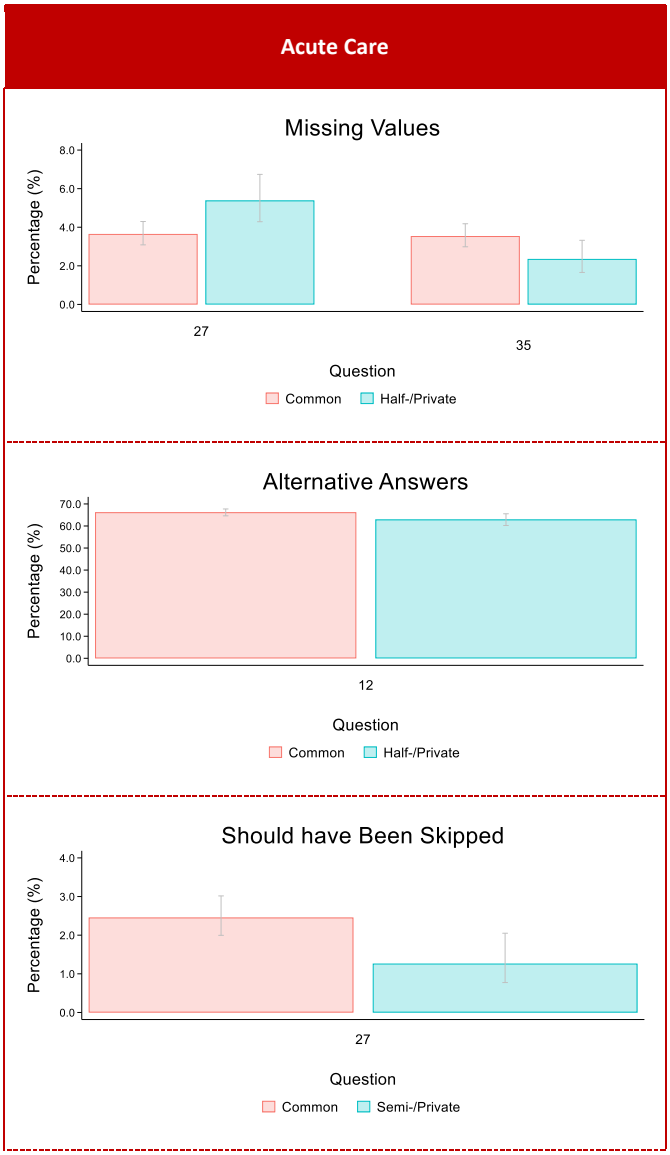


Figure 38 Influence of self-reported health on data quality indicators in the acute care pilot sample. Only questions showing significant differences are plotted.

Insurance Status

Figure 39 shows no evidence that insurance category influences how respondents complete the questionnaire. The only notable difference appears in patient experience ratings: respondents with half-private or private insurance coverage tend to give slightly higher scores.



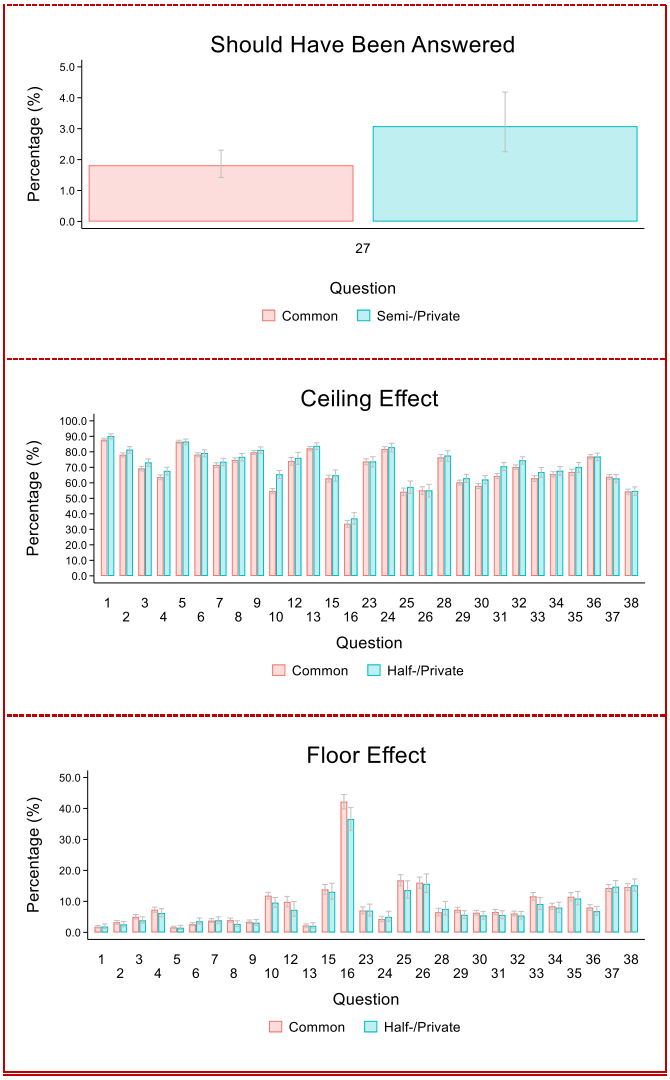
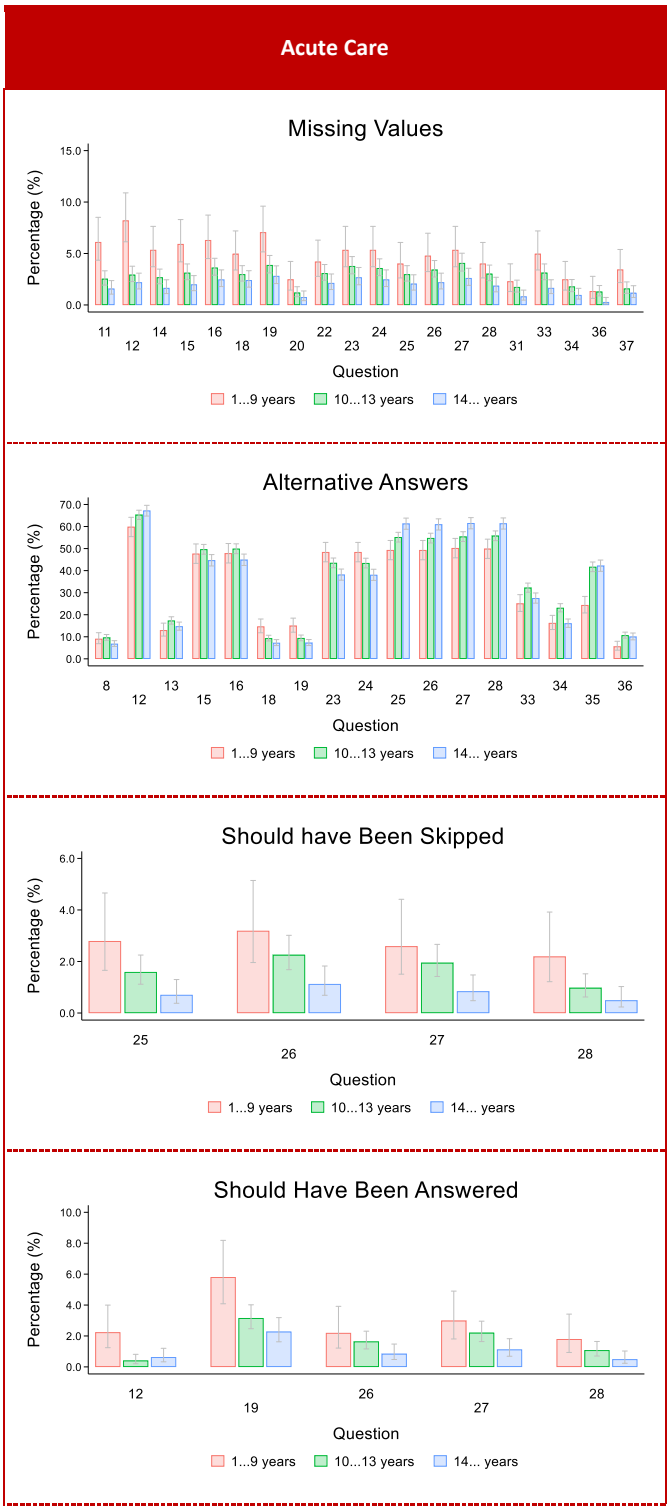


Figure 39 Influence of insurance category on data quality indicators in the acute care pilot sample. Only questions showing significant differences are plotted.

Education

Education level appears to influence the rate of missing values. However, all questions shown in Figure 40 belong to filtered sections, which may again suggest that the missing data stems from misunderstood skip patterns, particularly among respondents with lower education levels, who may be less familiar with more complex questionnaire formats. Simplifying the questionnaire's structure and layout may help address equity concerns.



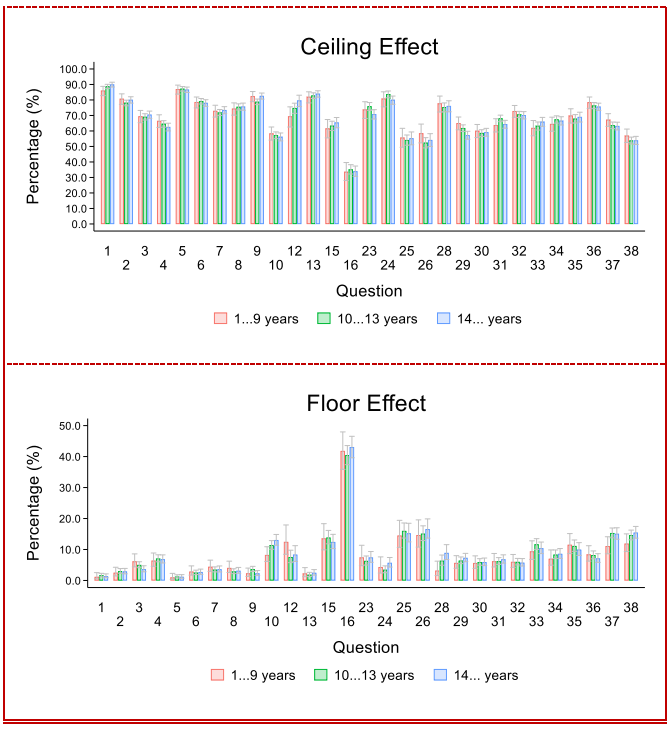


Figure 40 Influence of education level on data quality indicators in the acute care pilot sample. Only questions showing significant differences are plotted.

5.1.2.Questionnaire Shortening

Questionnaire Fatigue

As expected, there is a consistent trend of respondents missing the filter and questions and wrongly skipping questions with increasing age, as shown in Figure 41.

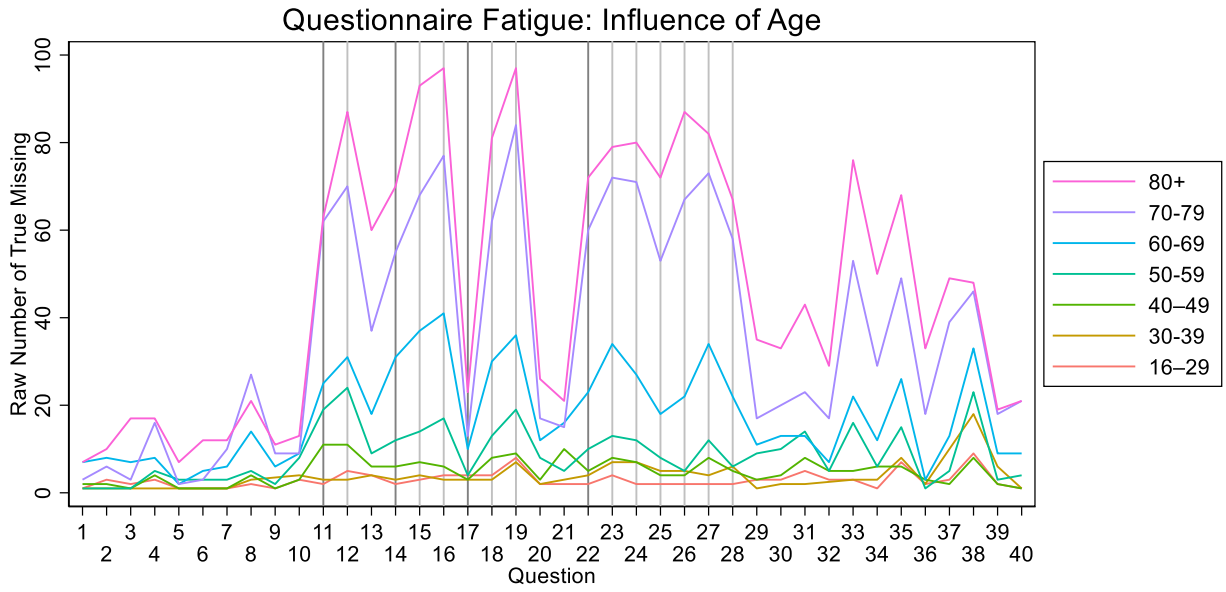


Figure 41 Influence of age on missing filter questions in the acute care pilot study.

Findings from acute care suggest that concerns about questionnaire fatigue and fear are minimal and therefore do not warrant major reductions in length.

Correlation

The heatmap in Figure 42 illustrates the correlation structure among PREMs questions. Brighter rectangles along the diagonal indicate stronger similarity within specific sections of the questionnaire, while brighter off-diagonal areas suggest potential redundancies. However, these rather weak correlations do not provide evidence to justify any shortening on their own in the case of PREMs items. Such insights will be further explored in the next section on global satisfaction.

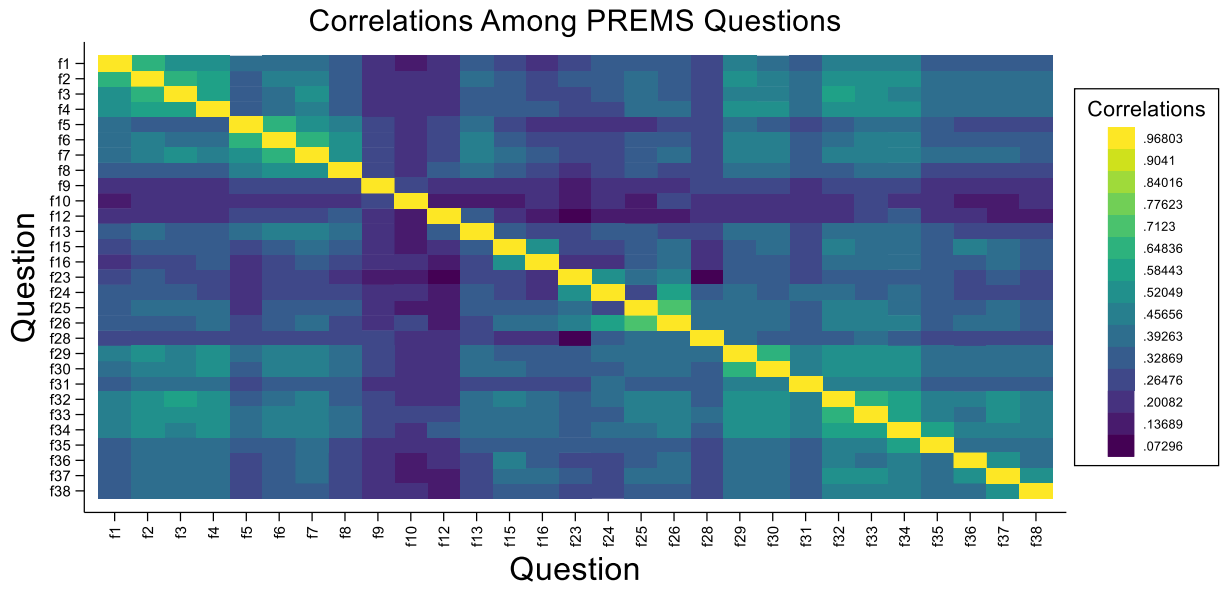


Figure 42 Heat plot of correlations between 4-point scale PREM questions in the acute care pilot study.

Global Satisfaction Questions

The proximity of question 20 (hospital rating) to 40 (global experience rating) in the PCA plot of Figure 43, along with the minimal contribution of question 39 (helpfulness of the hospital stay) when considered alongside question 40 in Figure 44, supports the decision to retain only question 21 (recommendation) and 40 (global experience).

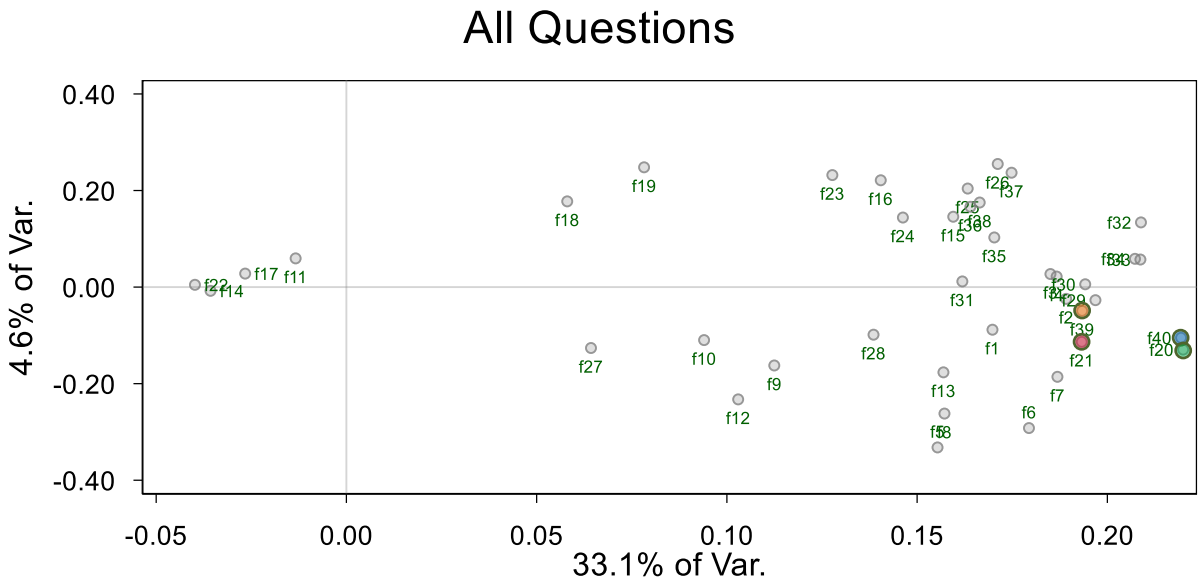


Figure 43 Position of global satisfaction questions in the principal component analysis (PCA) plane, in the acute care pilot study.

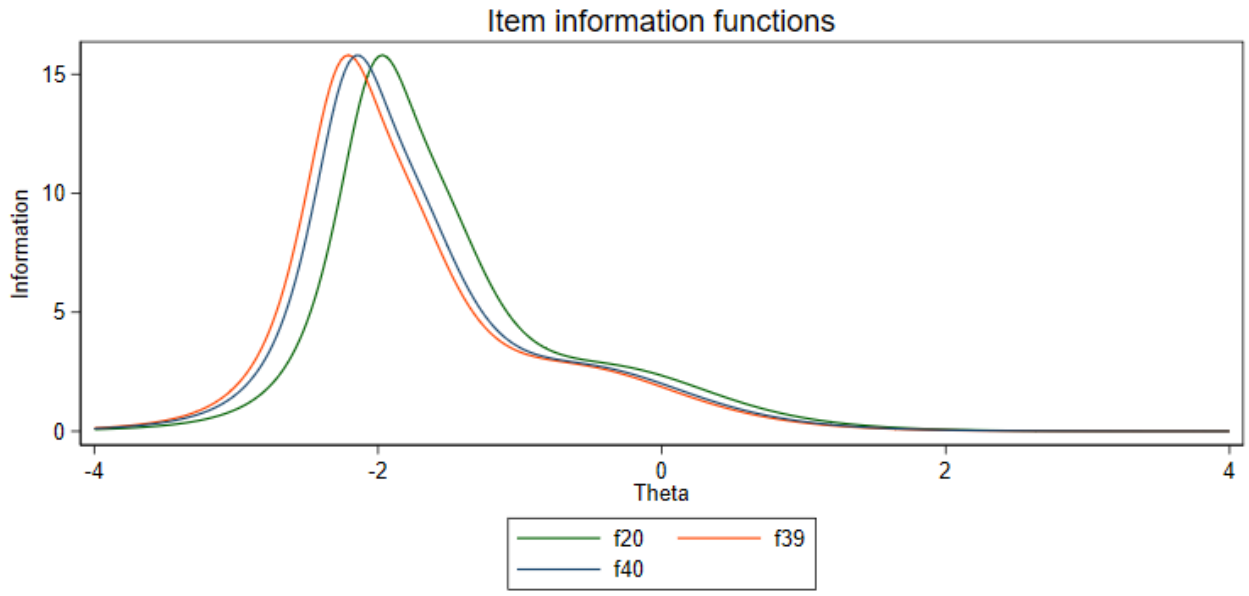


Figure 44 Contribution of each of the continuous scale items to the global satisfaction, based on item response theory (IRT) applied to the acute care pilot study. Question 21 on recommendation is not shown, as it uses a 4-point ordinal scale.

Summary Table

The original HCAHPS and CEPS-IC questionnaires underwent a similar, but more extensive, shortening process that reduced the number of questions by half. Table 19 presents the questions retained in these abbreviated, along with data summarizing the quality metrics of individual items, as detailed in Section 4.1.

Table 19 List of questions to be considered for shortening in the acute care questionnaire. Data quality is summarized by the rate of true missing values (i.e., no alternate answer selected), and potential scale saturation, as indicate by the ceiling effect, which may not necessarily be problematic in the context of extreme values boxes. The final two columns indicate which questions were retained in the shorter version of the original CPES-IC and the American HCAHPS, on which it was based.

Question	Type	Item-based criteria			
		True Miss.	Ceiling eff.	HCAPS	CPES-IC
1. Courtesy and respect (doc.)	PREMs	0.5%	88.1%	**6**	**4**
2. Careful listening (doc.)	PREMs	0.7%	78.7%	**7**	**5**
3. Understandable explanations (doc.)	PREMs	0.7%	69.9%	**8**	
4. Availability (doc.)	PREMs	1.2%	64.3%		
5. Courtesy and respect (nurs.)	PREMs	0.9%	86.2%	**1**	**1**
6. Careful listening (nurs.)	PREMs	1.0%	78.1%	**2**	**2**
7. Understandable explanations (nurs.)	PREMs	1.1%	71.7%	**3**	**3**
8. Availability (nurs.)	PREMs	1.9%	75.1%		
9. Cleanliness	PREMs	1.1%	80.1%	**10**	
10. Calm	PREMs	1.5%	57.5%	**11**	
11. Toilets?	Filter	4.2%		**12**	
12. Assistance for toileting	PREMs	5.1%	74.0%		
13. Pain control	PREMs	3.4%	82.3%		
14. New medicine?	Filter	4.3%		**18**	
15. Purpose of medication	PREMs	5.3%	63.1%		
16. Side effects	PREMs	5.7%	34.6%	**19**	
17. Destination?	Filter	2.0%		**20**	
18. Need for help	PREMs	5.0%	62.0%	**21**	
19. Symptom surveillance	PREMs	6.1%	60.9%	**22**	
20. Rating of hospital stay	Satisfaction	2.3%	35.5%	**5**	**14**
21. Recommendation	Satisfaction	2.3%	70.6%	**24**	
22. Emergency?	Filter	3.0%			
23. Information on admission	PREMs	5.4%	73.5%		
24. Admission organization	PREMs	5.1%	81.9%		
25. Condition information (emerg.)	PREMs	4.2%	54.8%		
26. Unfolding information (emerg.)	PREMs	4.9%	54.9%		
27. Hospital bed (emerg.)	PREMs	5.3%			
28. Transfer (emerg.)	PREMs	4.3%	76.4%		
29. Communication between staff	PREMs	3.0%	60.9%		

30. Information on treatment	PREMs	3.1%	58.9%		**8**
31. Punctuality	PREMs	3.6%	65.8%		
32. Information on health condition	PREMs	2.7%	70.9%		**13**
33. Worries	PREMs	4.4%	63.4%		
34. Decisions (pat.)	PREMs	3.5%	65.9%		**9**
35. Decisions (fam.)	PREMs	4.9%	67.6%		**10**
36. Understanding medication	PREMs	2.9%	76.7%		**11**
37. Continuining treatment	PREMs	3.8%	63.3%		**12**
38. Understanding condition	PREMs	5.2%	54.2%		
39. Helpfulness	Satisfaction	2.9%	54.5%		
40. Rating of experience	Satisfaction	2.9%	50.7%	**23**	

5.1.3.PREMs Score Construction

Initial Qualitative Attribution

Table 20 Categorization of CPES-IC questions into to ANQ's themes

Nr.	Question	Theme								
		Admission	Information	Involvement	Organization	Discharge	Confidence	Medication	Dignity	Other/notes (specify)
	YOUR CARE FROM DOCTORS									
1	During this hospital stay, how often did doctors treat you with courtesy and respect?						B		A/C/D	
2	During this hospital stay, how often did doctors listen carefully to you?		A/C				B		D	
3	During this hospital stay, how often did you receive understandable explanations from the doctors?		A/B/C/D							
4	During this hospital stay, how often were doctors available to answer your questions?		D		A/B/C					
	YOUR CARE FROM THE NURSING STAFF									
5	During this hospital stay, how often did the nursing staff treat you with courtesy and respect?						B		A/C/D	
6	During this hospital stay, how often did the nursing staff listen carefully to you?		A/C				B		D	
7	During this hospital stay, how often did you receive understandable explanations from the nursing staff?		A/B/C/D							

8	During this hospital stay, how often did you get help within the expected time after you asked for it (e.g., by pressing the call button)?				A/B/C/D					
	THE HOSPITAL ENVIRONMENT									
9	During this hospital stay, how often did the cleanliness of your room and bathroom meet your expectations?						A/B/C/D			
10	During this hospital stay, how often was the area around your room quiet at night?						A/B/C/D			
	YOUR EXPERIENCE IN THIS HOSPITAL									
12	How often did you get help within the expected time getting to the bathroom or in using a bedpan?				A/B/C/D					
13	During this hospital stay, how often did the hospital staff do everything they could to keep your pain under control?							A/B/C/d		
15	Before giving you any new medicine, how often did the hospital staff explain to you what the medicine was for?							A/B/C/D		
16	Before giving you any new medicine, how often did you receive understandable explanations from the hospital staff on possible side effects?							A/B/C/D		
	AFTER YOU LEFT THE HOSPITAL									
18	During this hospital stay, did a member of the hospital staff talk with you about whether you					A/B/C/D				

	would have the help you needed after you left the hospital?									
19	During this hospital stay, did you get any written information about what symptoms or health problems to look out for after you left the hospital?					A/B/C/D				
	YOUR OVERALL RATING OF THE HOSPITAL									
20	Your overall rating of the hospital listed in the cover letter. On a scale from 0 to 10, 0 means «Worst hospital», 10 means «Best hospital». (Check the box under the appropriate number.)									
21	Would you recommend this hospital to your friends and family?									
	YOUR ARRIVAL AT THE HOSPITAL									
23	Before coming to the hospital, did you have enough information about what was going to happen during the admission process?	A/B/C/D								
24	Was your admission into the hospital organized?	A/B/C/D								
25	When you were in the emergency department, did you get enough information about your condition and treatment?	A/B	C/D							
26	Were you given enough information about what was going to happen during your admission to the hospital?	A/C/D	B							
27	After you knew that you needed to be admitted to a hospital bed, did you	B/D			A/C					

	have to wait too long before getting there?									
28	Was your transfer from the emergency department into a hospital bed organized?				A/B/C/D					
	DURING YOUR HOSPITAL STAY									
29	Do you feel that there was good communication about your care between doctors, nursing staff and other hospital staff?				A/B/C		D			
30	How often did doctors, nursing staff and other hospital staff seem informed and up to date about your care?				A/B/C		D			
31	How often were examinations and procedures done when you were told they would be done?				A/B/C/D					
32	During this hospital stay, did you get all the information you needed about your health condition and treatments?		A/B/C/D							
33	Did you get the support you needed to help you with any anxieties, fears or worries you may have had during this hospital stay?						A/B		C/D	
34	Were you involved as much as you wanted to be in decisions about your care and treatments?			A/B/C/D						
35	Were your family or contact person involved as much as you wanted in decisions about your care and treatments?			A/B/C/D						

	YOUR HOSPITAL DISCHARGE									
36	Before you left the hospital, did you have a clear understanding about all of your prescribed medications? This also includes medication that you were already taking before your hospital stay.		B			C/D		A		
37	Did you receive enough information from hospital staff about what to do if you were worried about your health condition or treatments after you left the hospital?		B			A/C/D				
38	After you left the hospital, did you have a better understanding of your health condition than when you entered?		B			A/C/D				
	YOUR OVERALL RATINGS									
39	Your overall rating of the hospital listed in the cover letter. On a scale from 0 to 10, 0 means «Did not help», 10 means «Helped a lot». (Check the box under the appropriate number.)									
40	Your overall rating of the hospital listed in the cover letter. On a scale from 0 to 10, 0 means «I had a very poor experience», 10 means «I had a very good experience». (Check the box under the appropriate number.)									
35	Were your family or contact person involved as much as you wanted in decisions about your care and treatments?			A/B/C/D						

	YOUR HOSPITAL DISCHARGE									
36	Before you left the hospital, did you have a clear understanding about all of your prescribed medications? This also includes medication that you were already taking before your hospital stay.		B			C/D		A		
37	Did you receive enough information from hospital staff about what to do if you were worried about your health condition or treatments after you left the hospital?		B			A/C/D				
38	After you left the hospital, did you have a better understanding of your health condition than when you entered?		B			A/C/D				
	YOUR OVERALL RATINGS									
39	Your overall rating of the hospital listed in the cover letter. On a scale from 0 to 10, 0 means «Did not help», 10 means «Helped a lot». (Check the box under the appropriate number.)									
40	Your overall rating of the hospital listed in the cover letter. On a scale from 0 to 10, 0 means «I had a very poor experience», 10 means «I had a very good experience». (Check the box under the appropriate number.)									

Inter-judges' agreement

The inter-judges' agreement was assessed using Fleiss's Kappa, a statistical measure that evaluates the level of agreement among multiples judges classifying items into more than two mutually exclusive categories (Fleiss 1971). The overall kappa value was k = 0.67 ($p < 0.001$ IC95%: [0.62; 0.73]), which is conventionally interpreted as substantial agreement (Landis and Koch 1977, McHugh 2012). Agreement levels varied by theme, ranging from moderate agreement (k = 0.41) for *Information* to perfect agreement (k = 1.00) for *Involvement*. The kappa values for each theme, along with their signification levels and 95% confidence intervals, are presented in Table 21.

Table 21 Kappa agreement levels for the qualitative classification of PREMs items into thematic categories.

Theme	Kappa	Standard Error	z	Sig.	Lower Bound	Upper Bound
Admission	0.698	0.072	9.676	<.001	0.557	0.84
Information	0.407	0.072	5.645	<.001	0.266	0.549
Involvement	1	0.072	13.856	<.001	0.859	1.141
Organization	0.797	0.072	11.038	<.001	0.655	0.938
Discharge	0.762	0.072	10.557	<.001	0.62	0.903
Confidence	0.476	0.072	6.598	<.001	0.335	0.618
Medication	0.914	0.072	12.67	<.001	0.773	1.056
Dignity	0.421	0.072	5.84	<.001	0.28	0.563

Score Structure

The bottom- and top-box distribution of items within each theme is illustrated in Figure 45. While items inside *Information* theme, which contains the largest number of items, exhibits strong internal homogeneity. In contrast, the *Admission* theme, positioned on left side of the figure, reveals differing perceptions between elective and emergency patients. The *Not Used* block on the right displays boxes representing items that were not assigned to any theme.

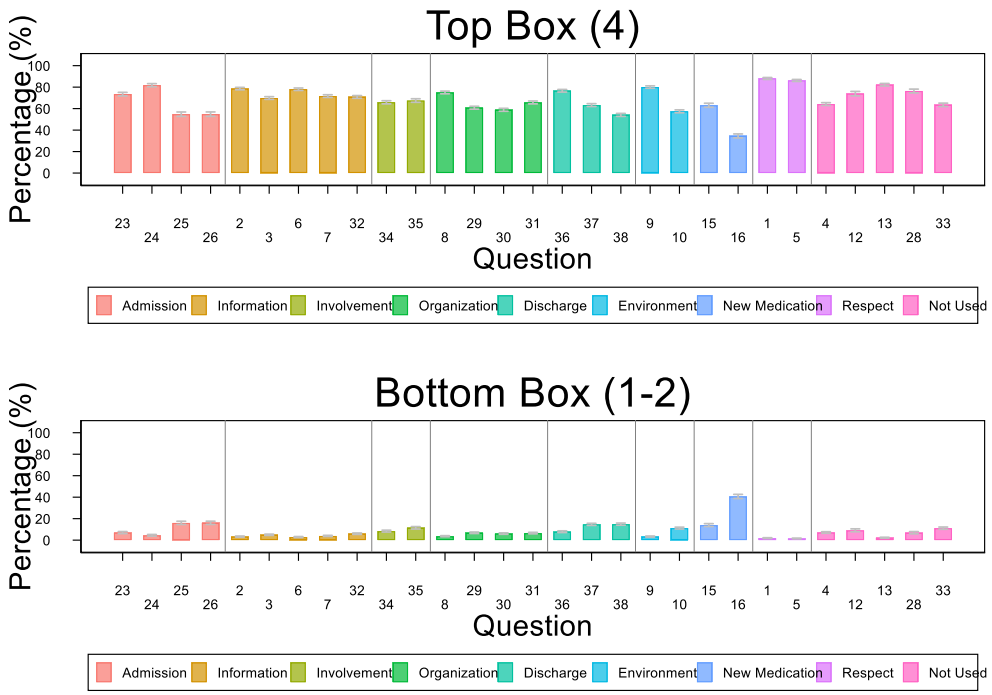


Figure 45 Distribution of responses at the extreme ends (top and bottom boxes) for each item within thematic domains in the acute care pilot study.

Quantitative Measures

A summary of theme-level metrics used to guide the item attribution process is presented in Table 22. It is important to note that the final decisions were not solely based on these criteria. Qualitative considerations among other factors, also influenced the selection process.

Table 22 Summary table of theme-level criteria for selecting questions in the acute care sample. Questions are grouped by thematic category and function. The table includes measures of a theme's global reliability and how it may change if the corresponding item was removed, distance from other questions detected on the principal component analysis (PCA) plane, redundance with other questions detected with Item response theory (IRT), correlation with other items.

Summary table	Theme-based criteria			
Theme Question	α Cronbach	PCA	IRT - IIF	Corr. ≥ 0.6
Admission	0.86			
23. Admission information	0.94			
24. Admission organization	0.92			
25. Condition information (emerg.)	0.65		26	26
26. Unfolding information (emerg.)	0.67		25	25
Information	0.83			
2. Careful listening (doc.)	0.79		3, 32	1, 3
3. Understandable explanations (doc.)	0.79		2, 32	2, 4
6. Careful listening (nurs.)	0.81			5
7. Understandable explanations (nurs.)	0.79			
32. Condition information	0.80	Far	2, 3	33, 34
Involvement	0.75			
34. Decisions (pat.)	0.75			32, 33
35. Decisions (fam.)	0.75			
Organization	0.77			
8. Rapid help (nurs.)	0.77		28	
29. Communication	0.65	Far	30	30
30. Treatment information	0.66	Far	29	29
31. Punctuality	0.74			
Discharge	0.73			
36. Understanding medication	0.70			
37. Continuining treatment	0.56			
38. Understanding condition	0.65			
Environment (Confidence)	0.41			
9. Cleanliness	0.41			
10. Calm	0.41			
New Medication (Medication)	0.68			
15. Purpose of medication	0.68			

16. Side effects	0.68			
Respect (Dignity)	0.58			
1. Courtesy and respect (doc.)	0.58		5	2
5. Courtesy and respect (nurs.)	0.58	Far	1	6
Other	0.73			
4. Availability (doc.)	0.67			3
12. Help for toilets	0.71			
13. Pain control	0.69			
28. Transfer (emerg.)	0.70		8	
33. Worries	0.61			32, 34
(Binary)				
18. Need for help		19		
19. Symptom surveillance		18		
Global Satisfaction				
20. Stay		21, 39, 40		
21. Recommendation		20, 39, 40		
39. Help		20, 21, 40		
40. Experience		20, 21, 39		
Skip				
11. Toilets				
14. New medicine				
17. Destination				
22. Emergency				
27. Hospital bed (emerg.)				

Similarity Among Questions

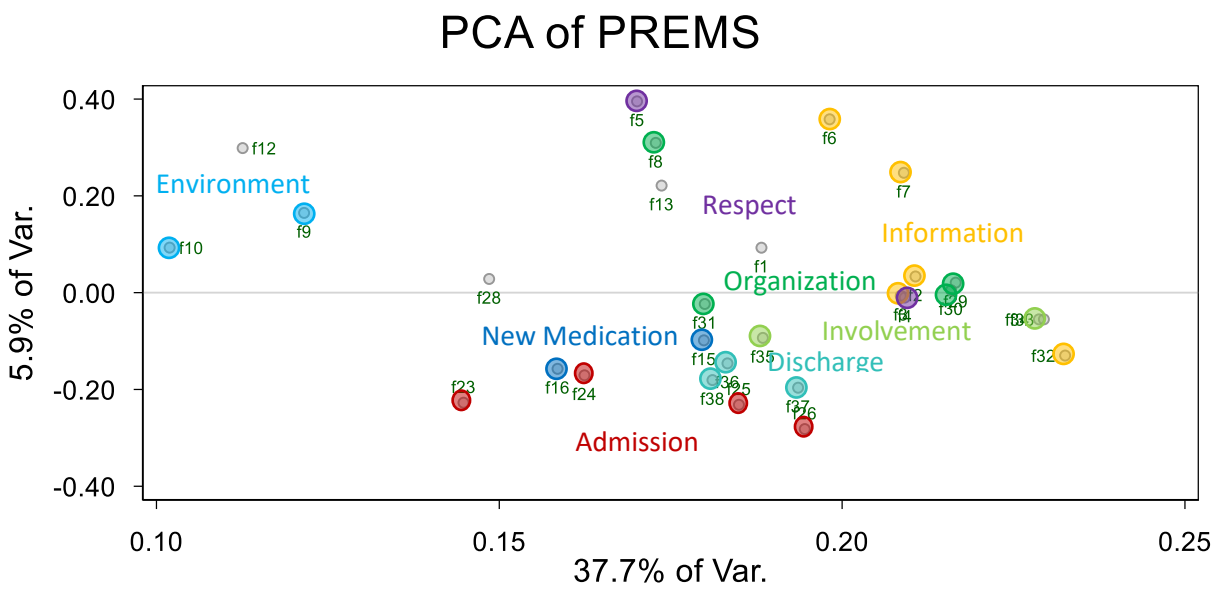
Similarities and differences among PREMs questions can be represented graphically on a principal component analysis (PCA) plane. This exploratory data analysis based on pairwise correlations of all PREMs questions defined on a similar four-items scale shows how themes are distributed across the PCA plane, as shown in Figure 46.

The axes of the PCA plot account for over 40% of the total variance. The horizontal axis can be interpreted as a gradient ranging from environmental aspects on the left to human-centered aspects on the right. Meanwhile, the vertical axis spans from routine experiences during the hospital stay at the top to transitional changes occurring at admission and discharge at the bottom.

Ideally, questions grouped by theme should appears as distinct, compact clusters on the PCA plot. Although this rarely occurs in practice, Figure 46 highlights themes that come close to meeting these criteria: *Environment* is clearly positioned on the left, *Discharge* forms a tight cluster in the center, and *Information* is distributed along the right edge. In contrast, the more populated *Organization* theme extends into the domains of other themes and *Respect* is dispersed across the upper right quadrant.

Questions shown in gray in Figure 46 were not assigned to any theme. Their isolation on the PCA plot (e.g., Question 12: *Help for bathroom*) can support qualitative interpretations, while excessive proximity (e.g., Question 33: *Worries*) may indicate potential redundancy in terms of correlation.

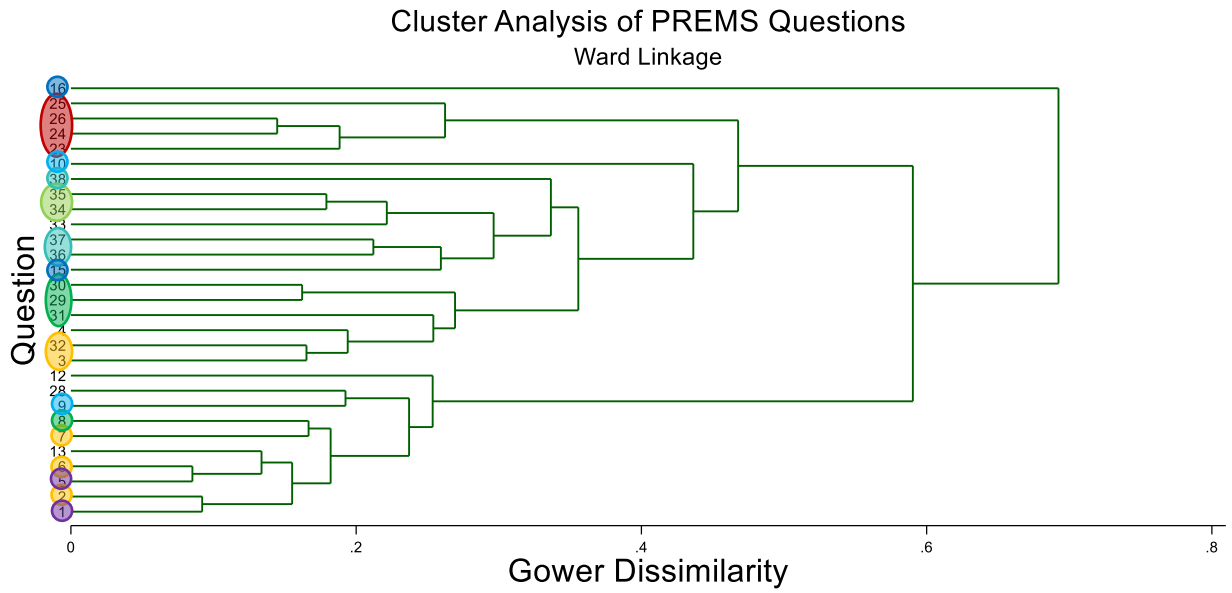
Figure 46 Principal component analysis (PCA) of correlations among 4-item PREMs questions in the acute care pilot sample. Colors represent thematic attributions.



As a confirmatory analysis, an automatic clustering procedure, incorporating the remaining principal components not shown in Figure 46, produced the groupings presented in Figure 47. The *Admission* theme is clearly clustered at the top, as is *Organization*, with the exception of Question 8: *Rapid Help*, which also stands apart on the PCA plot. The *Information* theme appears to consist of two subclusters, with Question 3: *Understandable explanations from doctors* and Question 32: *General information on health condition* isolated from the rest. Notably, these two questions also exhibit similar discriminative power according to item response analysis (Section on All eight themes exhibit the expected properties: low inter-item correlations, high item-score correlations, and consistency across the three linguistic regions. Alpha values fall within the optimal range of 0.70–0.80, except for the two-item themes Respect and *Environment*, which show lower values of 0.58 and 0.41, respectively.

Discriminative Power below).

Figure 47 Cluster analysis of correlations among 4-item PREMs questions in the acute care pilot sample. Colors represent thematic attributions, as in Figure 46.



Predictive Power of Questions

In the context of this pilot study, the four global satisfaction questions (20, 21, 39 & 40) serve as the gold standard for assessing overall patient satisfaction, as measured in previous PatZu questionnaires. Evaluating the correlation between a specific PREMs indicator and such broader satisfaction measure helps determine how each PREMs item aligns with the more diffuse concept of patient satisfaction.

The vertical bands in Figure 48 indicate that PREMs items tend to correlate fairly uniformly with all satisfaction questions. The darker horizontal band in the middle reveals that question 21 (net promoter score) and question 39 (overall helpfulness) show weaker correlations with PREMs items compared to question 20 (overall hospital rating) and question 40 (overall patient experience). Each pair of questions exhibits a similar correlation pattern, which ultimately supported the decision to select one representative from each pair and shorten the questionnaire.

On the other hand, themes such as *Environment*, represented by question 9 (cleanliness) and question 10 (calm), and *New Medication*, captured by question 15 (purpose) and question 16 (side effects), along with question 12 (help for bathroom), show weaker correlations with overall patient satisfaction. In contrast, question 33 (worries and anxieties), which is not assigned to any specific theme, question 34 (patient involvement), and question 29 (communication, from the *Organization* theme) exhibit the strongest correlations across all experience questions. Notably, the second question from the *Involvement* theme, which focuses on relatives, shows weaker correlations.

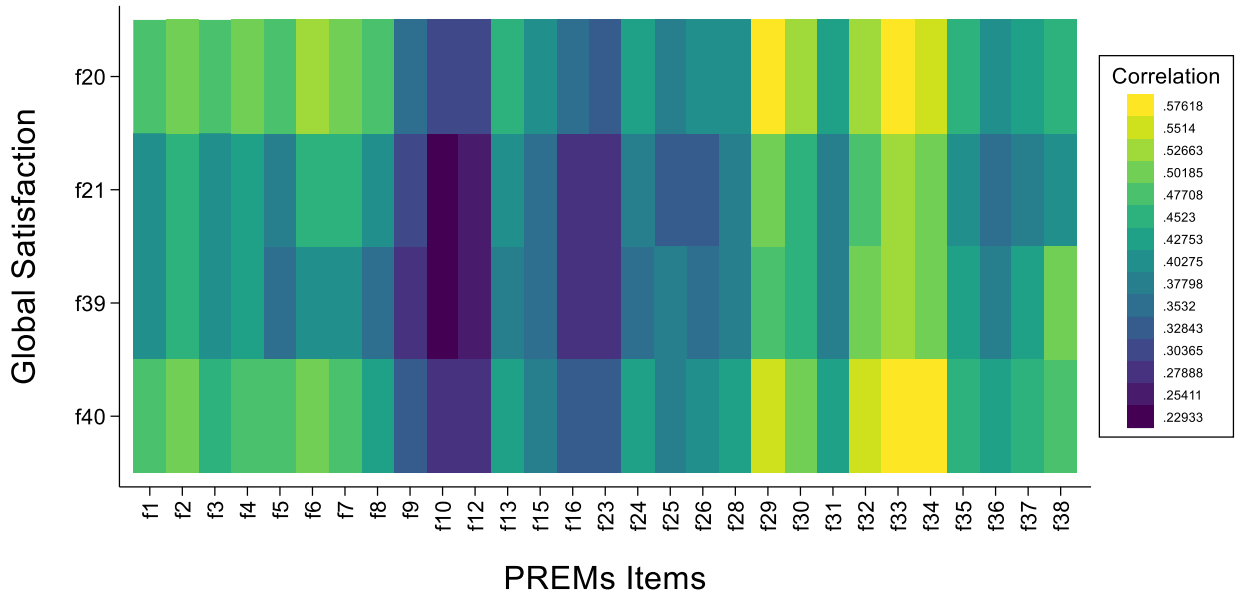


Figure 48 Correlation matrix of PREMs items with global satisfaction questions.

Availability of Themes

Missing values and alternative answers impact the availability of data for constructing thematic scores. Table 23 to Table 29 present the detailed distribution of missing responses per respondent within each theme.

Admission

The *Admission* theme represents a special case, as different questions are addressed to two distinct subsamples: elective and emergency patients. Consequently, the maximum expected response is two answers and two missing values (over 90% of respondents provided all relevant information). Those providing more responses likely bypassed the filter.

Table 23 Number of missing values or alternative responses across items within the *Admission* theme.

Missing Items	Respondents	%
all	212	4.0%
3	120	2.3%
2	4 815	91.3%
1	47	0.9%
none	81	1.5%

Information

Despite the large number of questions defining the Information theme, nearly all respondents (96%) provided a complete set of answers, with only less than 3% omitting one out of five.

Table 24 Number of missing values or alternative responses across items within the *Information* theme.

Missing Items	Respondents	%
all	9	0.2%
4	5	0.1%
3	23	0.4%
2	24	0.5%
1	151	2.9%
none	5 063	96.0%

Involvement

Only half of the sample responded to both questions on *Involvement*, largely because many participants preferred not to involve relatives. As a result, in one-third of the cases, the score is based solely on the patient's personal involvement. Fewer than 15% of respondents provided no information at all.

Table 25 Number of missing values or alternative responses across items within the *Involvement* theme.

Missing Items	Respondents	%
all	770	14.6%
1	1 763	33.4%
none	2 742	52.0%

Organization

The *Organization* theme also includes a large set of questions, yet a slightly higher proportion of respondents (less than 15%) tend to omit one out of four. Although missing values are relatively few and evenly distributed (see Figure 7), alternative questions to items 8 (rapid assistance) and 31 (punctuality) may have contributed to this pattern.

Table 26 Number of missing values or alternative responses across items within the *Organization* theme.

Missing Items	Respondents	%
all	34	0.6%
3	76	1.4%
2	105	2.0%
1	742	14.1%
none	4 318	81.9%

Discharge

More than 80% of respondents answered all three questions, while slightly over 10% omitted one. Missing values appear in 2% to 5% of questionnaires, and the 10% of respondents who chose the alternative answer to question 36 on medication may have contributed to this figure.

Table 27 Number of missing values or alternative responses across items within the *Discharge* theme.

Missing Items	Respondents	%
all	122	2.3%
2	62	1.2%
1	611	11.6%
none	4 480	84.9%

Environment

Nearly all respondents answered question 9 (cleanliness) and 10 (calm) related to the environment.

Table 28 Number of missing values or alternative responses across items within the *Environment* theme.

Missing Items	Respondents	%
all	40	0.8%
1	56	1.1%
none	5 179	98.2%

New Medication

As the name suggests, the *New Medication* theme applies only to patients who received a new medication during their hospital stay. Since this concerns only half of the respondents, the available data represents just that portion of the sample. However, within this subgroup, nearly all respondents answered both questions (see Table 28).

Table 29 Number of missing values or alternative responses across items within the *New Medication* theme.

Missing Items	Respondents	%
all	2 710	51.4%
1	40	0.8%
none	2 525	47.9%

Respect

The two questions on courtesy and respect on which the Respect theme is based were answered by a vast majority of respondents, as shown in Table 30.

Table 30 Number of missing values or alternative responses across items within the *Respect* theme.

Missing Items	Respondents	%
all	12	0.2%
1	50	0.9%
none	5 213	98.8%

Respondents generally answer all questions within a given theme. In themes containing a larger number of questions, such as *Information* and *Organization*, they may occasionally leave some responses blank, though this behavior remains minimal.

Three themes—*Information*, *Environment*, and *Respect*—achieve near-total completeness of available responses (close to 100%). *Discharge* and *Organization* show slightly lower response rates, exceeding 80%. The significantly lower rates observed for *Involvement* and *New Medications* are attributable to alternative response options and the use of filter questions (note that these figures reflect data availability and its influence on the sample sizes used. The relevance of the data or whether it should have been provided is not considered in this context).

Homogeneity Within Themes

How are top and bottom boxes distributed within themes? Since thematic scores are calculated by averaging multiple PREMs indicators, it is important that respondents demonstrate some consistency in their answers for the average score to accurately reflect their behavior.

Figure 49 shows that nearly half of the respondents consistently selected the highest response option for all items within each theme. Notably, the two-item Respect theme reached a consistency rate of 80%. On the other hand, with exception of the *New Medication* Theme, respondents do not consistently select the lower response options. While many may choose one or a few bottom-box answers, they rarely do so across all questions within a given theme.

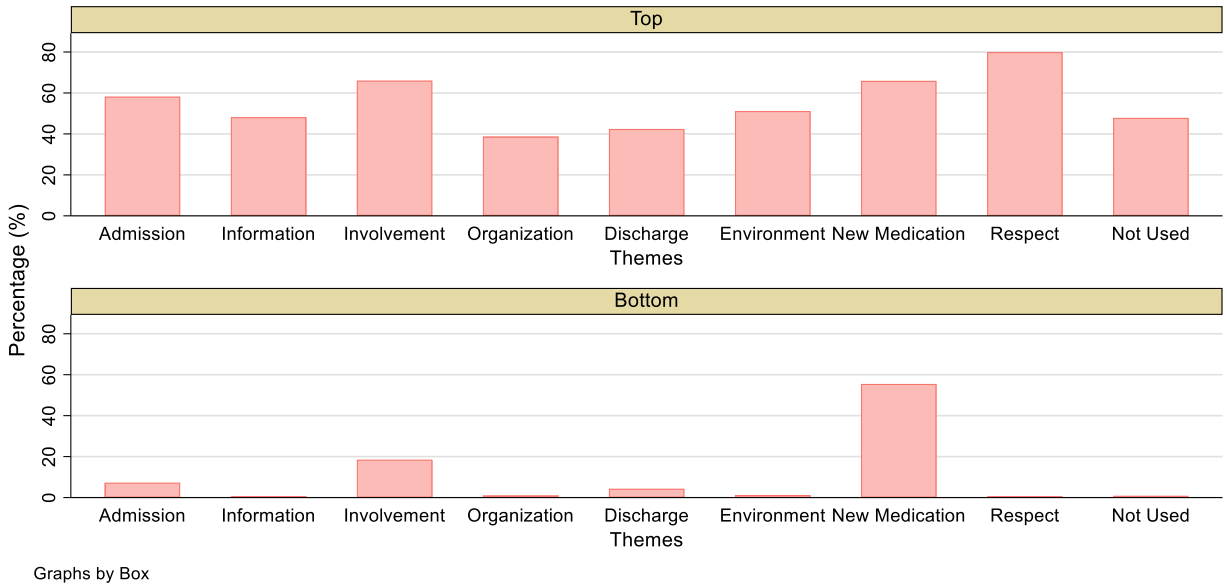


Figure 49 Proportion of respondents consistently flagged by the top-box or bottom-box responses across all answered items within a given theme.

Figure 50 provides a more detailed view, revealing that respondents who answered all questions also tended to give similar responses across all items.

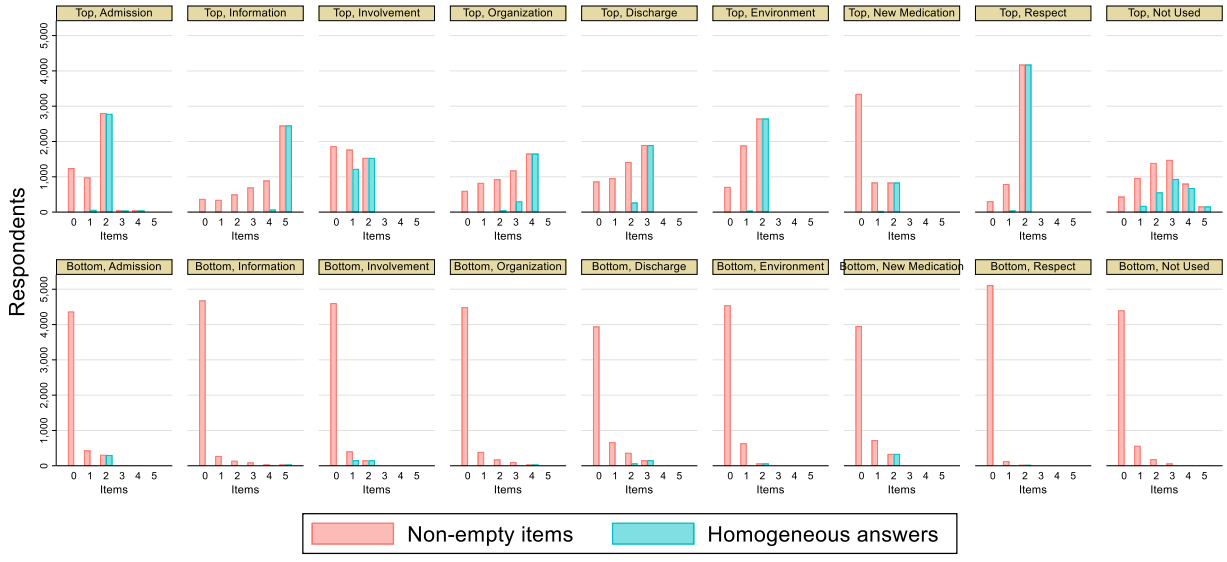


Figure 50 Number of items flagged as top-or bottom box within each theme.

Defining indicators as top and bottom boxes doubles the number of figures presented, compared to the PatZu reports, which only included averages scores. If a single summary statistic must be chosen, the more consistent response pattern favors the use of the top box.

Internal Consistency of Themes

In the following sections, we use Cronbach's alpha to assess the internal consistency of themes and visualize item-item (expected to be lower) and item-score (expected to be higher) correlations using heatmaps. A commonly cited rule of thumb suggests that an alpha value of 0.8 is considered optimal (Lance, Butts et al. 2006). The tables also demonstrate the effect of removing individual items on the overall theme score.

Admission

Items from the *Admission* theme fall into two categories: those targeting elective patients (questions 23 and 24) and those targeting emergency cases (questions 25 and 26). The emergency block has a greater impact on the overall alpha value (Table 31). It also shows higher inter-item correlations (see Figure 51). The theme shows a consistent correlation structure across all three linguistic regions.

Table 31 Internal consistency of the *Admission* theme. The overall Cronbach's alpha is presented on the first line, followed by values obtained after excluding individual items.

Question	Items	Variance	Cronbach's α	Difference
Admission	4	0.31	0.86	
23. Admission information	3	0.47	0.94	0.09
24. Admission organization	3	0.48	0.92	0.06
25. Condition information (emerg.)	3	0.19	0.65	-0.21
26. Unfolding information (emerg.)	3	0.18	0.67	-0.19

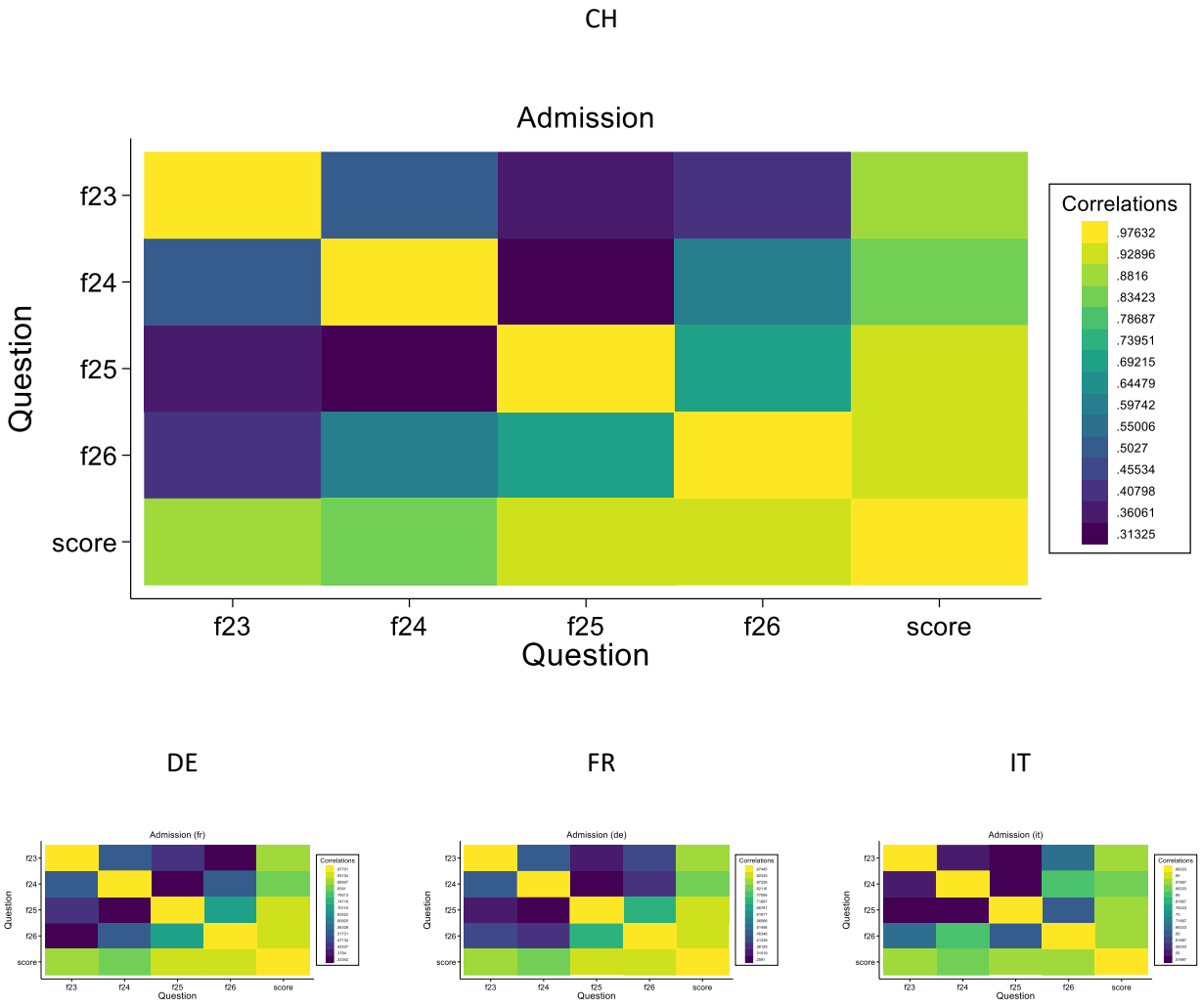


Figure 51 Correlations between items and the overall score within the *Admission* theme. The upper row displays results for the entire sample, while the lower row shows language-specific subsamples. Ideally, correlations should be higher (yellow) between each item and the overall score, and lower (blue) between individual items.

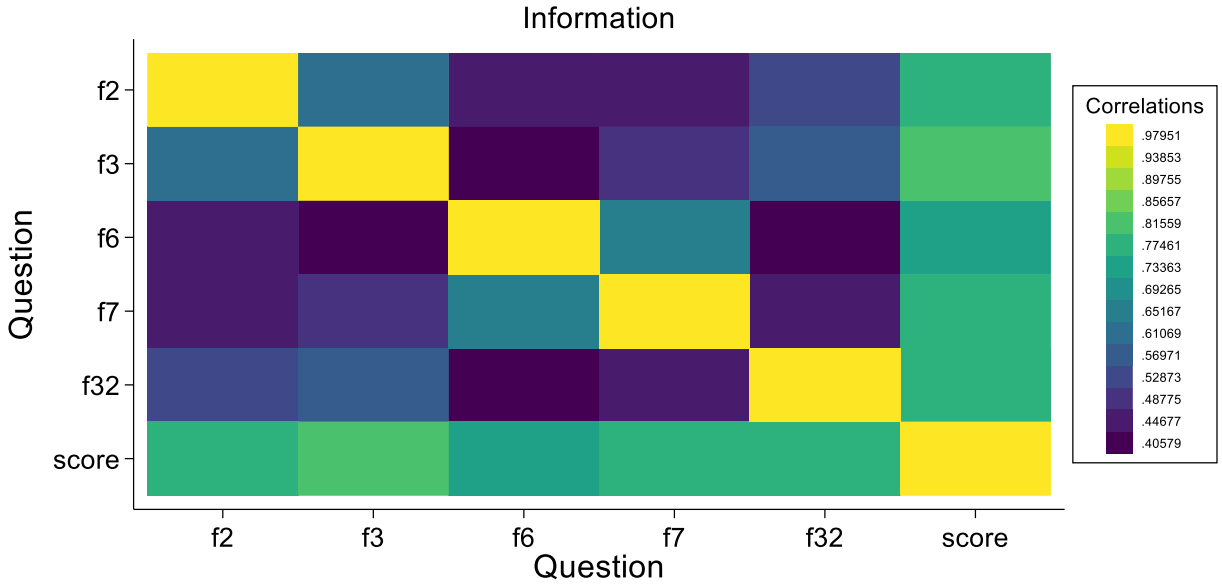
Information

Figure 52 shows a correlation structure that is stable across regions and reveals slightly stronger correlations among questions addressing the same professional groups. Table 32 highlights the minimal impact of removing individual items on the overall score. It is worth noting that this theme includes a larger number of items, which may reduce the interpretability of the alpha value as an indicator of reliability.

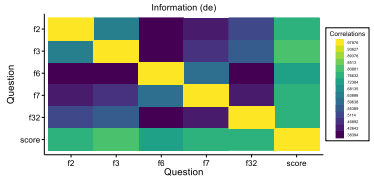
Table 32 Internal consistency of the *Information* theme. The overall Cronbach's alpha is presented on the first line, followed by values obtained after excluding individual items.

Question	Items	Variance	Cronbach's α	Difference
Information	5	0.16	0.83	
2. Careful listening (doc.)	4	0.16	0.79	-0.04
3. Understandable explanations (doc.)	4	0.15	0.79	-0.04
6. Careful listening (nurs.)	4	0.17	0.81	-0.02
7. Understandable explanations (nurs.)	4	0.16	0.79	-0.04
32. Condition information	4	0.15	0.80	-0.03

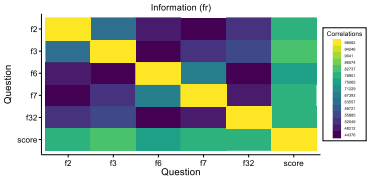
CH



DE



FR



IT

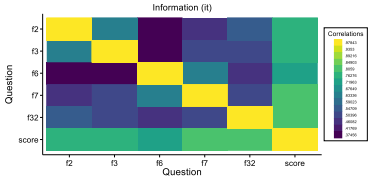


Figure 52 Correlations between items and the overall score within the *Information* theme. The upper row displays results for the entire sample, while the lower row shows language-specific subsamples. Ideally, correlations should be higher (yellow) between each item and the overall score, and lower (blue) between individual items.

Involvement

Since the Involvement theme consists of only two items, it is not possible to assess the impact of removing one. The overall alpha value falls within the expected range (Table 33), and the correlation structure is as anticipated (Figure 53), showing consistency across all three regions.

Table 33 Internal consistency of the *involvement* theme. The overall Cronbach's alpha is reported on the first line. Since the theme consists of only two items, discarding either would prevent the construction of an aggregated score.

Question	Items	Variance	Cronbach's α	Difference
Involvement	2	0.36	0.75	
34. Decisions (pat.)	1			
35. Decisions (fam.)	1			

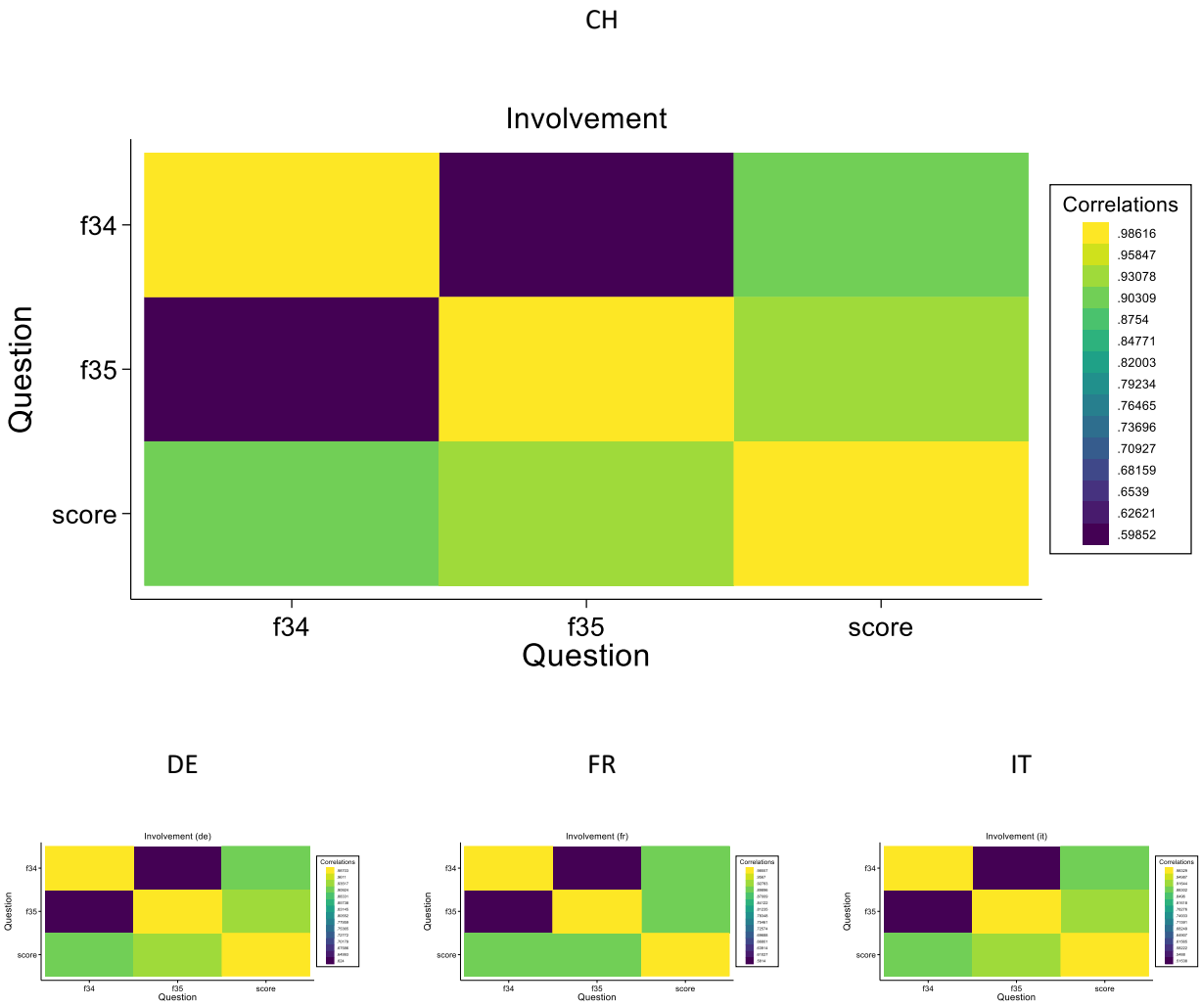


Figure 53 Correlations between items and the overall score within the *Involvement* theme. The upper row displays results for the entire sample, while the lower row shows language-specific subsamples. Ideally, correlations should be higher (yellow) between each item and the overall score, and lower (blue) between individual items.

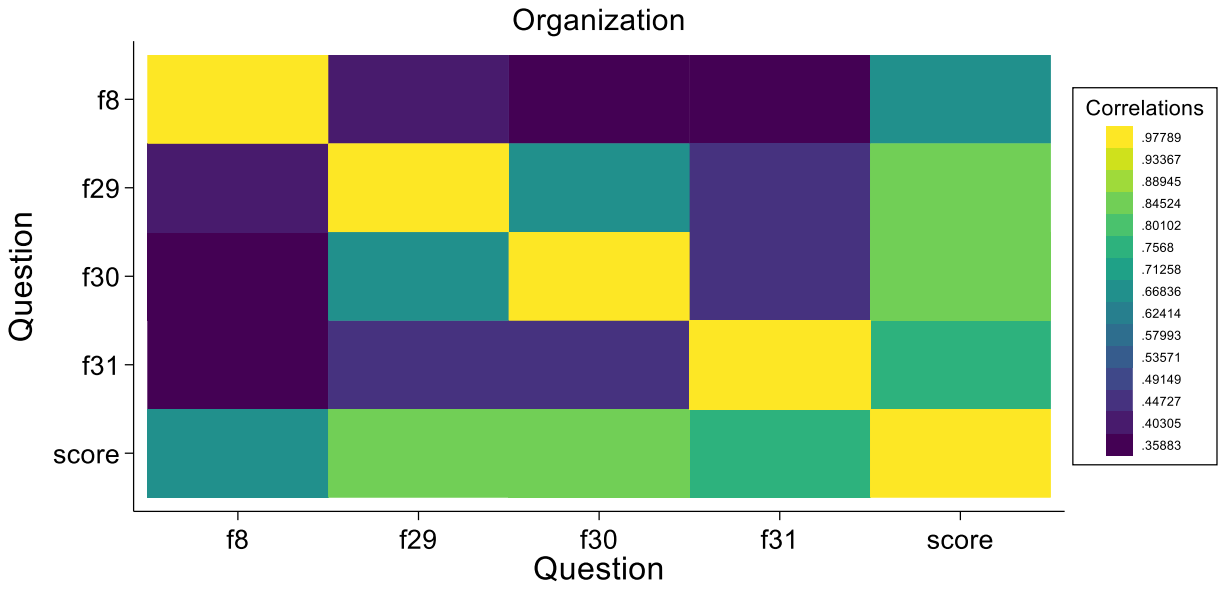
Organization

Successive questions from the same section ("During your hospital stay") tend to show slightly stronger correlations than question 8, which is addressed to the nursing staff (Figure 54). This results in a modest decrease in the overall alpha score when those questions are removed, as illustrated in Table 34.

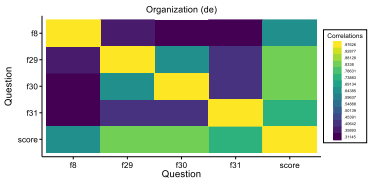
Table 34 Internal consistency of the *organization* theme. The overall Cronbach's alpha is presented on the first line, followed by values obtained after excluding individual items.

Question	Items	Variance	Cronbach's α	Difference
Organization	4	0.17	0.77	
8. Rapid help (nurs.)	3	0.22	0.77	0.00
29. Communication	3	0.14	0.65	-0.11
30. Treatment information	3	0.15	0.66	-0.11
31. Punctuality	3	0.18	0.74	-0.03

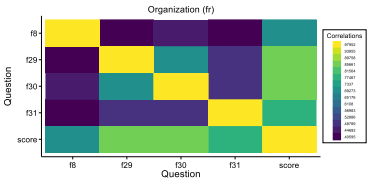
CH



DE



FR



IT

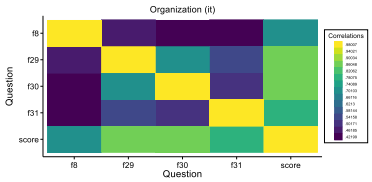


Figure 54 Correlations between items and the overall score within the *Organization* theme. The upper row displays results for the entire sample, while the lower row shows language-specific subsamples. Ideally, correlations should be higher (yellow) between each item and the overall score, and lower (blue) between individual items.

Discharge

The three questions defining the *Discharge* theme appear consecutively in the questionnaire section bearing the same title. Their inter-item correlation is modest (Figure 55), and their impact on the overall alpha value is moderate (Table 35).

Table 35 Internal consistency of the *discharge* theme. The overall Cronbach's alpha is presented on the first line, followed by values obtained after excluding individual items.

Question	Items	Variance	Cronbach's α	Difference
Discharge	3	0.32	0.73	
36. Understanding medication	2	0.40	0.70	-0.04
37. Continuining treatment	2	0.23	0.56	-0.17
38. Understanding condition	2	0.31	0.65	-0.08

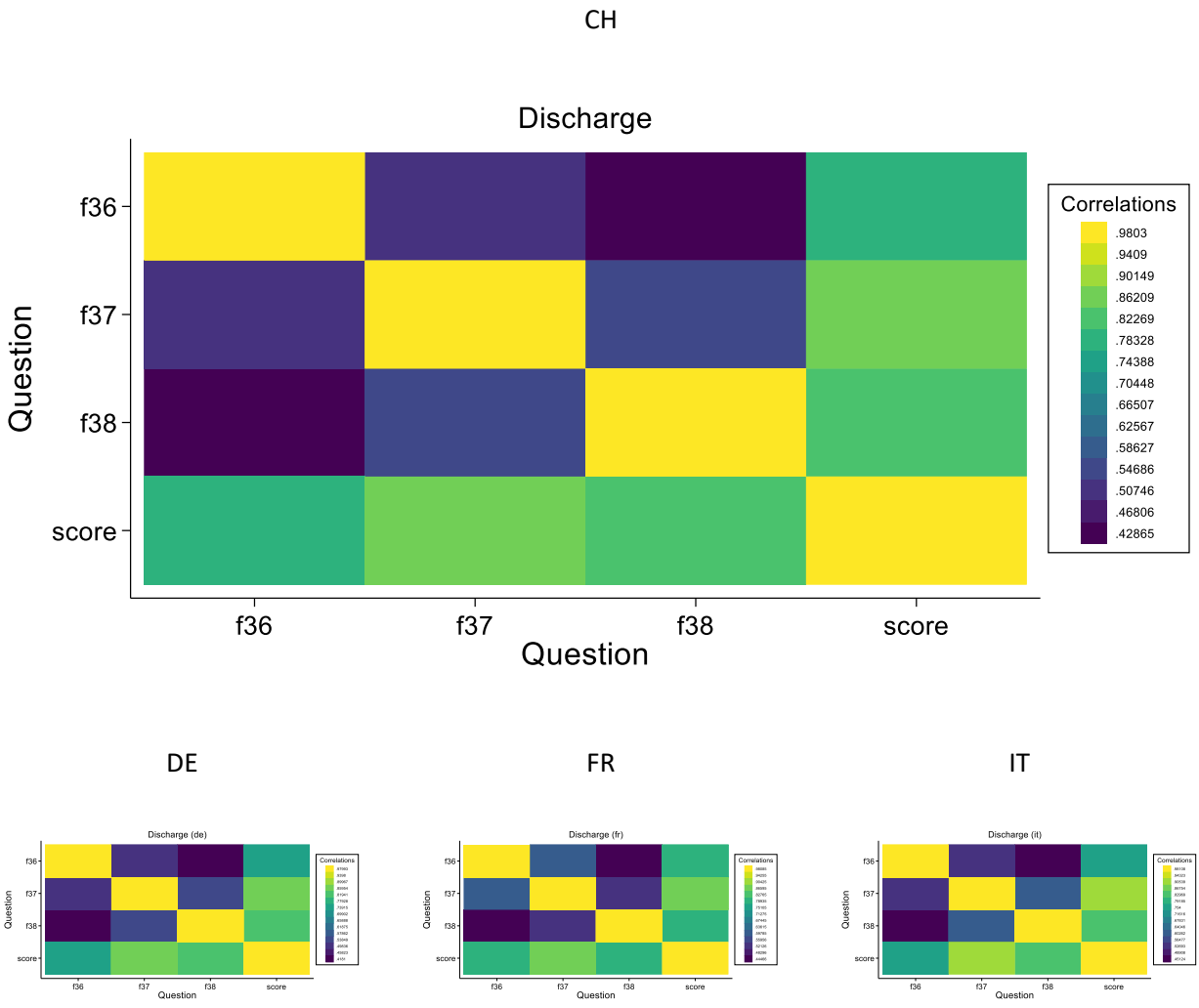


Figure 55 Correlations between items and the overall score within the *Discharge* theme. The upper row displays results for the entire sample, while the lower row shows language-specific subsamples. Ideally, correlations should be higher (yellow) between each item and the overall score, and lower (blue) between individual items.

Environment

The Environment theme is defined in the corresponding section of the questionnaire. As seen from Figure 56, inter-item correlations are low, while correlations with the overall score are high and consistent across all regions. However, the alpha score remains low (Table 36).

Table 36 Internal consistency of the *Environment* theme. The overall Cronbach's alpha is reported on the first line. Since the theme consists of only two items, discarding either would prevent the construction of an aggregated score.

Question	Questions	Variance	Cronbach's α	Difference
Environment	2	0.11	0.41	
9. Cleanliness	1			
10. Calm	1			

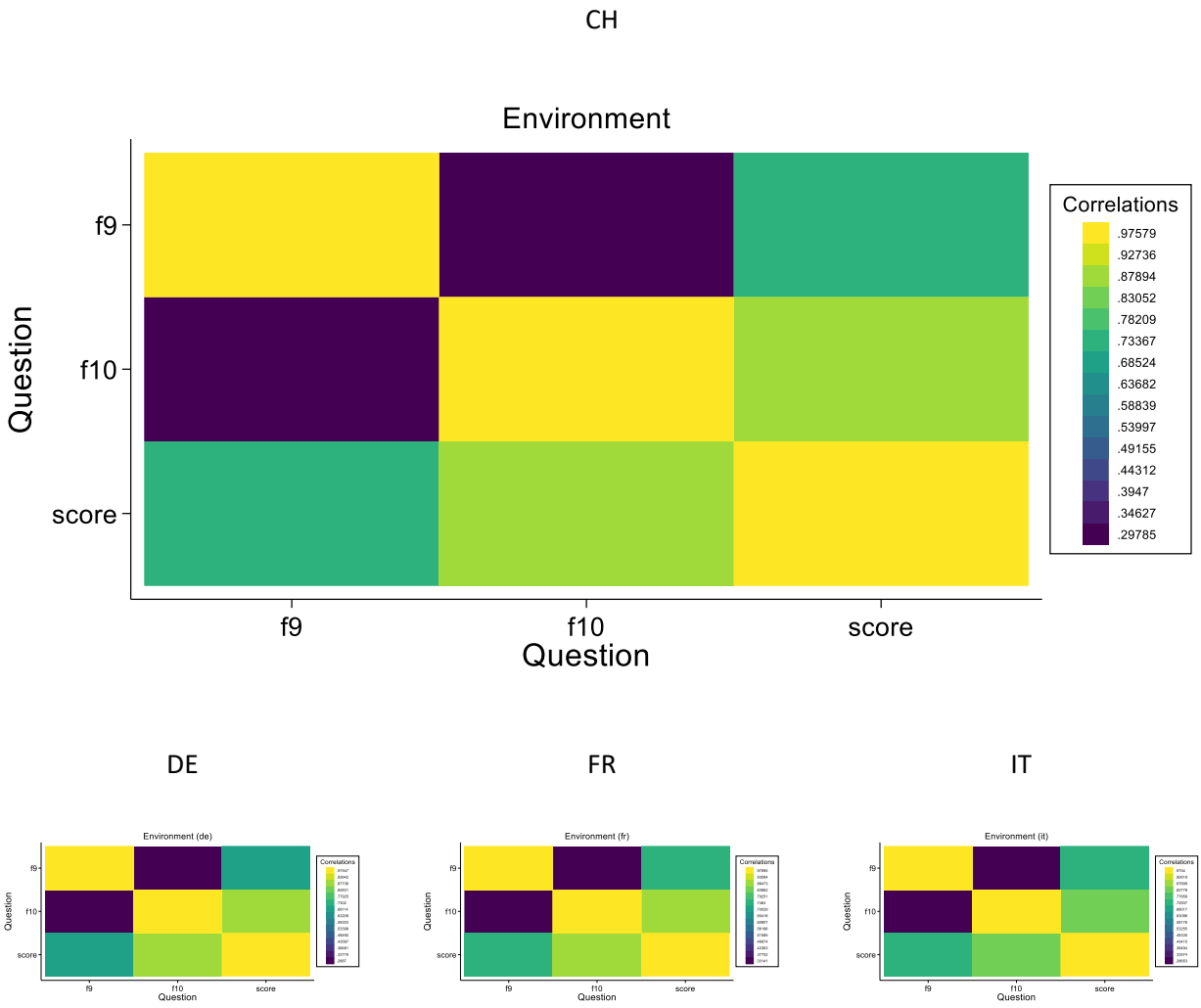


Figure 56 Correlations between items and the overall score within the *Environment* theme. The upper row displays results for the entire sample, while the lower row shows language-specific subsamples. Ideally, correlations should be higher (yellow) between each item and the overall score, and lower (blue) between individual items.

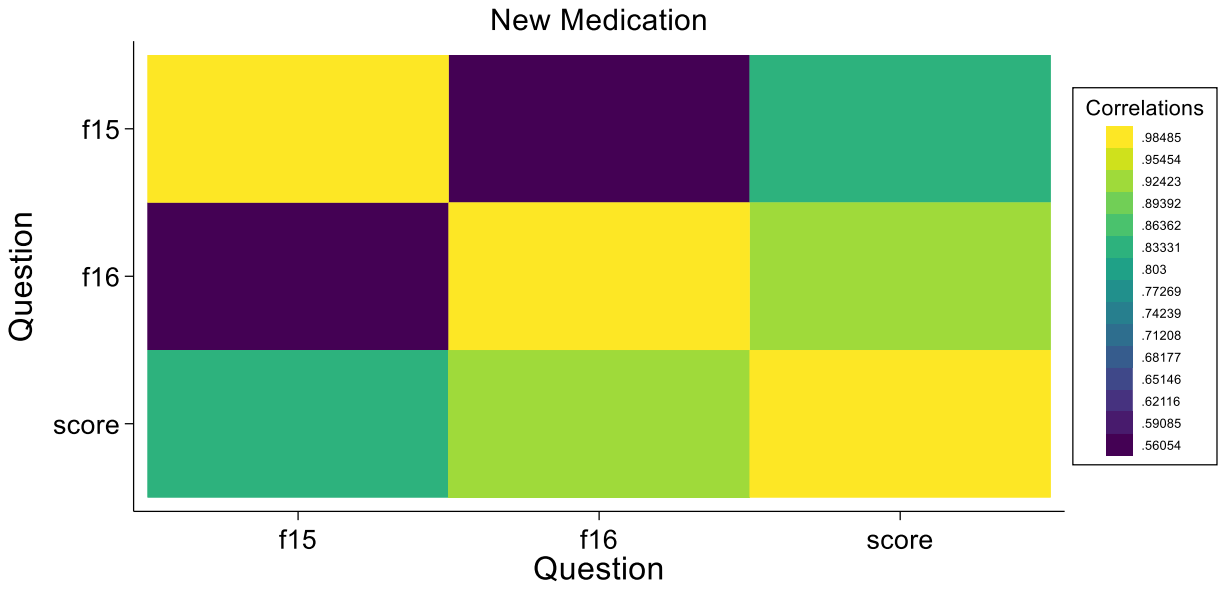
New Medication

Similar to the *Environment* theme, the *New Medication* theme is based on questions from the corresponding section of the questionnaire and exhibits comparable values. The correlation structure is consistent across all regions (Figure 57), and the alpha value falls more clearly within the expected range (Table 37).

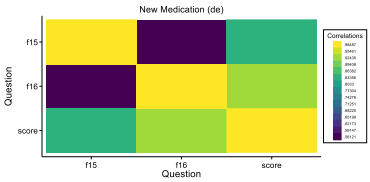
Table 37 Internal consistency of the *New Medication* theme. The overall Cronbach's alpha is reported on the first line. Since the theme consists of only two items, discarding either would prevent the construction of an aggregated score.

Question	Questions	Variance	Cronbach's α	Difference
New Medication	2	0.55	0.68	
15. Purpose of medication	1			
16. Side effects	1			

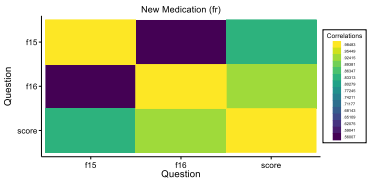
CH



DE



FR



IT

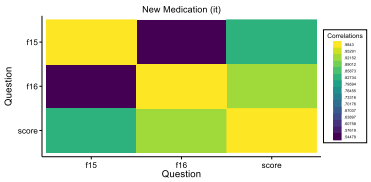


Figure 57 Correlations between items and the overall score within the *New Medication* theme. The upper row displays results for the entire sample, while the lower row shows language-specific subsamples. Ideally, correlations should be higher (yellow) between each item and the overall score, and lower (blue) between individual items.

Respect

The *Respect* theme is based on identical questions directed at doctors (question 1) and nursing staff (question 5). The correlation patterns appear as expected (Figure 58), and the overall alpha value falls slightly below the optimal range (Table 38).

Table 38 Internal consistency of the *Respect* theme. The overall Cronbach's alpha is reported on the first line. Since the theme consists of only two items, discarding either would prevent the construction of an aggregated score.

Question	Questions	Variance	Cronbach's α	Difference
Respect	2	0.07	0.58	
1. Courtesy and respect (doc.)	1			
5. Courtesy and respect (nurs.)	1			

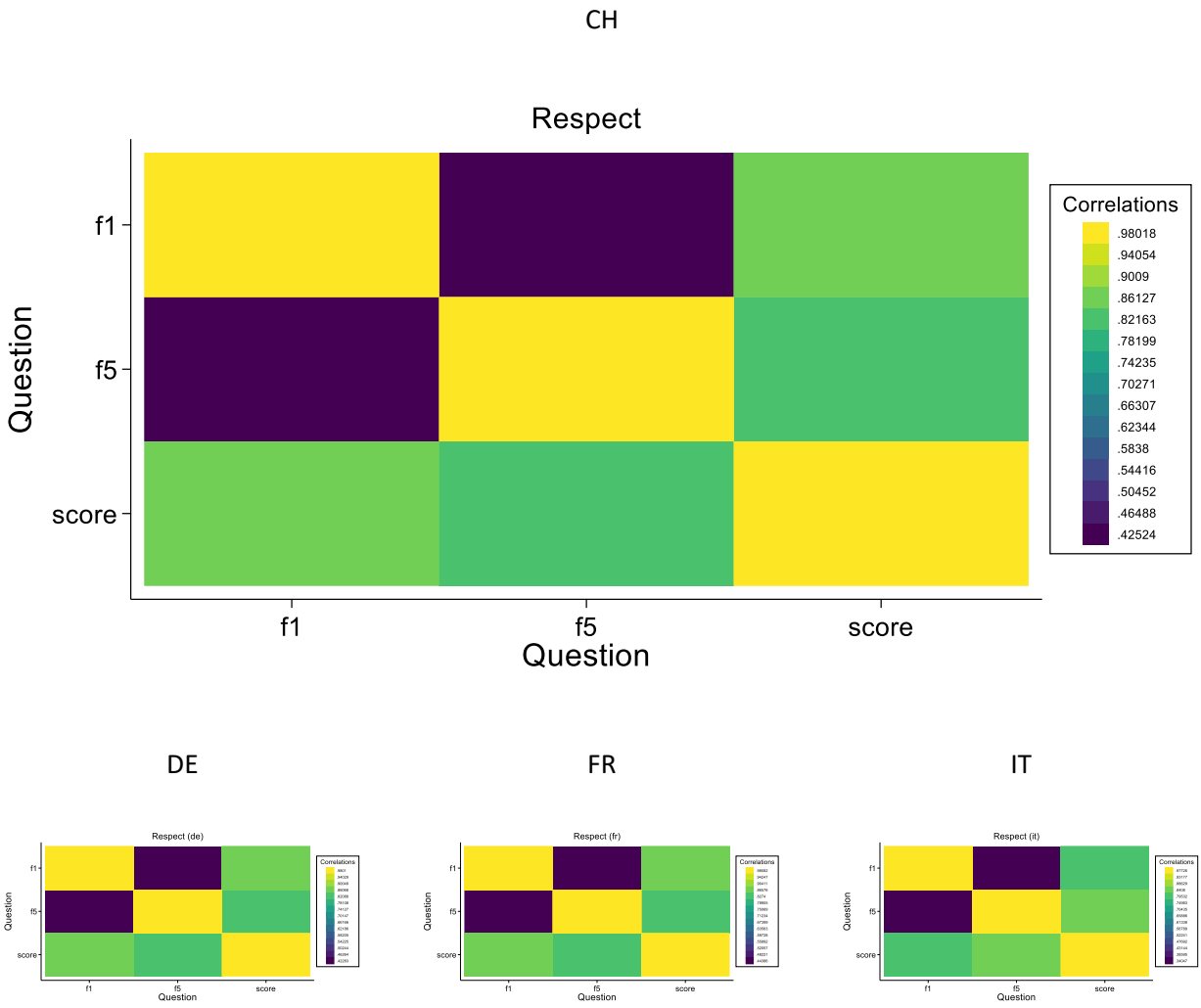


Figure 58 Correlations between items and the overall score within the *Respect* theme. The upper row displays results for the entire sample, while the lower row shows language-specific subsamples. Ideally, correlations should be higher (yellow) between each item and the overall score, and lower (blue) between individual items.

All eight themes exhibit the expected properties: low inter-item correlations, high item-score correlations, and consistency across the three linguistic regions. Alpha values fall within the optimal range of 0.70–0.80, **except for** the two-item themes *Respect* and *Environment*, which show lower values of 0.58 and 0.41, respectively.

Discriminative Power

Item Response Theory (IRT) tools can be used to evaluate how each item within a theme contributes to distinguishing between patients who report better or worse experiences on the overall score.

Admission

In Figure 59, the item information functions (IIFs) for Questions 25 and 26 overlap, suggesting potential redundancy between the two questions addressing emergency admissions.

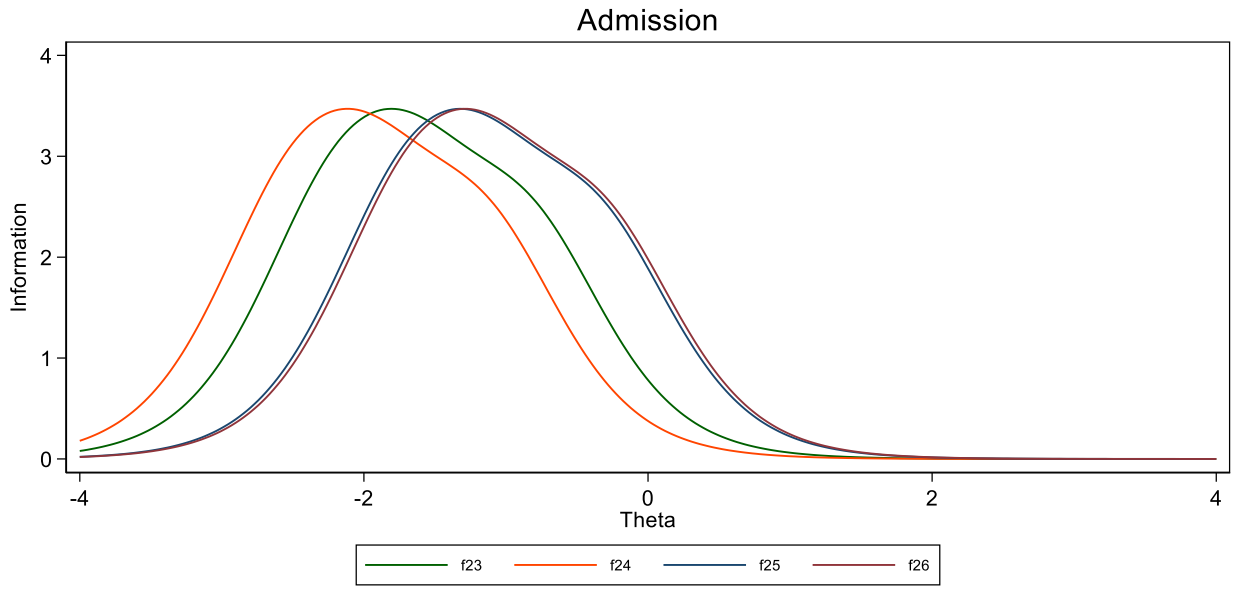


Figure 59 Discriminative power of items within the *Admission* theme.

Information

In Figure 60, two pairs of item information functions (IIFs) show overlap, suggesting that questions about attentive listening by doctors (Question 2) and nursing staff (Question 6) may be redundant. In contrast, questions about understandable explanations (Questions 3 and 7) do not exhibit such redundancy. However, the doctor's understandable explanations (Question 3) closely resemble the general understanding of the patient's condition (Question 32).

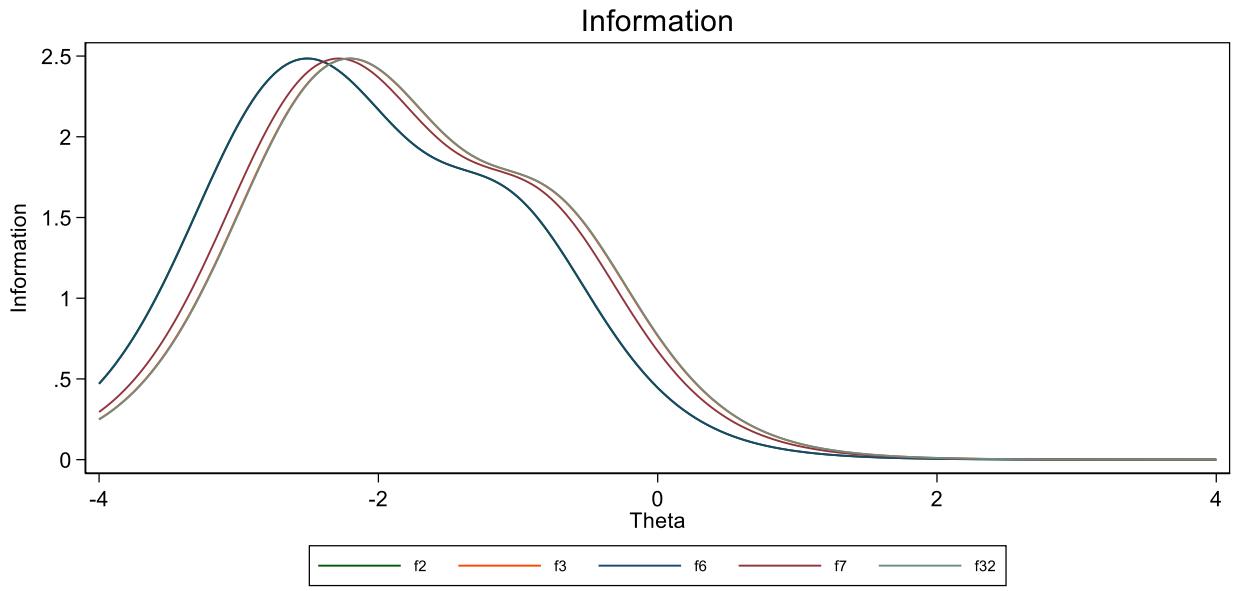


Figure 60 Discriminative power of items within the *Information* theme.

Involvement

In Figure 61, the item information functions (IIFs) for both pairs do not overlap, suggesting that the two questions are not redundant.

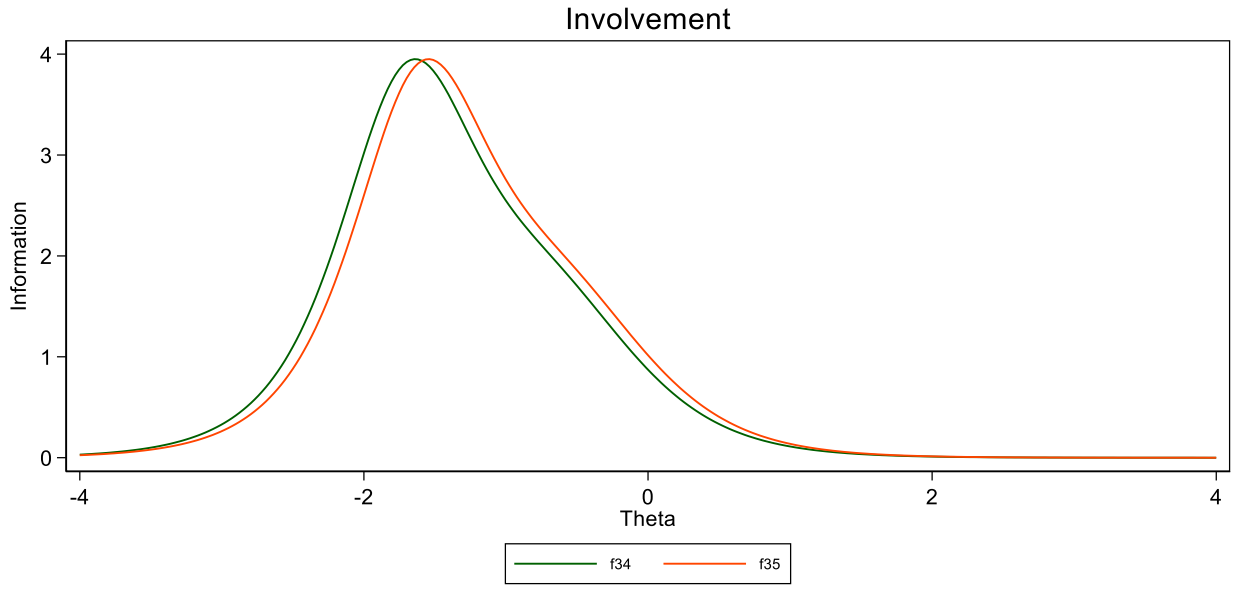


Figure 61 Discriminative power of items within the *Involvement* theme.

Organization

In Figure 62, the item information functions (IIFs) for Question 29, concerning communication, and Question 30, regarding treatment information, show overlap, suggesting potential redundancy.

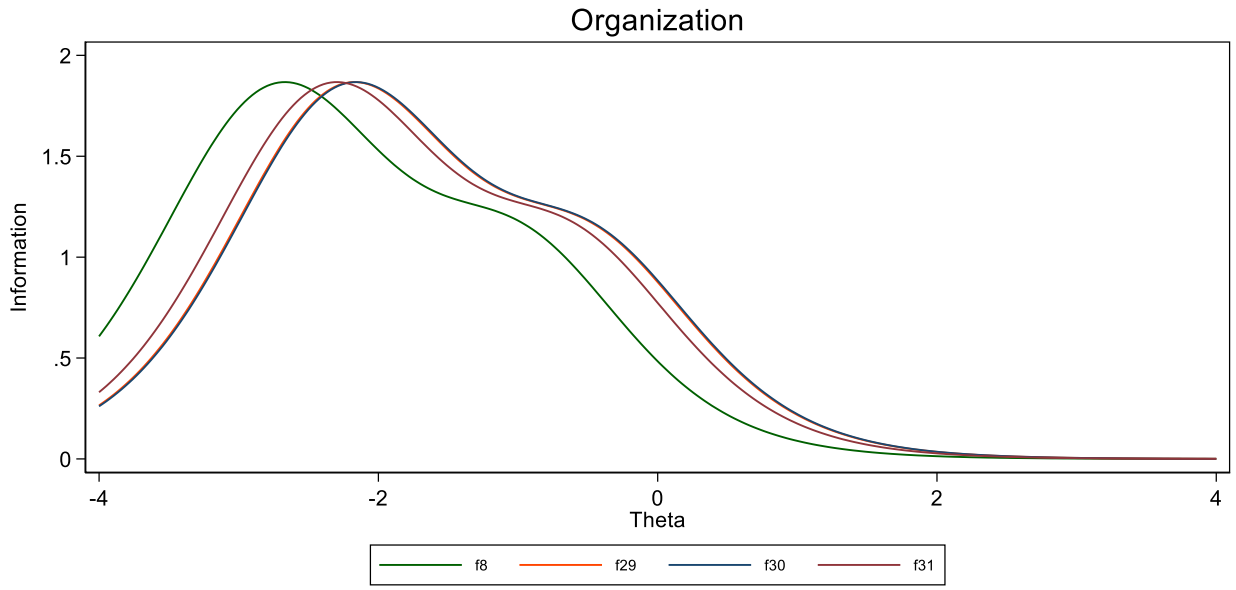


Figure 62 Discriminative power of items within the *Organization* theme.

Discharge

In Figure 63, the item information functions (IIFs) do not overlap, suggesting that the items are likely not redundant.

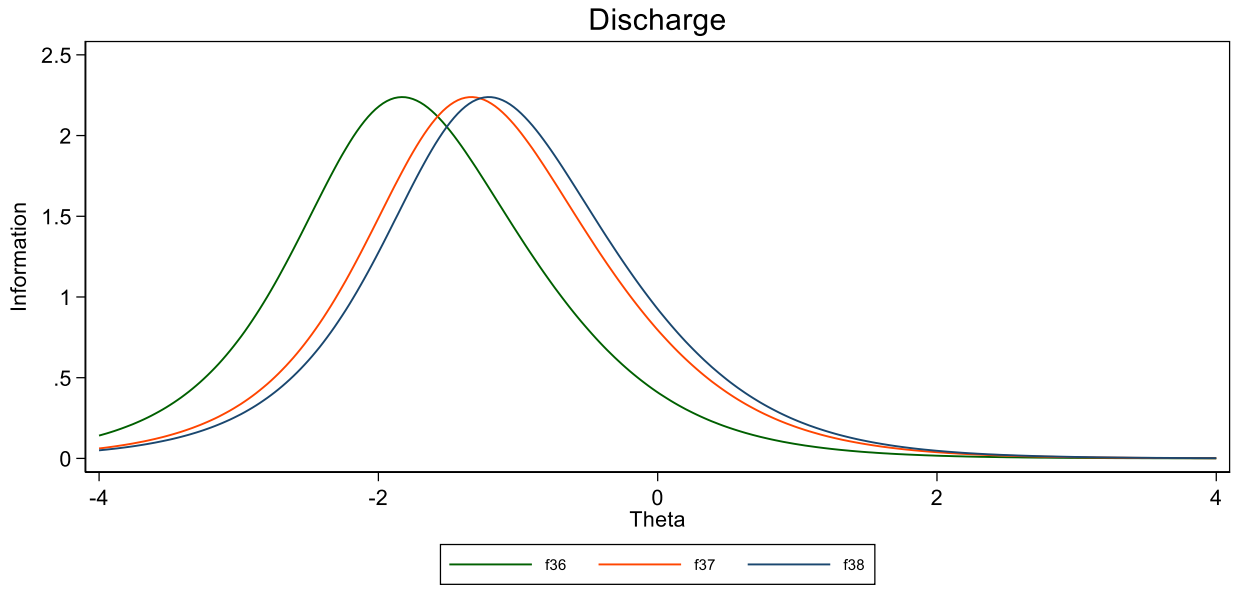


Figure 63 Discriminative power of items within the *Discharge* theme.

Environment

In Figure 64 the item information functions (IIFs) clearly do not overlap, indicating that the items are likely not redundant.

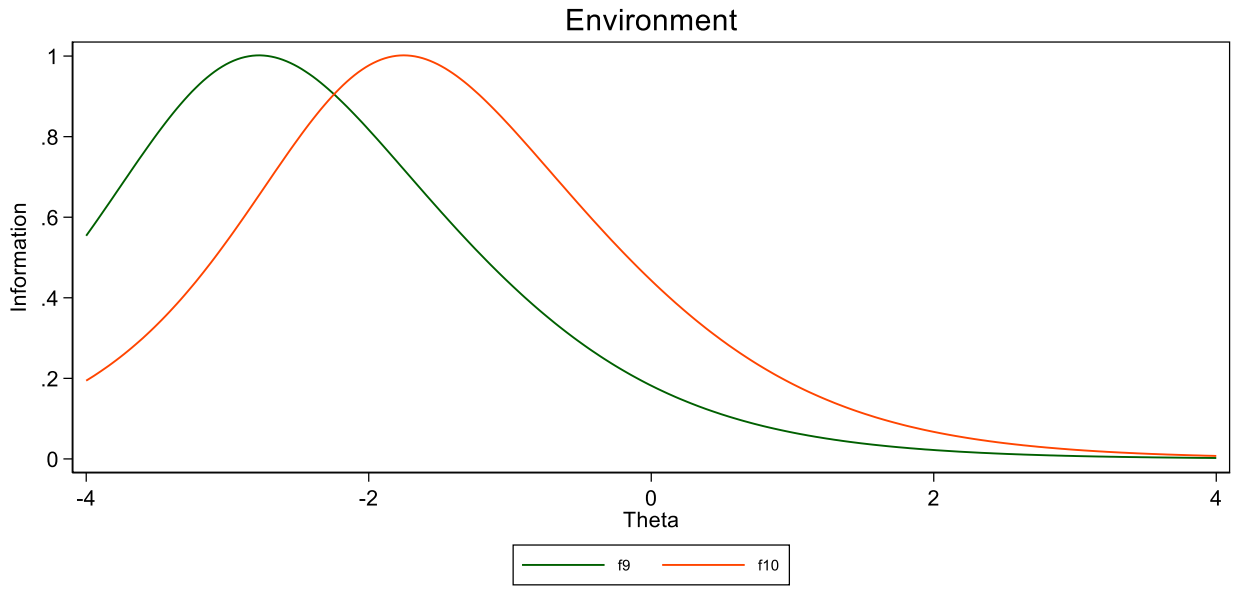


Figure 64 Discriminative power of items within the *Environment* theme.

New Medication

In Figure 65. the item information functions (IIFs) do not overlap, suggesting that the two questions are likely non-redundant.

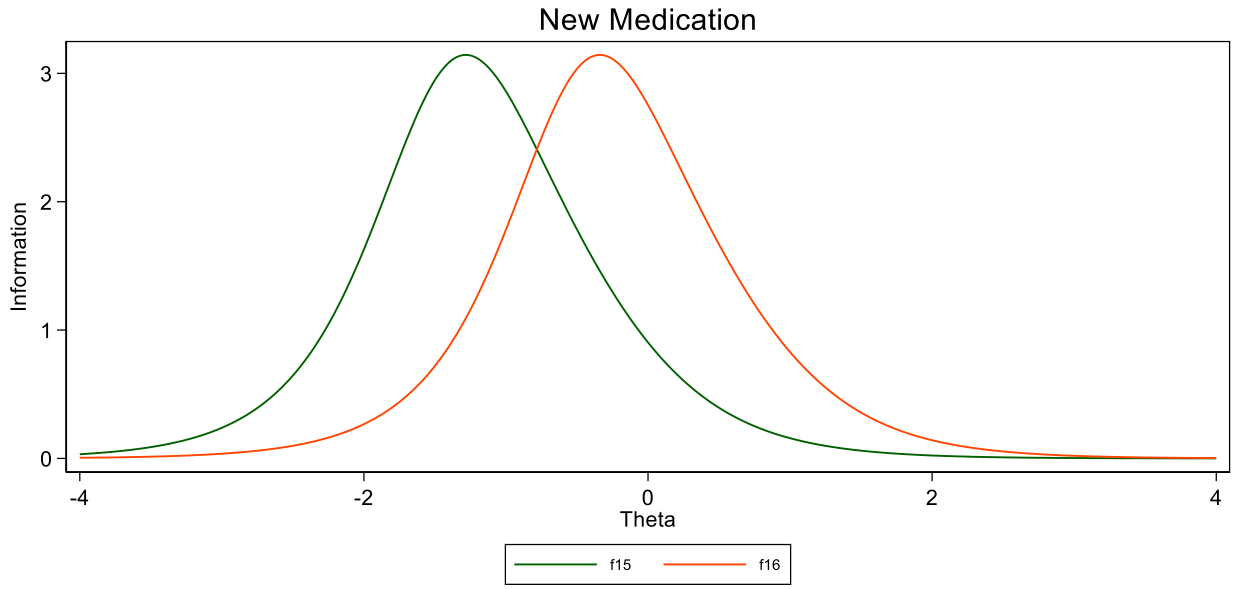


Figure 65 Discriminating power of items inside the *new medication* theme.

Respect

In Figure 66, the item information functions (IIFs) do not overlap, suggesting that the two questions are likely non-redundant.

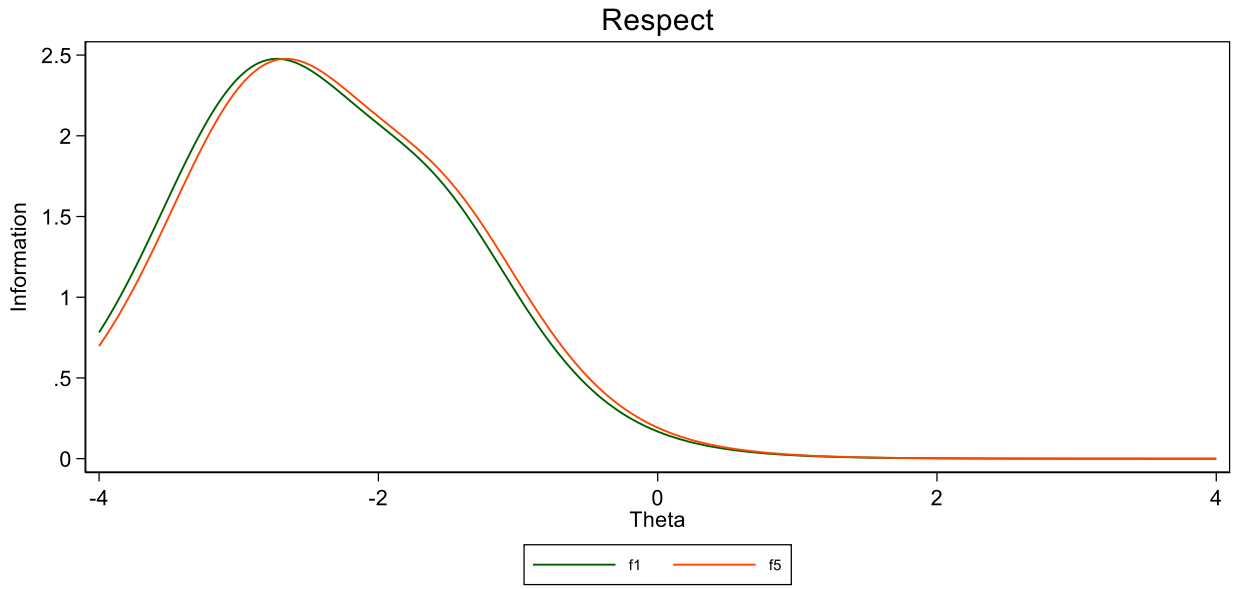


Figure 66 Discriminating power of items inside the *respect* theme.

Influence of Data Collection

Figure 67 shows that national PREMs scores are not influenced by the data collection process. Nearly identical results were observed between the paper and electronic versions, despite minor socio-demographic differences. There is no indication that difficulties in understanding and completing the questionnaire affects patients' experience ratings-most respondents did not report difficulties, hence the smaller confidence intervals.

The only notable variation appears across linguistic versions, with consistently lower scores from the French-speaking sample. This discrepancy likely reflects a well-known tendency among these patients to express more dissatisfaction, rather than issues with translation.

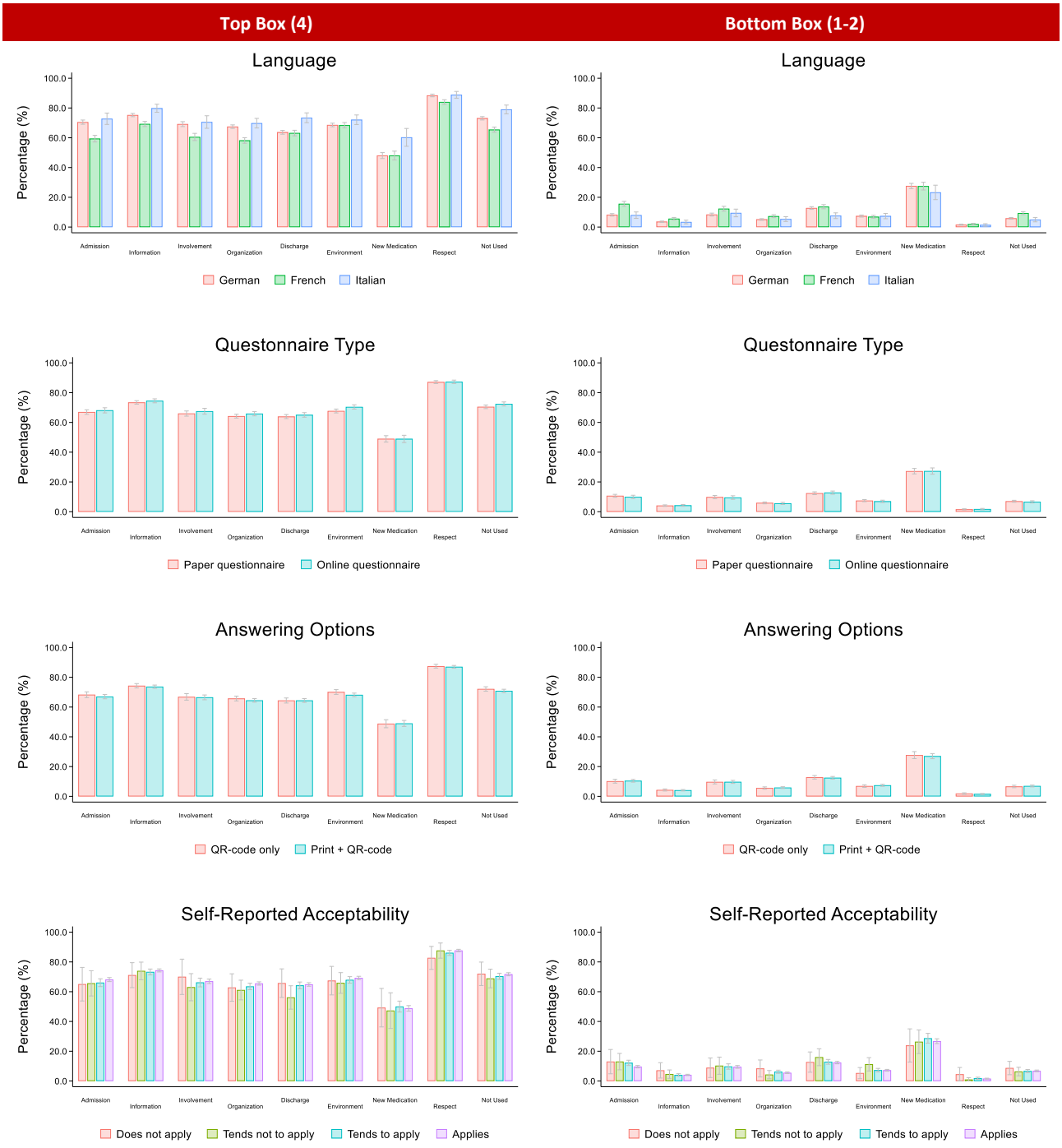
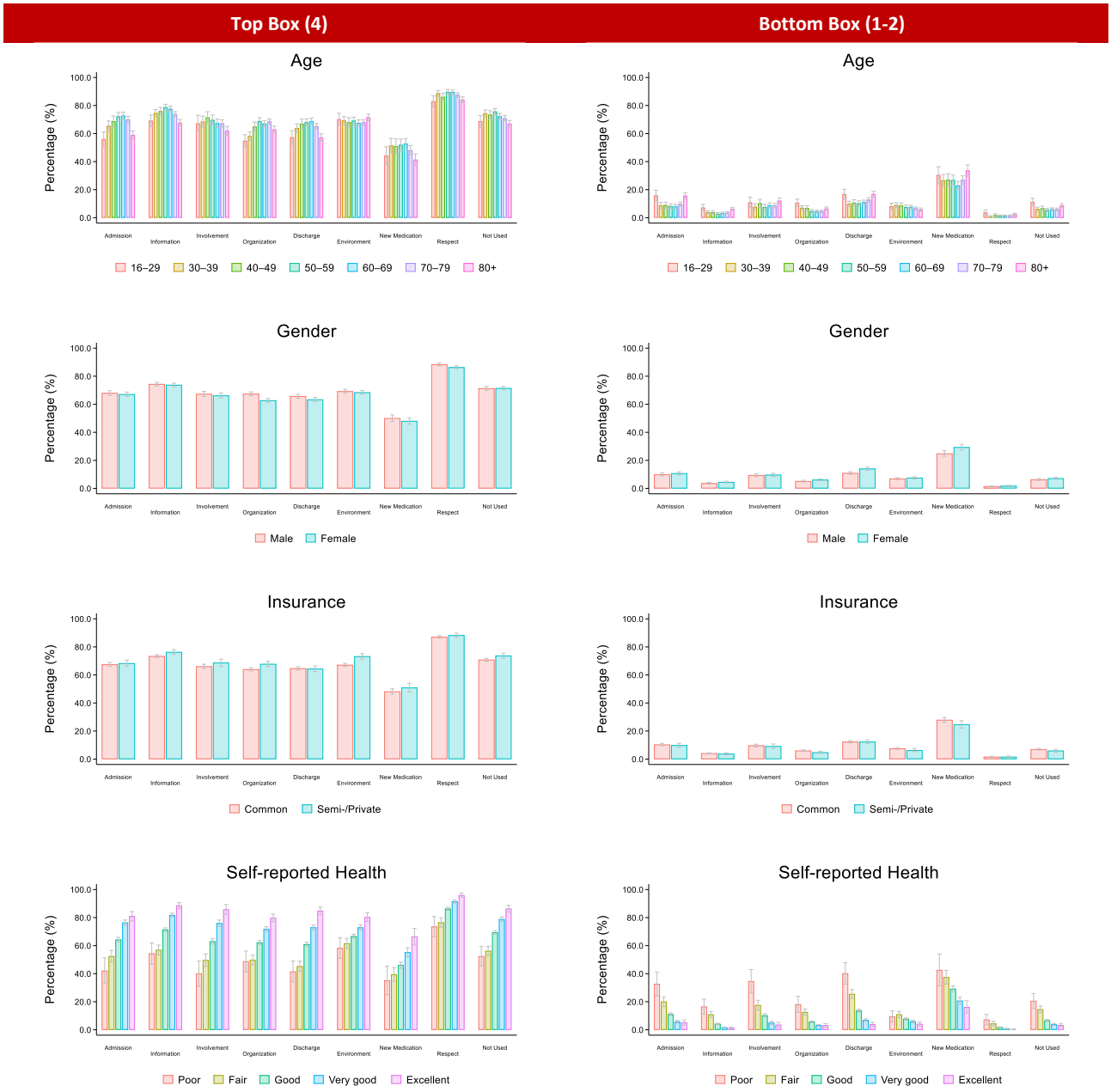


Figure 67 Differences in national averages of top and bottom box proportions of the eight thematic categories according to different questionnaire versions and data collection processes, as measured on the acute care pilot sample.

Influence of Socio-Demographic Factors

More differences emerge when comparing various socio-demographic subsamples. Figure 68 clearly shows that patients in better health tend to rate their experience more positively across all thematic categories. Age also plays a role, with younger and older patients giving lower ratings than those in their 50s and 60s. For the other three variables, the patterns are less pronounced. Although the differences are small and not statistically significant, they are consistent across all thematic categories: females, individuals with lower education levels, and patients with basic insurance coverage tend to rate their experience slightly lower than their respective counterparts.



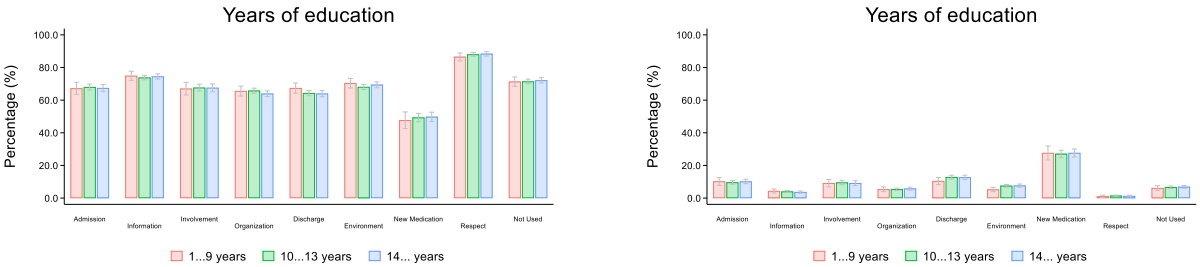


Figure 68 Differences in national averages of top and bottom box proportions of the eight thematic categories according to socio-demographic factors, as measured on the acute care pilot sample.

5.2. Detail of Results in Psychiatry and Rehabilitation

The same analyses conducted in acute care were also performed in the psychiatry and rehabilitation domains, albeit less thoroughly. For the sake of completeness and transparency, detailed results are provided in the following sections.

5.2.1. Data Quality Measurements

Sample

The psychiatry and rehabilitation studies were designed to replicate and confirm findings from acute care. To reduce the burden on hospitals, the sample size was limited to approximately one-quarter of that used in acute care. Despite the reduced sample size, hospitals were selected with attention to diversity and representativeness. Table 18 outlines the distribution across linguistic regions and hospital types.

Table 39 Hospitals participating in the Swiss PREMs Pilot Studies in acute care, by linguistic region and hospital type.

	Psychiatry		Rehabilitation	
Total Hospitals	37		34	
In German-speaking part	30	81.1%	22	64.7%
In French-speaking part	5	13.5%	10	29.4%
In Italian-speaking part	2	5.4%	2	5.9%
n/a	4	10.8%	2	5.9%
K111 – University Hospital	3	8.1%		
K112 – Medical Center Level 2			5	14.7%
K121 – Primary Care Level 3			1	2.9%
K122 – Primary Care Level 4	1	2.7%	1	2.9%
K123 – Primary Care Level 5				
K211 – Psychiatric Clinic Level 1	14	37.8%		
K212 – Psychiatric Clinic Level 2	10	27.0%		
K221 – Rehabilitation Clinic	5	13.5%	22	64.7%
K234 – Geriatric Clinic			2	5.9%
K235 – Specialized Clinic			1	2.9%

Missig Values

Horizontal patterns are more pronounced than vertical ones in Figure 69, suggesting that certain psychiatric clinics have patients who tend to respond to fewer questions than others. In this case, the higher rate of missing values is not necessarily due to smaller sample sizes, as some of these clinics returned more than twenty questionnaires.

In contrast, vertical patterns are more evident in rehabilitation (see Figure 70), indicating that missing values are more likely associated with specific questions. Particularly those from the *Your experience in this clinic* section appear across all domains of care (see Figure 24 for the acute care case).

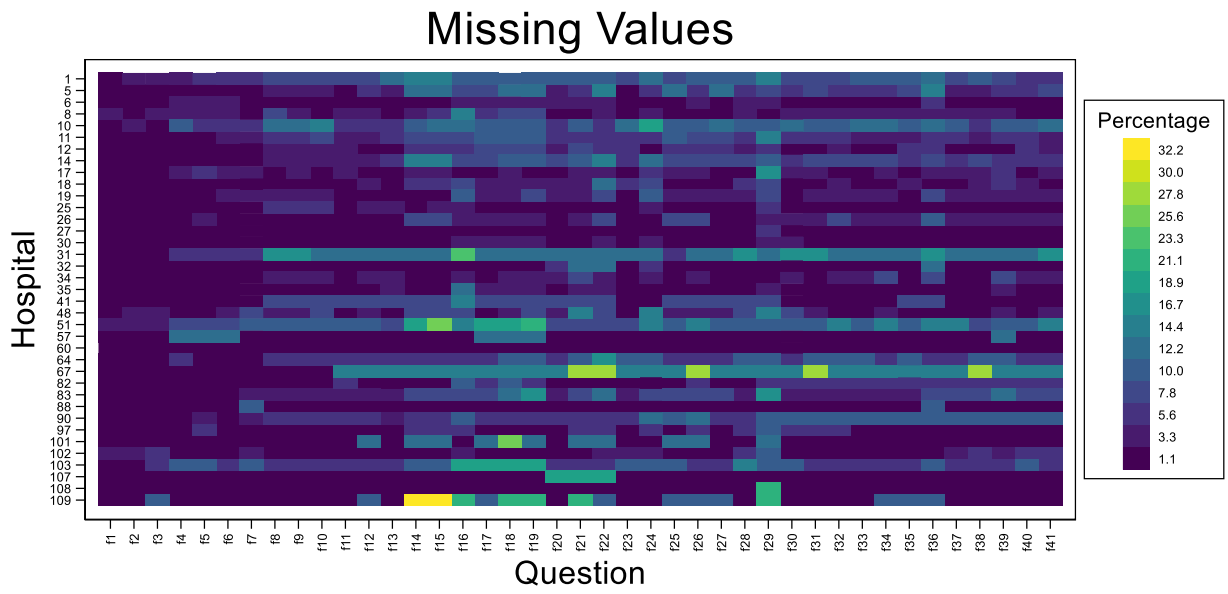


Figure 69 Heat plot of missing values in the psychiatry pilot study. Each row corresponds to a hospital, each column to a question.

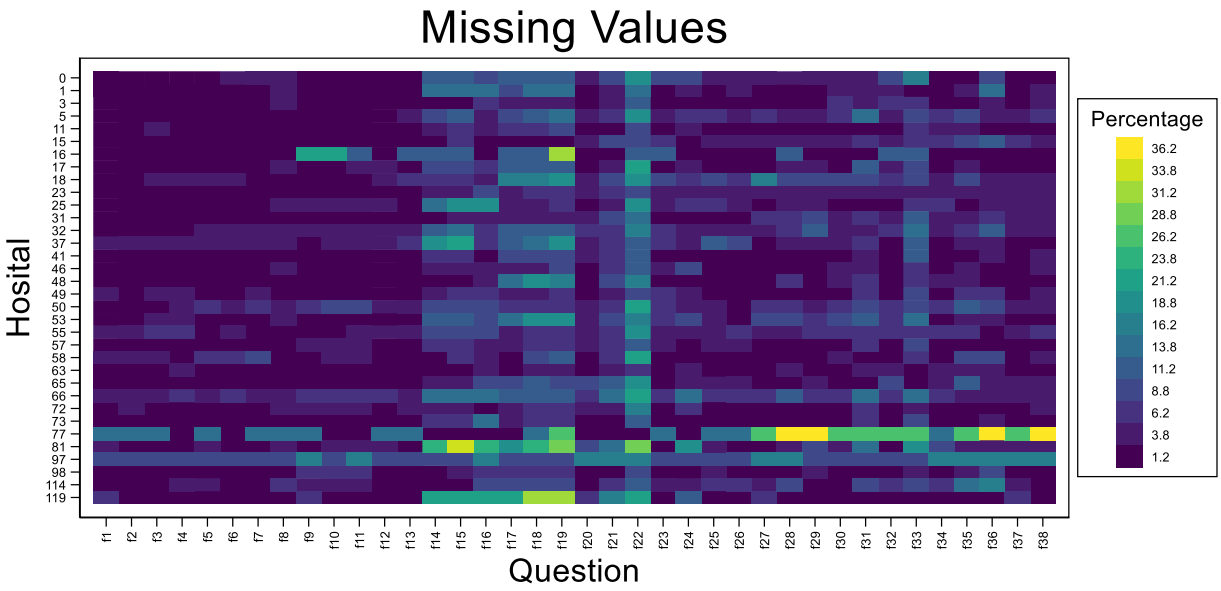


Figure 70 Heat plot of missing values in the rehabilitation pilot study. Each row corresponds to a hospital, each column to a question.

Alternative Answers

Patients across all hospitals consistently select alternative responses. The lack question regarding assistance with toileting appears prominently irrelevant in all psychiatric clinics (see Figure 71), and remains strongly, but less systematically, irrelevant across rehabilitation clinics (see Figure 72), suggesting a somewhat greater need for such assistance in certain facilities. Other questions reinforce the overall pattern illustrated in Figure 9.

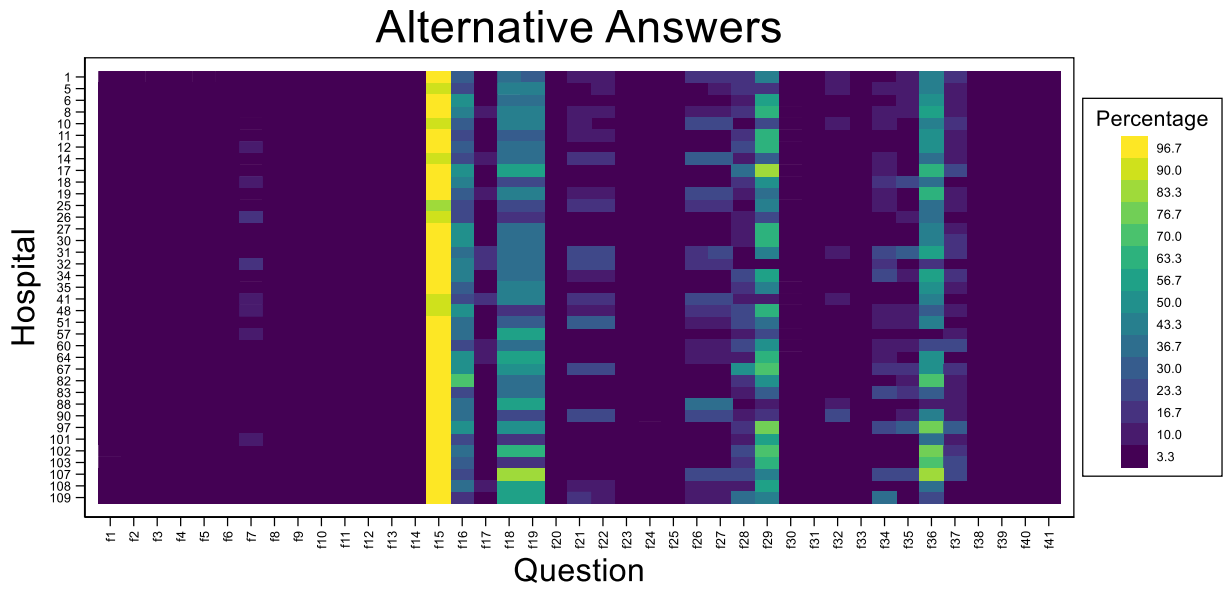


Figure 71 Heat plot of alternative answers in the psychiatry pilot study. Each row corresponds to a hospital, each column to a question.

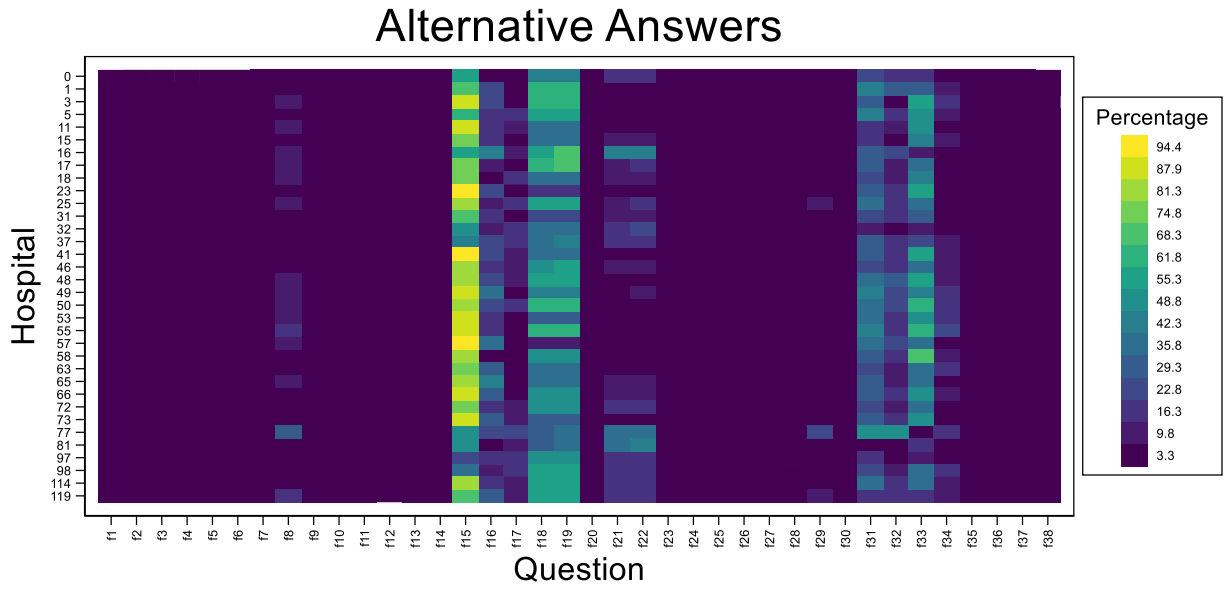


Figure 72 Heat plot of alternative answers in the rehabilitation pilot study. Each row corresponds to a hospital, each column to a question.

Influence of Data Collection

QR-Code Only Versus Paper + QR-Code

The main difference between the two collection modes is the response rates. When the freedom to choose the questionnaire is restricted (i.e. allowing only electronic forms sent by QR-code), almost half of the sample is lost in psychiatry as well as in rehabilitation (see Table 40)

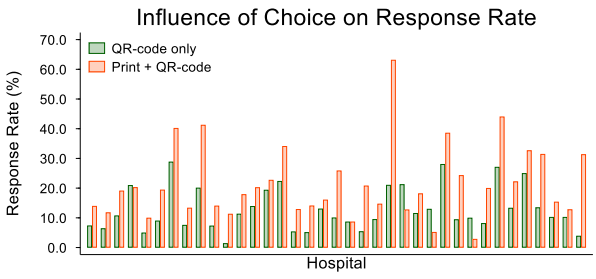
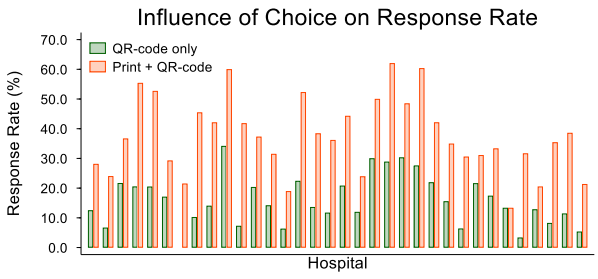
Table 40 Response rates measured in the pilot study when respondents are left free to choose between the paper and online questionnaires (top) or are restricted to the online form (bottom).

Collection Mode	Psychiatry	Rehabilitation
QR-Code + Paper	18.3%	37.9%
QR-Code Only	10.7%	16.5%

Differences in missing values and alternative answers are mainly due to respondents misunderstanding filters and questions to skip in the paper questionnaire, see Table 41. Thess differences are explained more in detail in the next section.

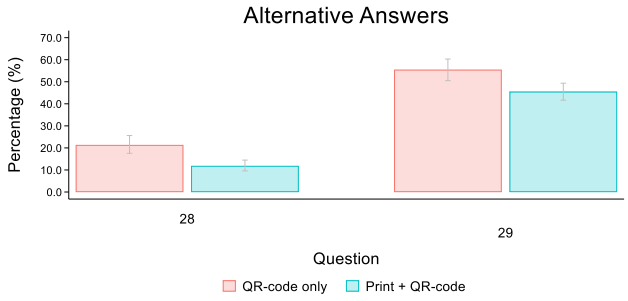
Ceiling and floor effects do not differ significantly between the QR-only and QR-code+paper collection modes.

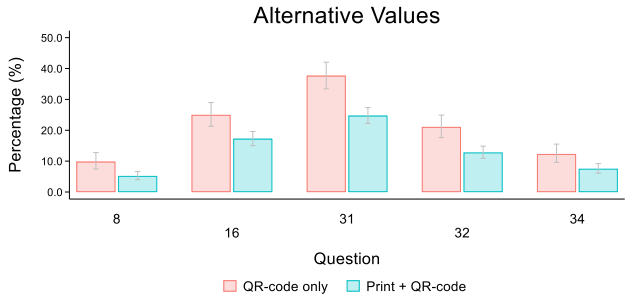
Table 41 Influence of the mode of contact on data quality metrics. The general trend can be observed in every hospital participating in the pilot study: the response rate is higher when respondents have the choice between paper and electronic questionnaires (top). Questions where significant differences were observed between the two collection modes regarding missing values and alternative answers appear on the rows below.

Psychiatry	Rehabilitation
	

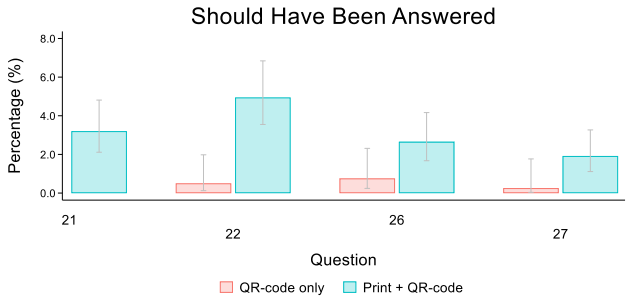

Missing Values
Percentage (%)
25.0
20.0
15.0
10.0
5.0
0.0
3 4 8 14 15 16 17 18 19 21 22 24 31 33
Question
QR-code only
Print + QR-code

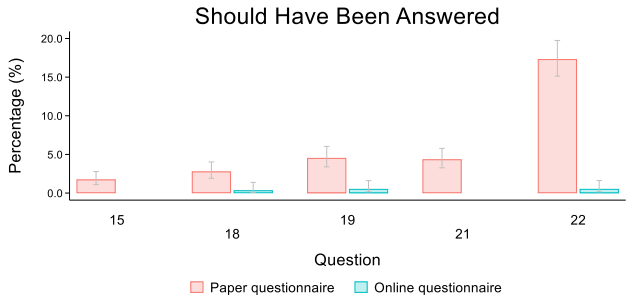

Missing Values
Percentage (%)
25.0
20.0
15.0
10.0
5.0
0.0
3 4 8 14 15 16 17 18 19 21 22 24 31 33
Question
QR-code only
Print + QR-code

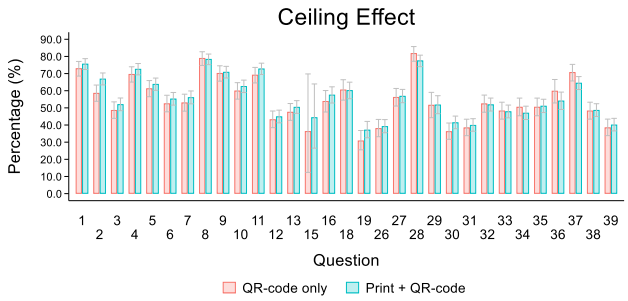

Alternative Answers
Percentage (%)
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
28 29
Question
QR-code only
Print + QR-code

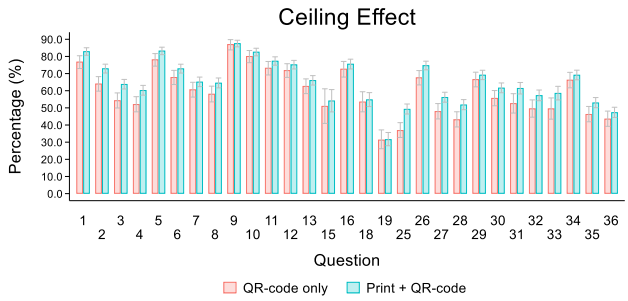

Alternative Values
Percentage (%)
50.0
40.0
30.0
20.0
10.0
0.0
8 16 31 32 34
Question
QR-code only
Print + QR-code

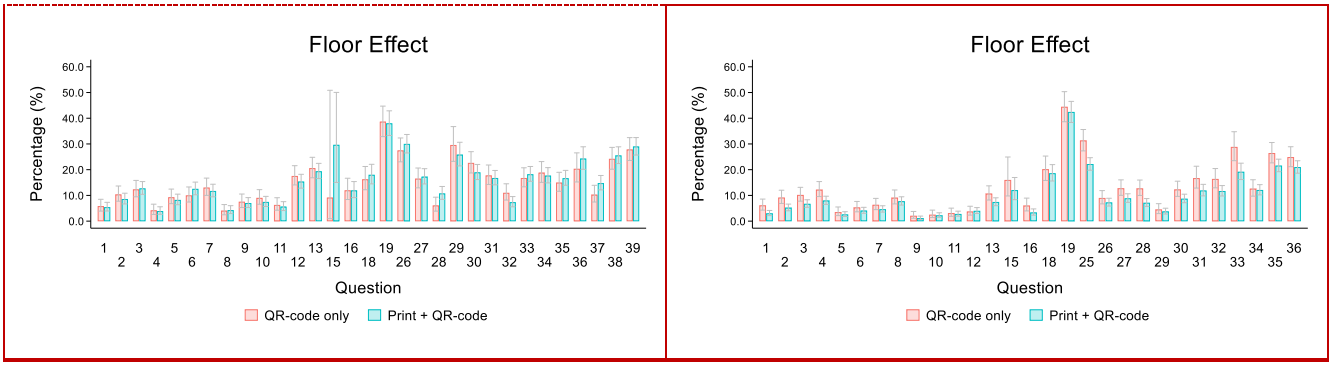

Should have Been Skipped
Percentage (%)
6.0
4.0
2.0
0.0
26 27
Question


Should Have Been Answered
Percentage (%)
8.0
6.0
4.0
2.0
0.0
21 22 26 27
Question
QR-code only
Print + QR-code


Should Have Been Answered
Percentage (%)
20.0
15.0
10.0
5.0
0.0
15 18 19 21 22
Question
Paper questionnaire
Online questionnaire


Ceiling Effect
Percentage (%)
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 15 16 18 19 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Question
QR-code only
Print + QR-code


Ceiling Effect
Percentage (%)
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 15 16 18 19 25 26 27 28 29 30 31 32 33 34 35 36
Question
QR-code only
Print + QR-code



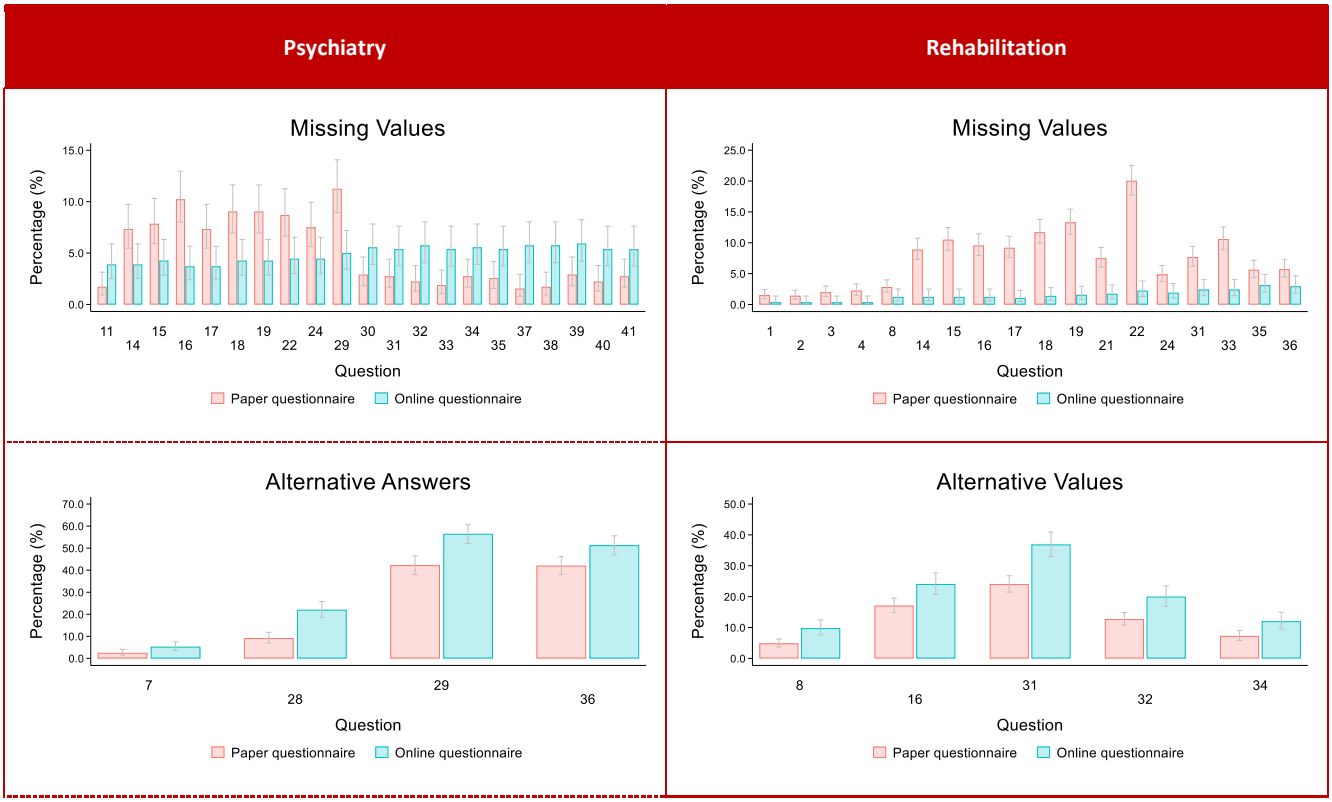
Online Forms Versus Paper Questionnaires

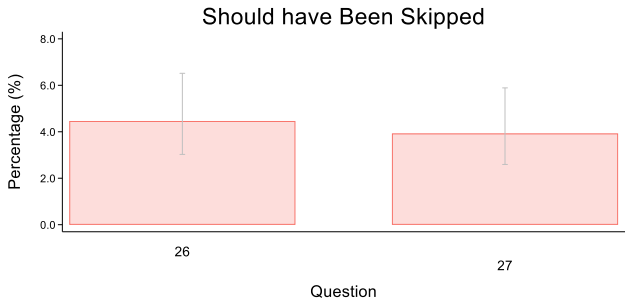
While the preceding section addresses how different ways of addressing respondents may affect data quality, this section addresses the effects of the questionnaire type: paper or online. This distinction is related to the previous one, but the two questionnaire types were mixed in the results of the Paper + QR-Code subsample.

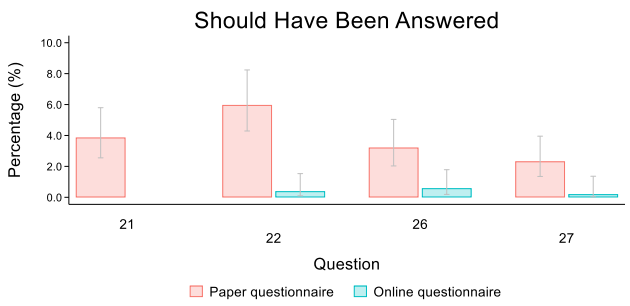
Respondents were guided by the online form and could not misunderstand the questionnaire's structure. These problems appear in the paper questionnaires. Respondents of the online form show higher rates of alternative answers ("question skipped") while respondents filling the paper questionnaire may have overseen the filter and forgot to answer questions (missing value), see Figure 73.

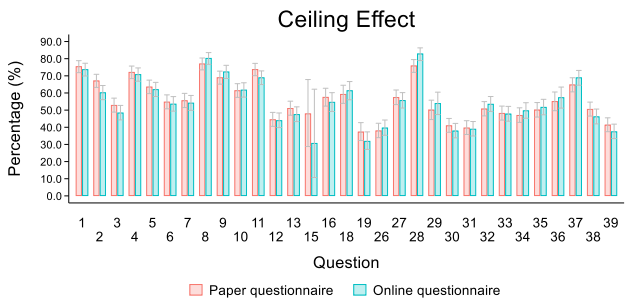
Ceiling and floor effects are not affected by the questionnaire type.

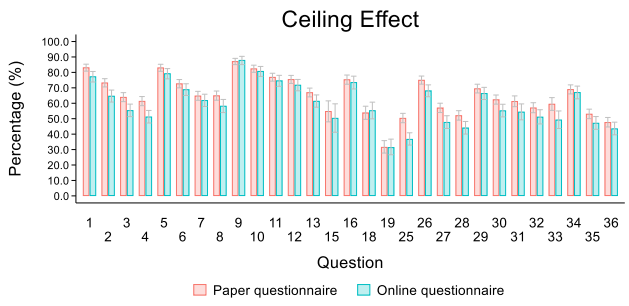
Figure 73 Influence of the type of questionnaire on data quality metrics. Questions where significant differences were observed between the two collection modes regarding missing values (top) and alternative answers (middle). Answers provided despite filters appear on the bottom line.

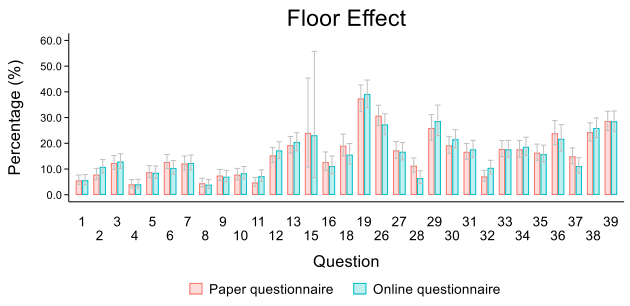


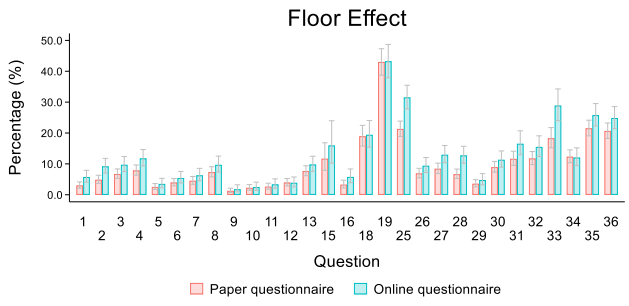












Translations

The different translations of the questionnaires yield comparable results. The only notable difference in Figure 74 is the significantly higher proportion of German-speaking respondents who indicated that question 29 (concerning unjustified restrictions of freedom of movement in in psychiatry care) was not applicable. This may reflect cultural differences rather than flaws in the translation process.

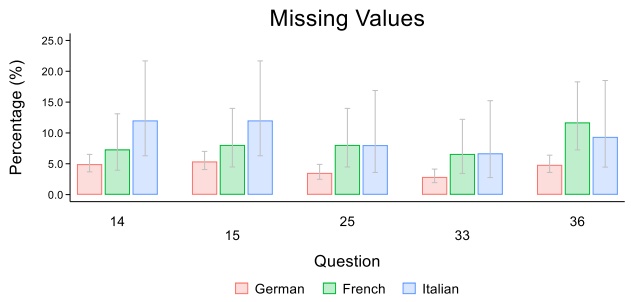
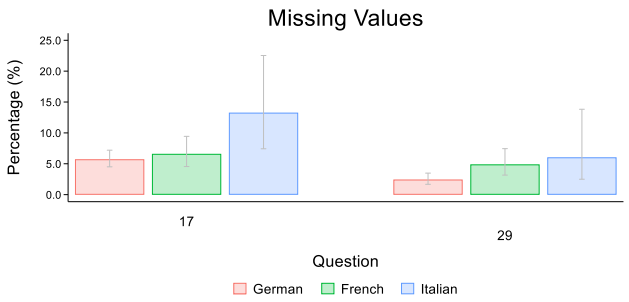
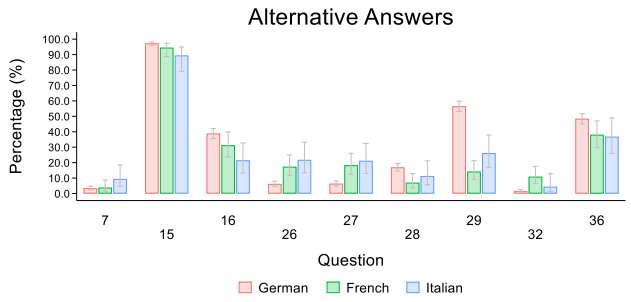
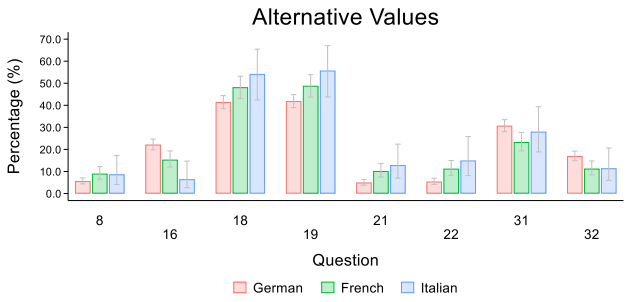
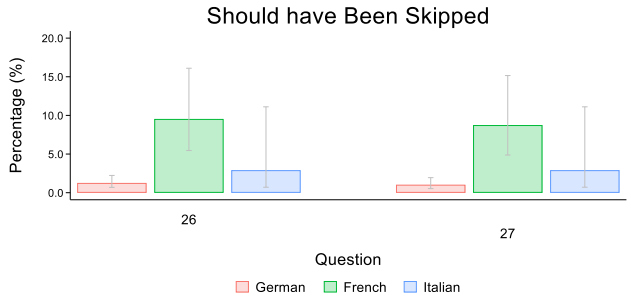
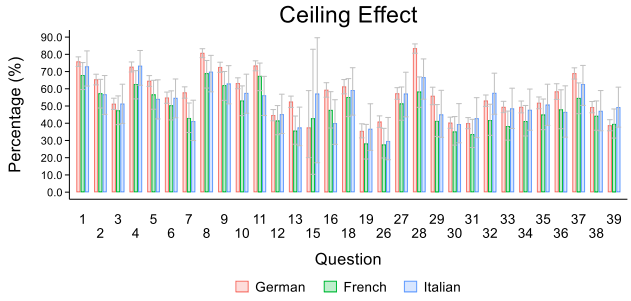
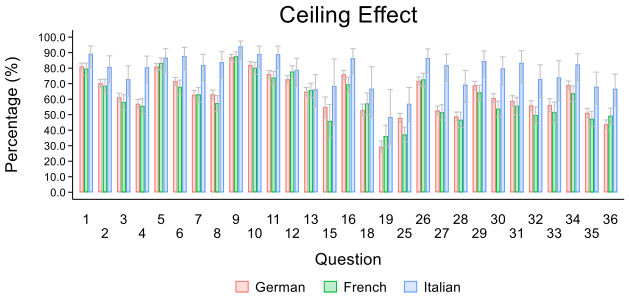
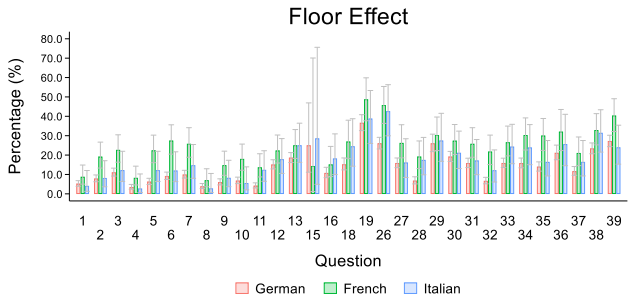
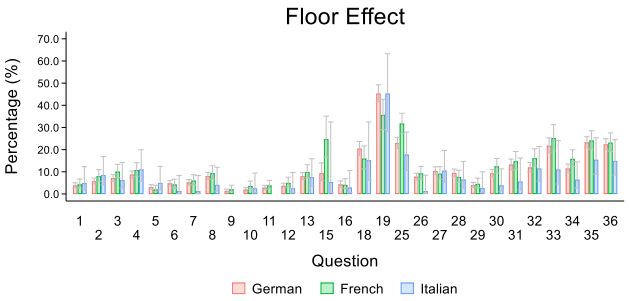
Psychiatry	Rehabilitation
	
	
	
	
	

Figure 74 Questions showing significant cross-linguistic differences in data quality-metrics in the psychiatry and rehabilitation studies. Proportion of missing values (top), proportion of alternative answers (middle) and proportions of responses that should not have been skipped due to the filter question (bottom).

Self-Reported Acceptability

The final control question on self-reported acceptability, assessing the extent to which the statement "It was easy for me to answer the questionnaire" applies, does not appear to influence the proportion of missing values or alternative responses in Figure 75. The group of respondents who reported the greatest difficulty is the only one that deviates from the others. However, it is also the smallest group, resulting in wide confidence intervals that prevent firm conclusions. Nevertheless, results are similar to those observed in acute care (see Figure 35).

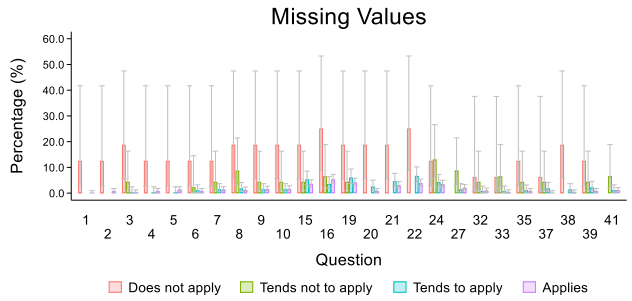
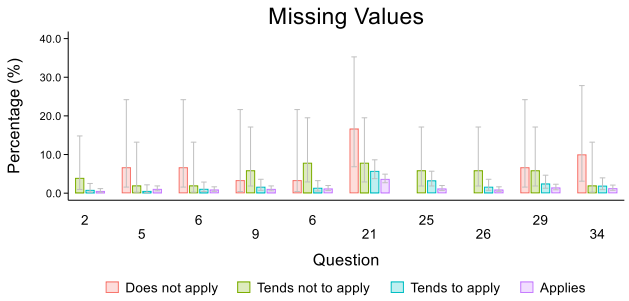
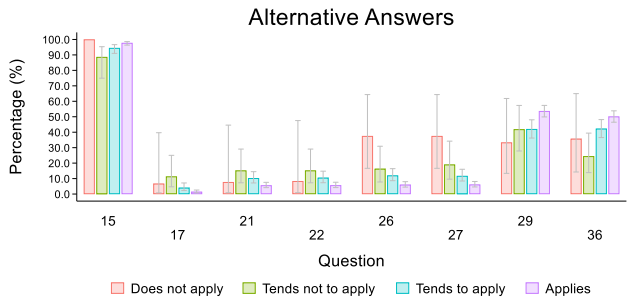
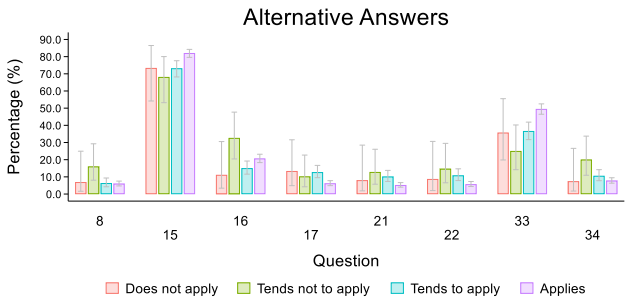
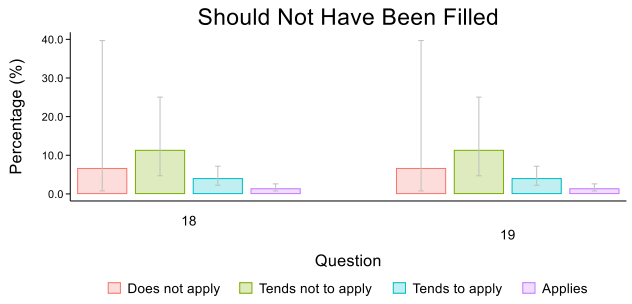
Psychiatry	Rehabilitation
	
	
	

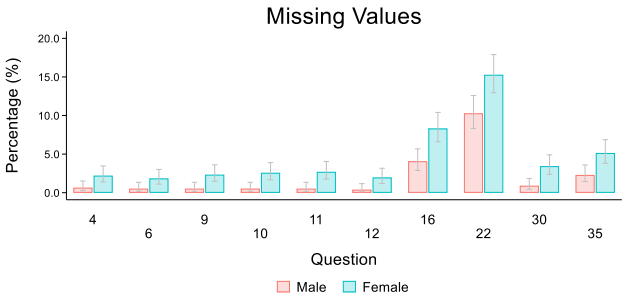
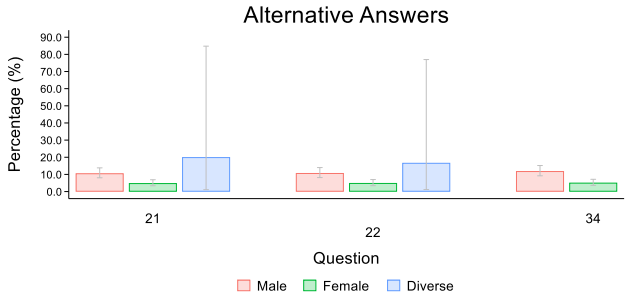
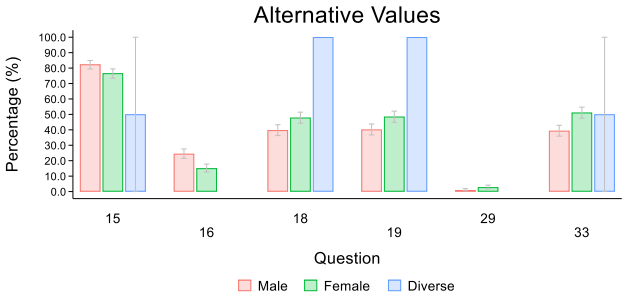
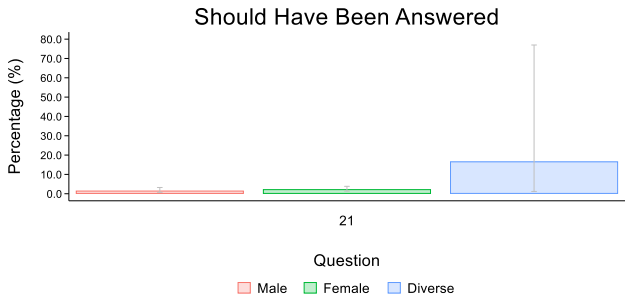
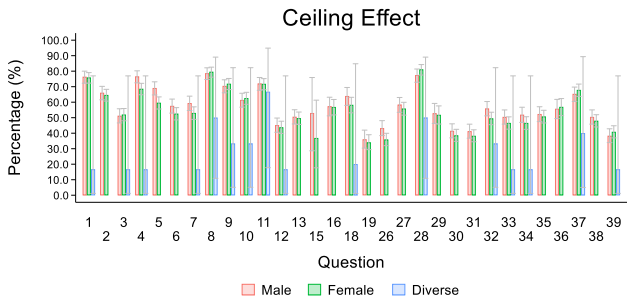
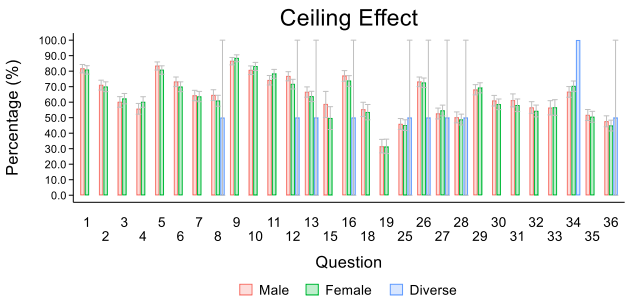
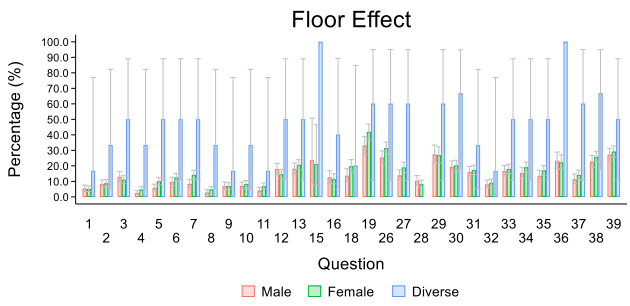
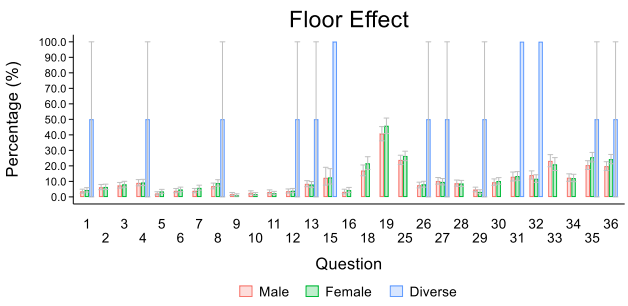
Figure 75 Questions showing differences in data quality-metrics across levels of self-reported acceptability. Proportion of missing values (top), proportion of alternative answers (middle) and proportions of responses that should not have been skipped due to the filter question (bottom).

Influence of Socio-Economic Factors

Gender

Most of the observed differences originate in the third “Diverse” category, which has been chosen by a very small number of respondents: 6 respondents in psychiatry (0.6% of persons having declared a gender) and 2 (0.1%) in rehabilitation. Their PREMs ratings is systematically less favorable than the two other categories, but the very great difference in values is due to proportions being computed on such a small collective, see Figure 76.

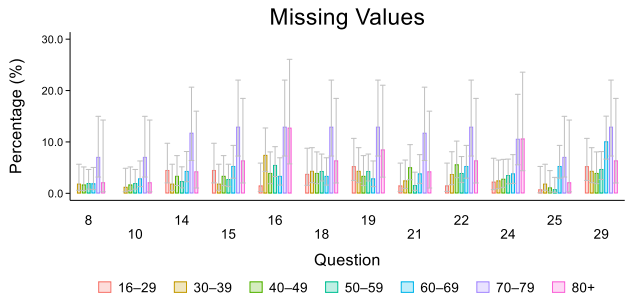
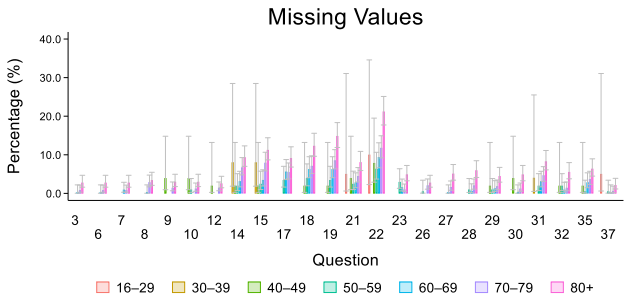
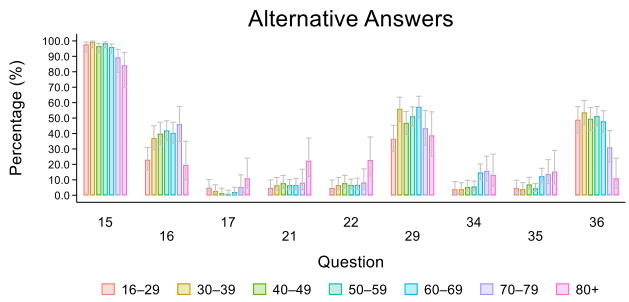
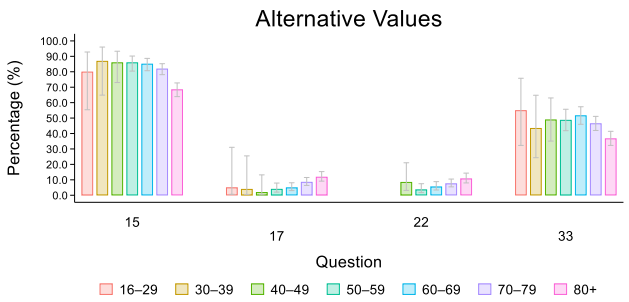
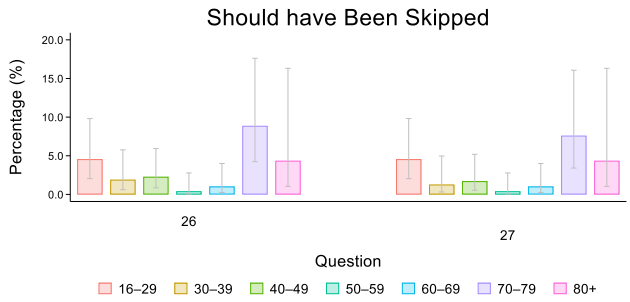
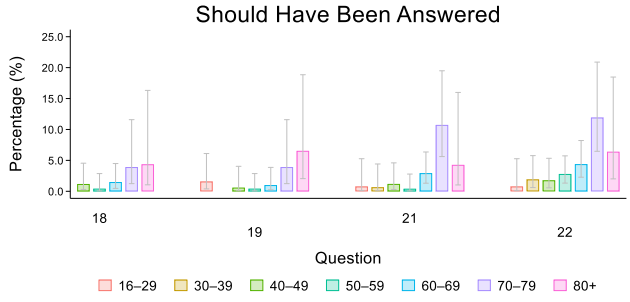
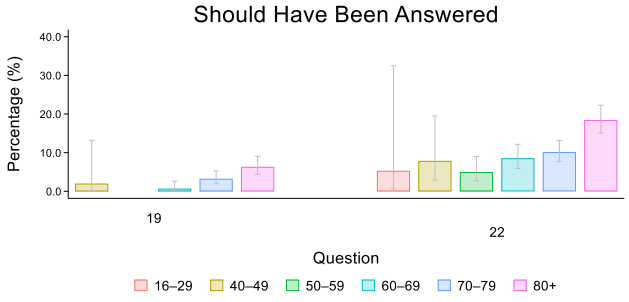
Figure 76 Gender differences in data quality metrics observed in the pilot samples from Psychiatry and Rehabilitation.

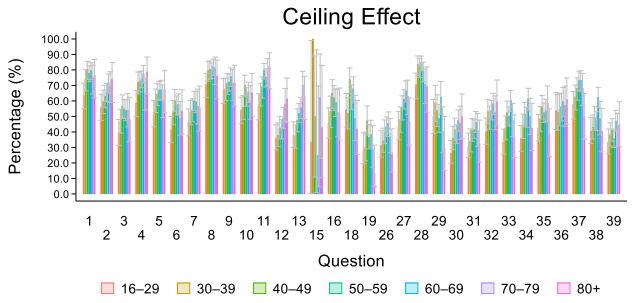
Psychiatry	Rehabilitation
	
	
	
	
	

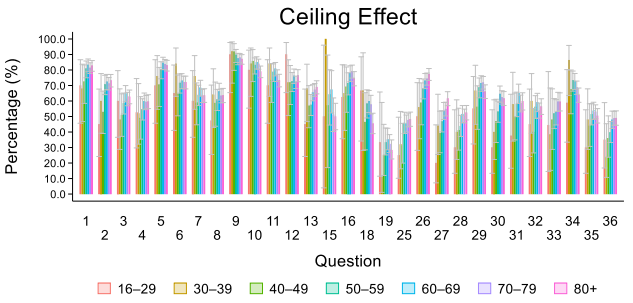
Age

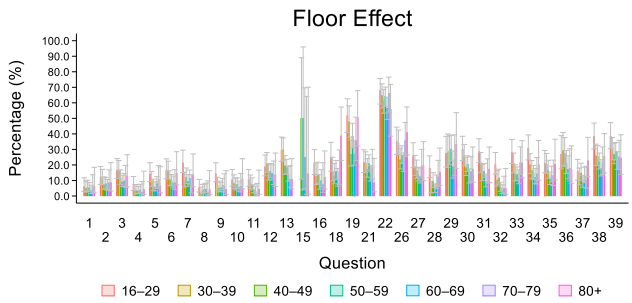
We observe a systematic trend in missing values in older respondents leaving more questions blank. This may also be related to the difficulties in understanding the filter and skip questions in the paper questionnaire (see Figure 41).

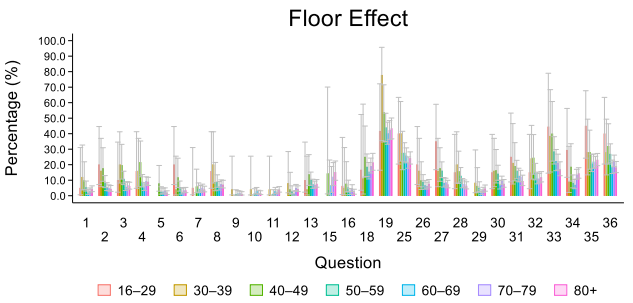
Figure 77 Age differences in data quality metrics observed in the pilot samples from Psychiatry and Rehabilitation.

Psychiatry	Rehabilitation
	
	
	
	


Ceiling Effect
Percentage (%)
Question
16–29
30–39
40–49
50–59
60–69
70–79
80+

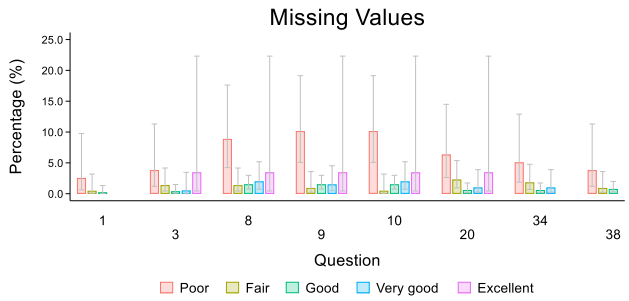
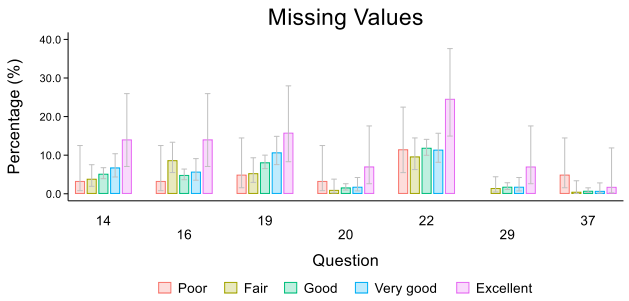
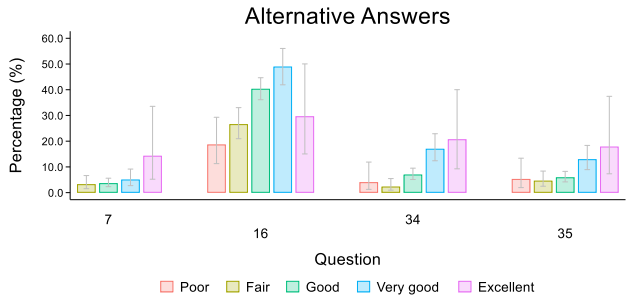
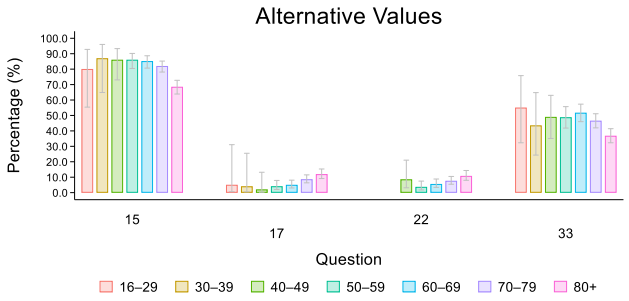
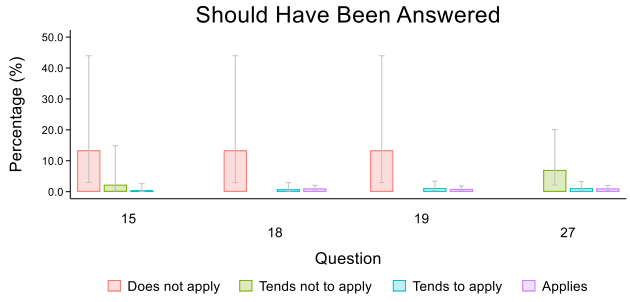
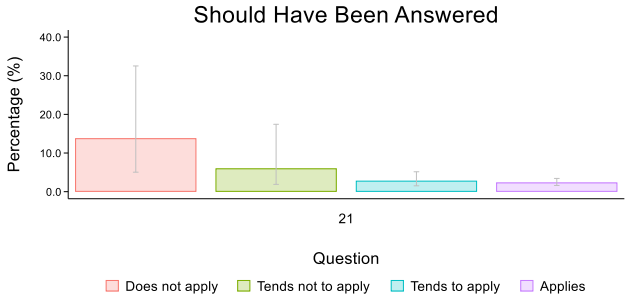

Ceiling Effect
Percentage (%)
Question
16–29
30–39
40–49
50–59
60–69
70–79
80+


Floor Effect
Percentage (%)
Question
16–29
30–39
40–49
50–59
60–69
70–79
80+


Floor Effect
Percentage (%)
Question
16–29
30–39
40–49
50–59
60–69
70–79
80+

Self-Reported Health

Figure 78 does not indicate a consistent influence of self-reported health on data quality metrics. Only a few questions exhibit significant differences, primarily within the sparsely populated extreme categories. The trend observed previously in acute care (see Figure 38), where improving health status is associated with more favorable patient experiences, is also confirmed psychiatry and rehabilitation.

Psychiatry	Rehabilitation
	
	
	
	

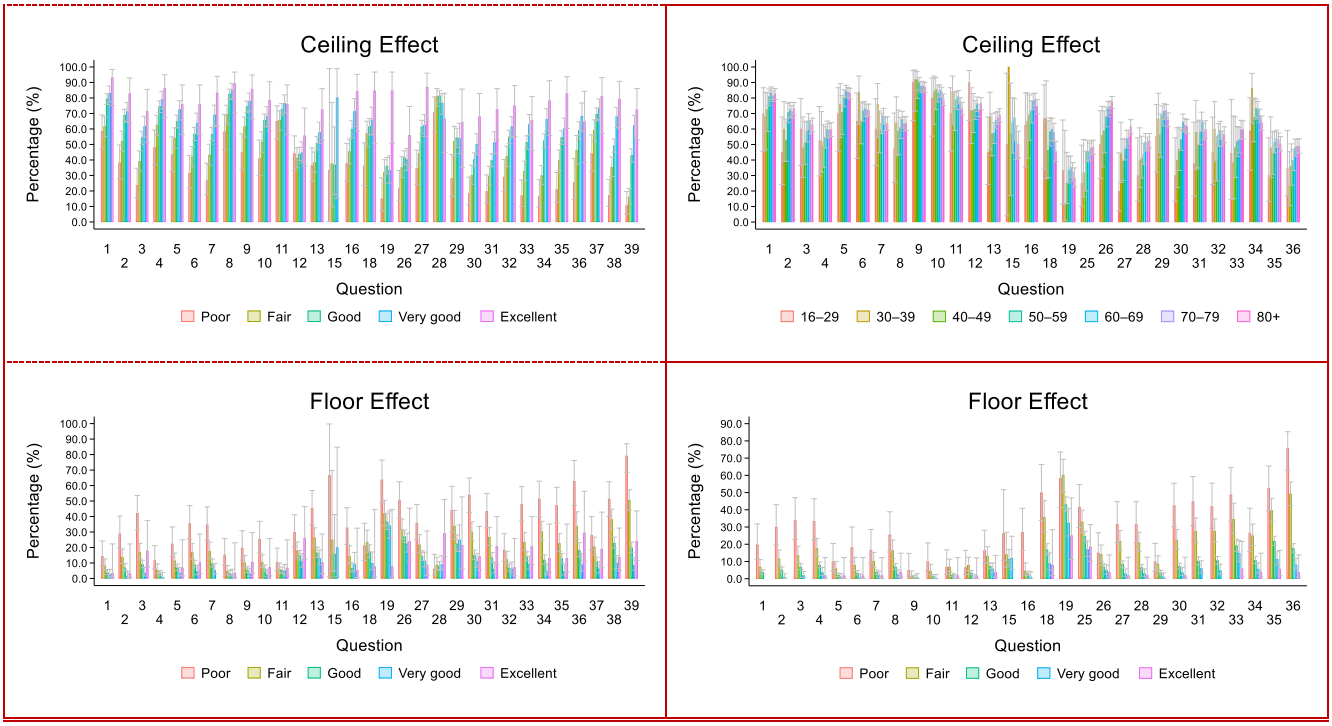


Figure 78 Self-reported health differences in data quality metrics observed in the pilot samples from Psychiatry and Rehabilitation.

Insurance Status

Apart from the consistent effect of insurance coverage (already noted in acute care (see Figure 37), where patients with semi-private or private insurance tend to rate their experience more favorably) Figure 79 shows little variation in data quality metrics.

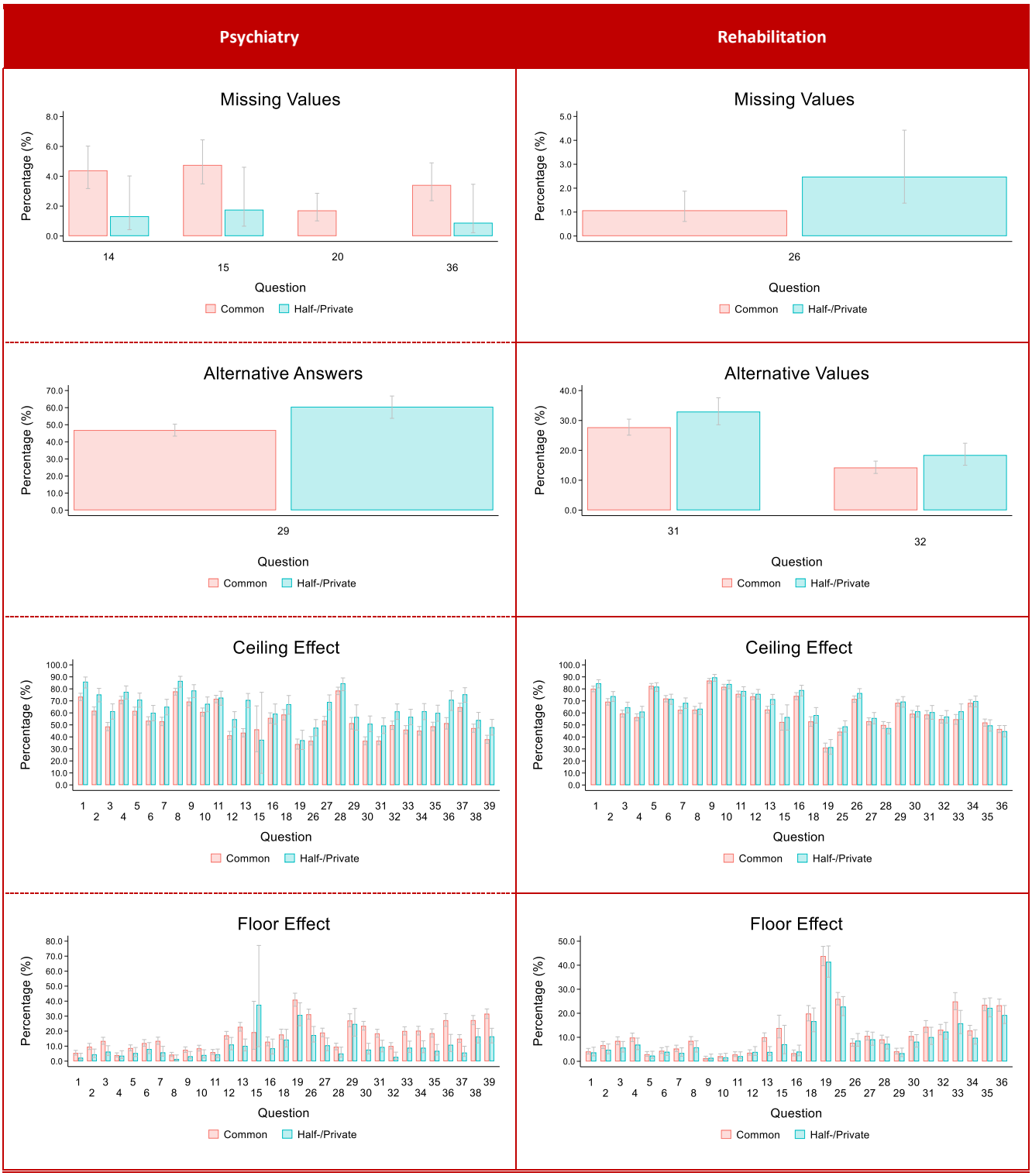
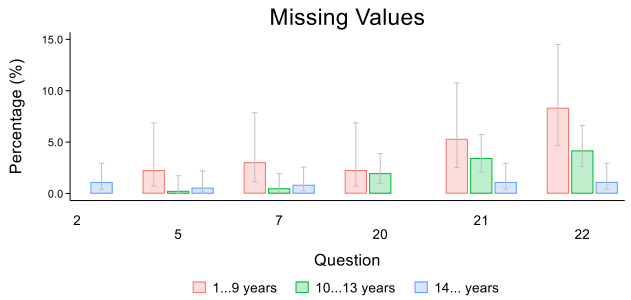
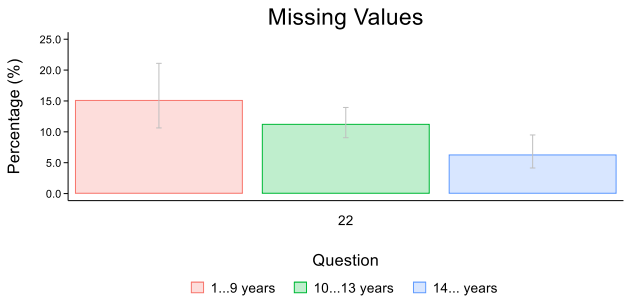
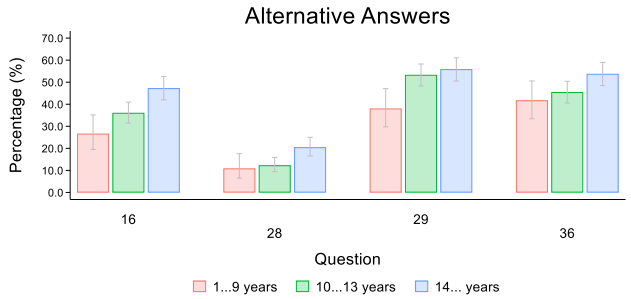
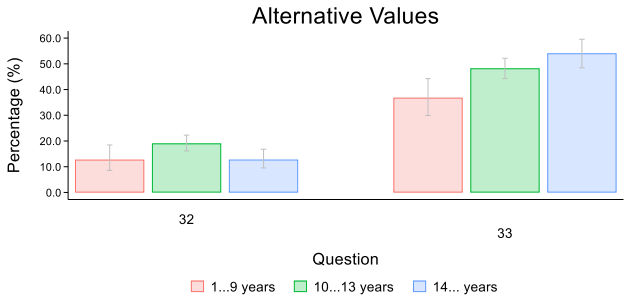
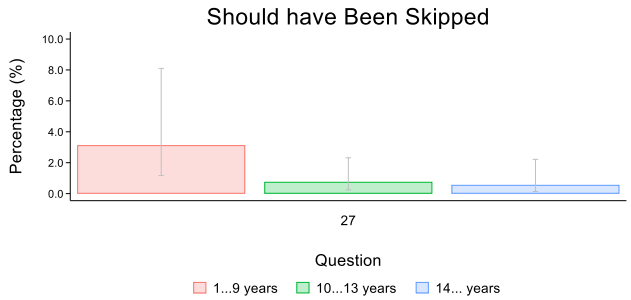
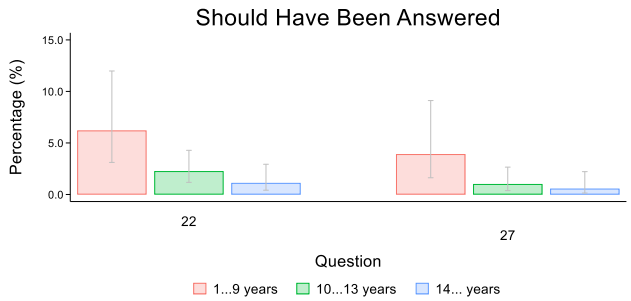
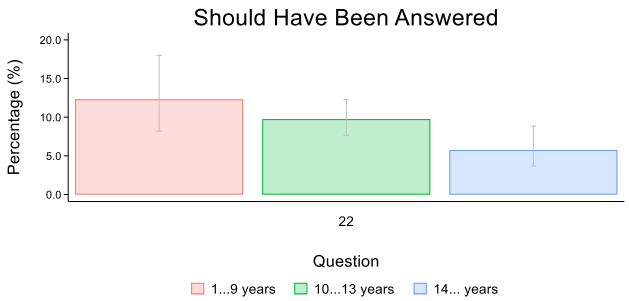


Figure 79 Insurance status-related differences in data quality metrics observed in the pilot samples from Psychiatry and Rehabilitation.

Education

The impact of education level on data quality is negligible, as shown in Figure 80. As previously observed in acute care (see Figure 44), respondents with fewer years of education may be less familiar with completing questionnaires and tend to exhibit slightly higher error rates following filter questions.

Psychiatry	Rehabilitation
	
	
	
	

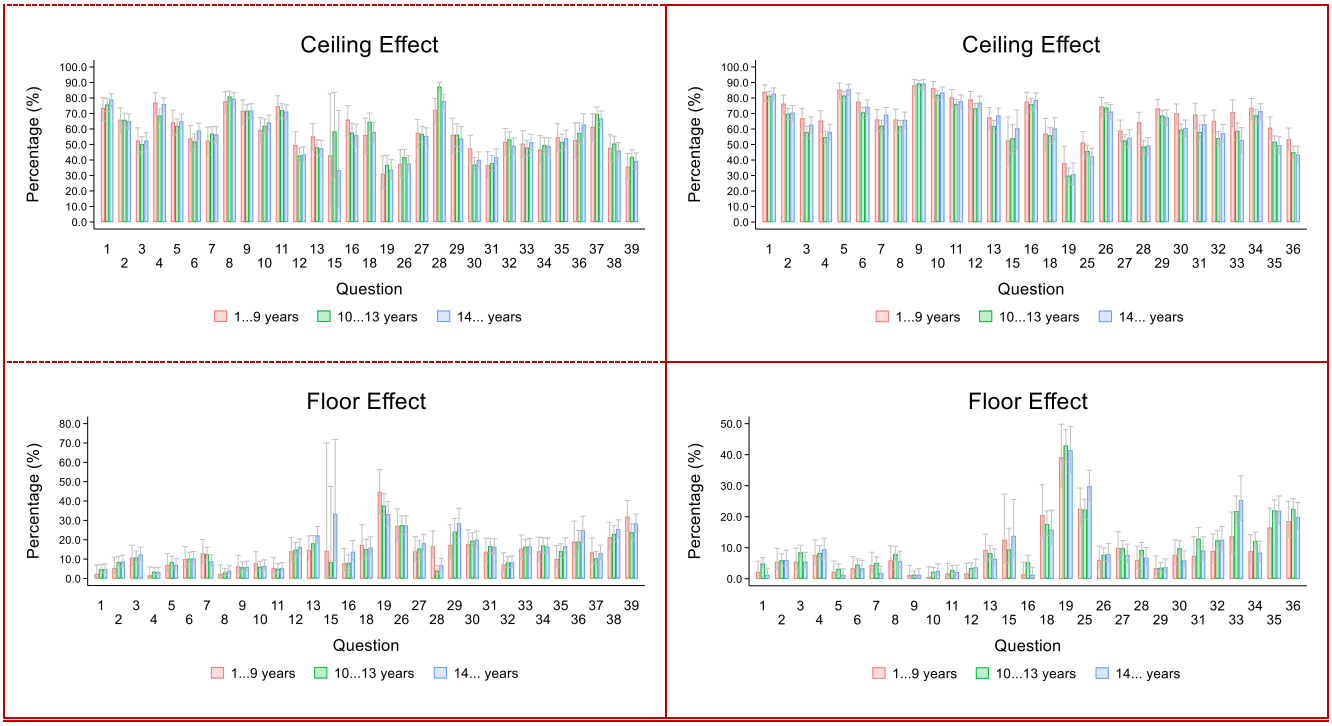
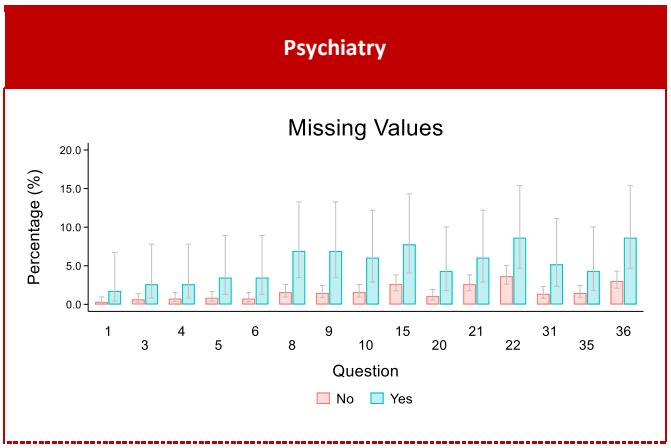


Figure 80 Education level-related differences in data quality metrics observed in the pilot samples from Psychiatry and Rehabilitation.

Placement

This variable is specific to the psychiatry questionnaire and asks respondents whether they were admitted to the clinic involuntarily. The patterns shown in Figure 81 confirm those previously observed in the PatZu measurements: patients who were admitted against their will tend to rate their experience less favorably and are more likely to leave questions unanswered.



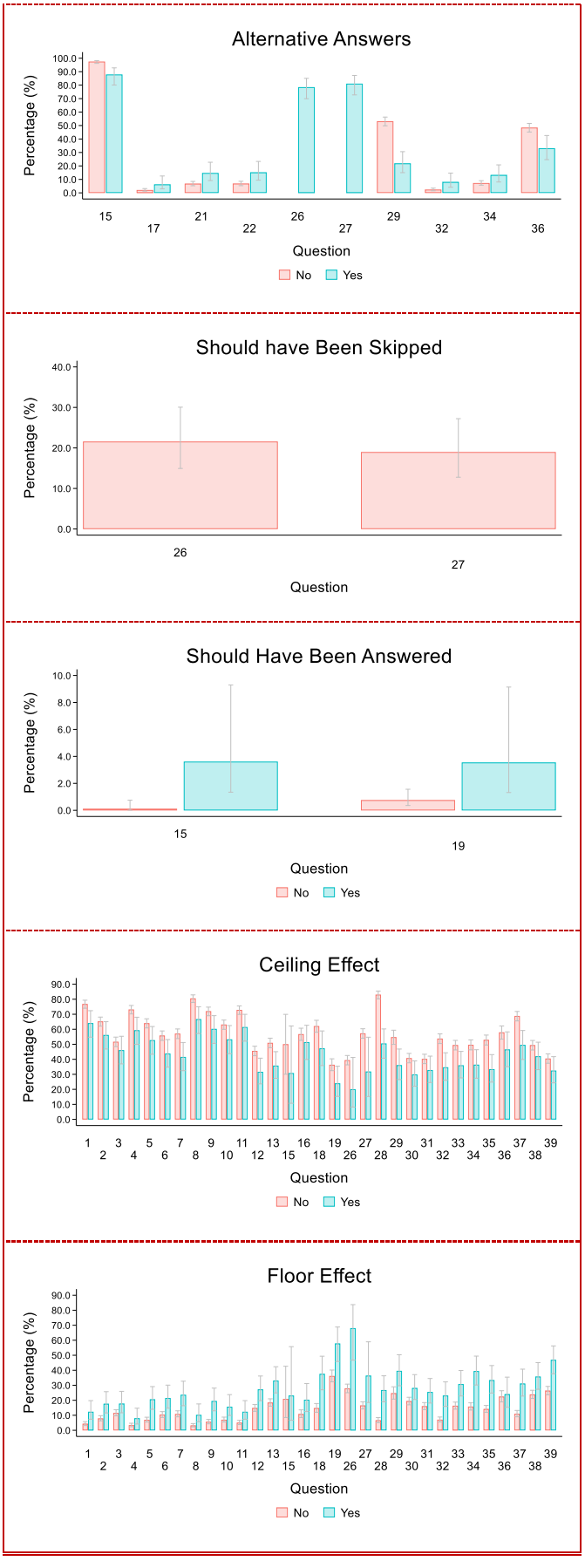


Figure 81 Placement-related differences in data quality metrics observed in the pilot sample from Psychiatry.

5.2.2.Questionnaire Shortening

Questionnaire Fatigue

Figure 82 confirms that the increase in missing values across the questionnaires in both psychiatry and rehabilitation is minimal, typically only one or two respondents by question. These missing values are less the result of respondents dropping out and more due to difficulties in understanding the filter question and subsequent questions to skip in the paper version (age plays also a role). The patterns observed in rehabilitation closely mirror those seen in acute care (see Figure 38). However, in psychiatry, the curve representing respondents completing the electronic form rises more sharply in the initial sections focused on psychiatrists and nursing staff, then levels off almost completely thereafter.

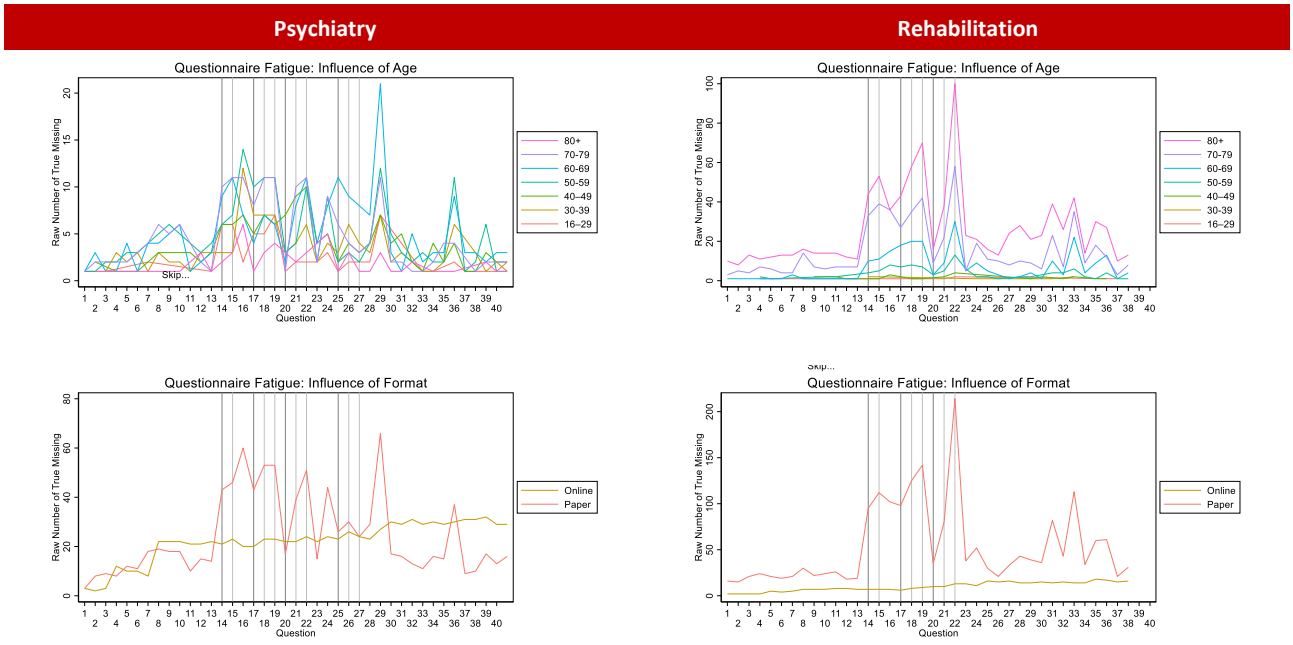


Figure 82 Influence of age and questionnaire format on missing filter questions in the acute care pilot study.

Correlation

Figure 83 illustrates the correlation structure among PREMs questions. In psychiatry, the three professional categories form a homogeneous block (top left), whereas in rehabilitation, inter-category correlations are weaker, with each category appearing to form its own distinct block. The section related to the clinic stay (bottom right) forms a separate block, which shows stronger correlations with the staff block in both studies.

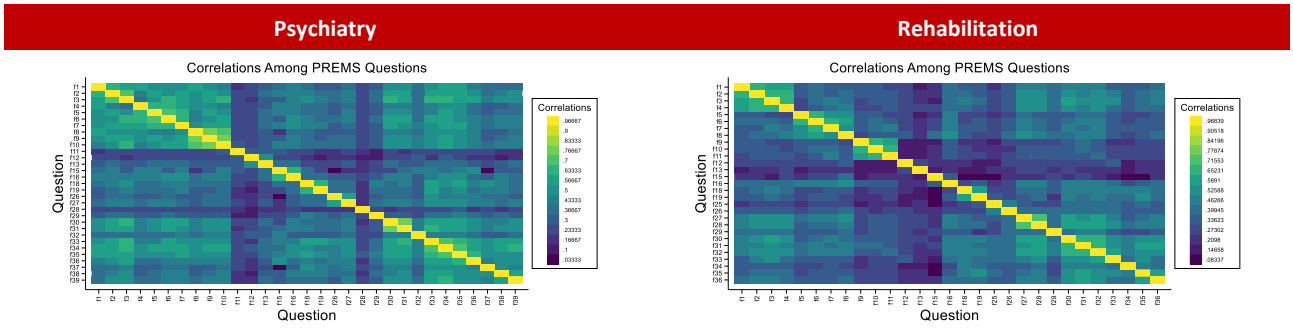


Figure 83 Heat plot of correlations between 4-point scale PREM questions in the psychiatry and rehabilitation pilot studies.

Global Satisfaction Questions

Figure 84 reveals patterns in psychiatry and rehabilitation that closely resemble those seen in the PCA plot for acute care (see Figure 43) and the item response theory information functions (see Figure 44). A similar rationale can be applied to divide the four global satisfaction questions into two pairs and retain only one question from each pair.

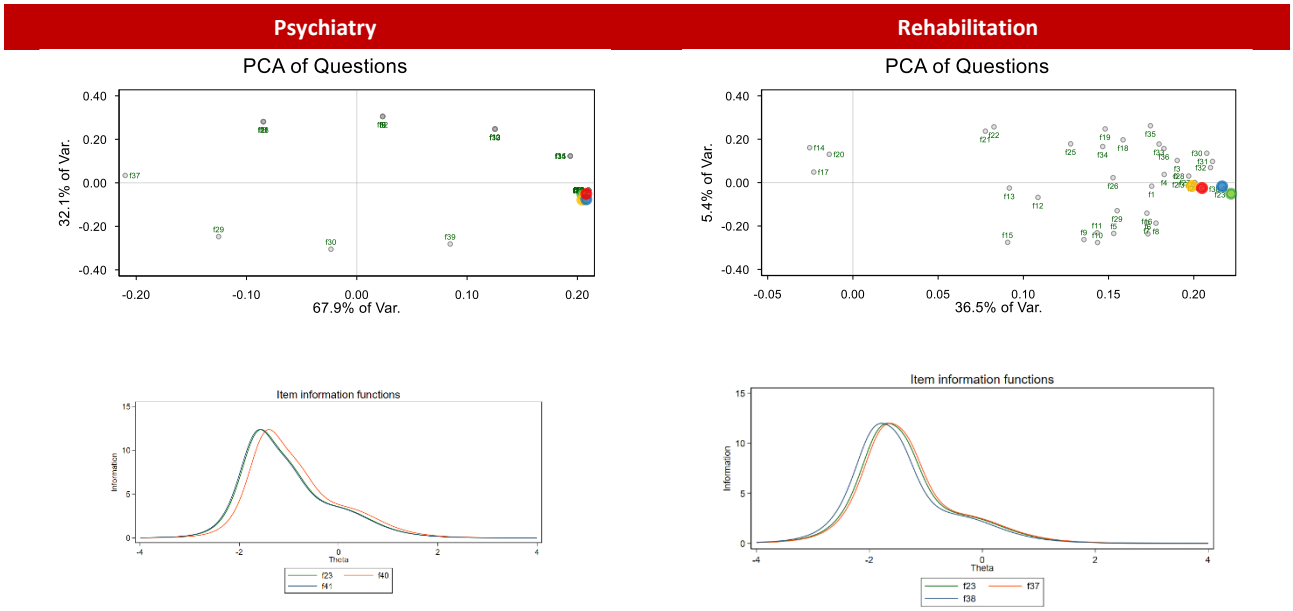


Figure 84 Position of global satisfaction questions in the principal component analysis (PCA) plane, in the psychiatry and rehabilitation pilot studies (top row). Item Information from IRT on bottom row.

Summary Tables

One notable observation in Table 42 is the exceptionally high rates of "does not apply" responses to the question concerning assistance with toileting. This is more pertinent in the acute care setting, which justified its exclusion from the final psychiatry and rehabilitation versions of the questionnaire.

Table 42 List of questions to considered for shortening in the psychiatry and rehabilitation questionnaires. Data quality is summarized by the rate of true missing values (i.e., no alternate answer selected), and potential scale saturation, as indicated by the ceiling effect, which may not necessarily be problematic in the context of extreme values boxes. The proportion of alternative responses shows the amount of non-missing escape responses.

		Psychiatry				Rehabilitation			
Item	Type	No.	Miss. Val.	Ceil. eff.	Altern. Ans.	No.	Miss. Val.	Ceil. eff.	Altern. Ans.
Courtesy and resp. (doc.)	PREMs	1	0.5%	74.7%		1	1.1%	81.1%	
Careful listening (doc.)	PREMs	2	0.9%	63.9%		2	1.0%	70.3%	
Underst. expl. (doc.)	PREMs	3	1.1%	50.8%		3	1.4%	61.0%	
Availability (doc.)	PREMs					4	1.6%	57.8%	
Courtesy and resp. (nurs.)	PREMs	4	1.8%	71.6%		5	1.6%	81.8%	
Careful listening (nurs.)	PREMs	5	2.0%	62.9%		6	1.4%	71.4%	
Understand. expl. (nurs.)	PREMs	6	1.9%	54.3%		7	1.6%	63.9%	
Availability (nurs.)	PREMs	7	2.3%	55.0%	3.8%	8	2.2%	62.7%	
Courtesy and resp. (ther.)	PREMs	8	3.6%	78.7%		9	1.8%	87.5%	6.6%
Careful listening (ther.)	PREMs	9	3.5%	70.7%		10	1.9%	81.9%	
Understand. expl. (ther.)	PREMs	10	3.5%	61.7%		11	2.1%	76.2%	
Cleanliness	PREMs	11	2.7%	71.6%		12	1.6%	74.3%	
Calm	PREMs	12	3.2%	44.3%		13	1.6%	65.1%	
Safe space	PREMs	13	3.2%	49.5%					
Toilets?	Filter	14	5.7%			14	6.2%		
Assistance for toileting	PREMs	15	6.1%	42.1%	96.4%	15	7.2%	53.3%	79.2%
Pain control	PREMs	16	7.1%	56.3%	36.7%	16	6.6%	74.8%	19.7%
New medicine?	Filter	17	5.6%		2.6%	17	6.3%		7.8%
Purpose of medication	PREMs	18	6.7%	60.4%	37.8%	18	8.1%	54.4%	43.7%
Side effects	PREMs	19	6.7%	34.7%	37.8%	19	9.2%	31.6%	44.2%
Destination?	Filter	20	3.5%			20	2.7%		
Need for help	PREMs	21	5.4%		7.5%	21	5.5%		6.6%
Symptom surveillance	PREMs	22	6.6%		7.6%	22	13.8%		7.2%
Rating of hospital stay	Satis.	23	3.3%			23	3.1%		
Recommendation?	Satis.	24	6.0%			24	3.8%		
Placement	Filter	25	4.3%						
Information on admission	PREMs	26	5.0%	38.8%	8.5%	25	2.8%	45.6%	
Admission organization	PREMs	27	4.3%	56.7%	8.7%	26	2.2%	72.7%	
Restriction frequency	PREMs	28	4.6%	79.0%	15.3%				
Expl. about restriction	PREMs	20	8.2%	51.8%	49.3%				
Comm. between staff	PREMs	30	4.2%	39.6%		27	3.0%	53.8%	
Information on treatment	PREMs	31	4.0%	39.4%		28	3.5%	49.3%	
Punctuality	PREMs	32	3.9%	52.1%	2.9%	29	3.2%	68.5%	1.8%

Inform. on health cond.	PREMs	33	3.5%	48.1%		30	3.1%	59.9%	
Worries	PREMs	34	4.1%	48.3%	7.9%	31	5.8%	59.1%	28.7%
Decisions (pat.)	PREMs	35	3.9%	50.9%	7.3%	32	3.5%	55.2%	15.3%
Decisions (fam.)	PREMs	36	5.9%	56.2%	46.5%	33	7.7%	56.0%	44.8%
Understanding medication	PREMs	37	3.5%	66.8%	11.8%	34	2.9%	68.4%	8.9%
Continuing treatment	PREMs	38	3.6%	48.6%		35	4.7%	51.0%	
Understanding condition	PREMs	30	4.3%	39.6%		36	4.7%	46.2%	
Helpfulness	Satis.	40	3.7%			37	2.2%		
Rating of experience	Satis.	41	4.0%			38	2.8%		

5.2.3.PREMs Score Construction

The construction of PREMs themes was initially based on the acute care pilot sample. The same items were used to generate scores in the psychiatry and rehabilitation domains. Additional domain-specific items were included where appropriate: questions for therapists were added to relevant themes in both psychiatry and rehabilitation. The question about safe spaces was incorporated into the *Environment* theme in psychiatry, while questions regarding freedom of movement were considered too distinct and therefore excluded, sett Table 15. Analyses were then replicated in these two domains to ensure consistency of results and to confirm the validity of the construct across settings.

Score Structure

Top- and bottom-box structure in psychiatry (Figure 85) and rehabilitation (Figure 86) are similar to the acute care case (Figure 45).

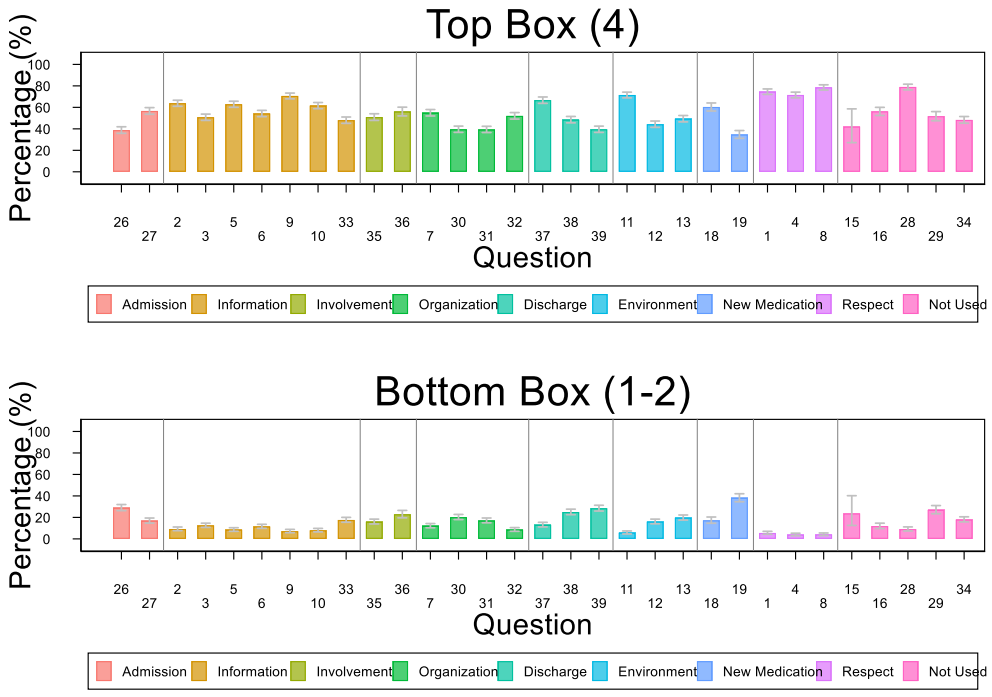


Figure 85 Distribution of responses at the extreme ends (top and bottom boxes) for each item within thematic domains in the psychiatry pilot study.

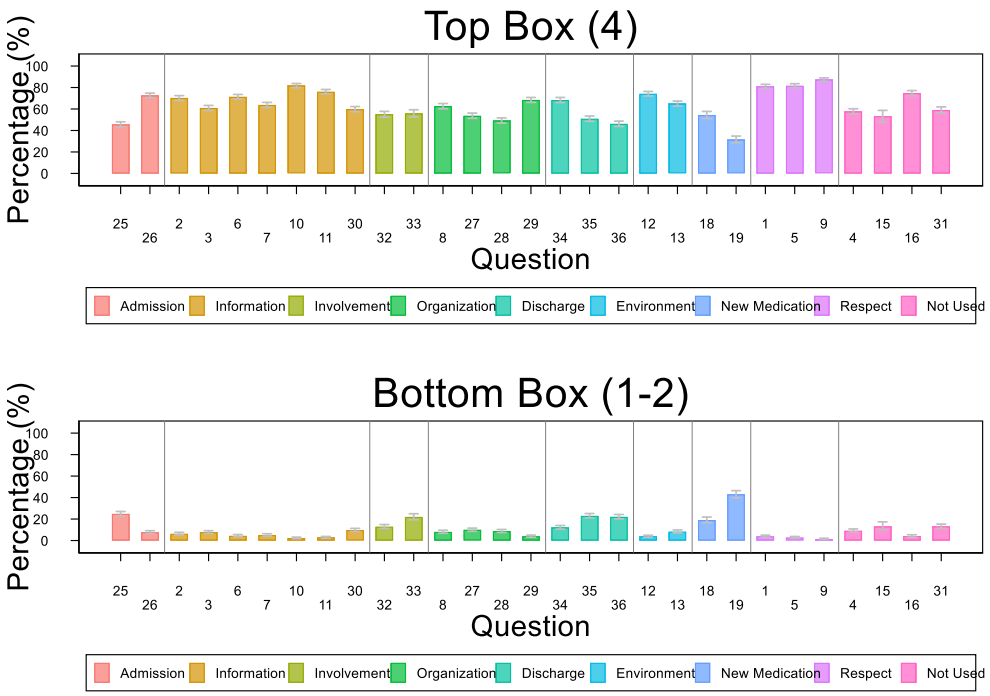


Figure 86 Distribution of responses at the extreme ends (top and bottom boxes) for each item within thematic domains in the rehabilitation pilot study.

Similarity Among Questions

Replication of principal component analyses (PCA) yields similar results, with a central cluster comprising the two most populated themes, *Information* and *Organization*, along with the less populated *Involvement* theme, remaining relatively stable.

However, some divergences emerge in the psychiatry study: there is a greater overlap of questions (Figure 88), and the *Environment* cluster shifts dramatically from the far-left periphery to the more crowded right side of the plane. Question 13, concerning safe spaces, was added to the psychiatry questionnaire and lies further apart from questions 11 and 12, which address cleanliness and calm, respectively. The same applies to questions 28 and 29 concerning restriction of freedom of movement: they clearly stand apart, thus providing additional information.

In contrast, acute care and rehabilitation show more similarities (Figure 87), but questions 9 to 11 (targeting the therapist group) form a quasi-distinct cluster at the bottom of the graph.

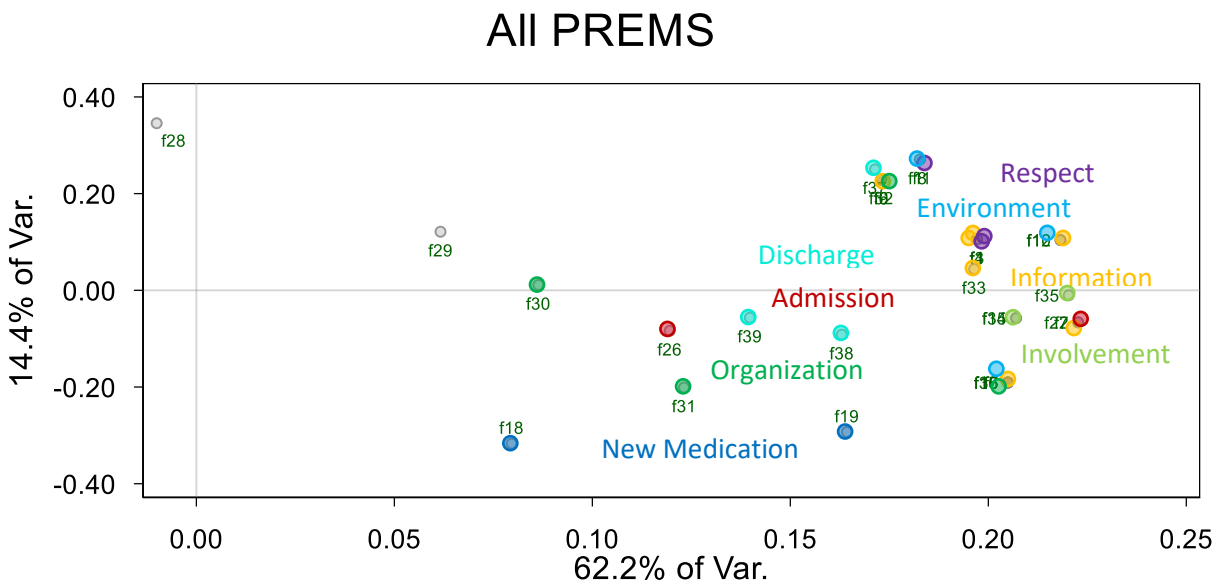


Figure 87 Distribution of PREMs Items across the PCA plane in the psychiatry pilot study.

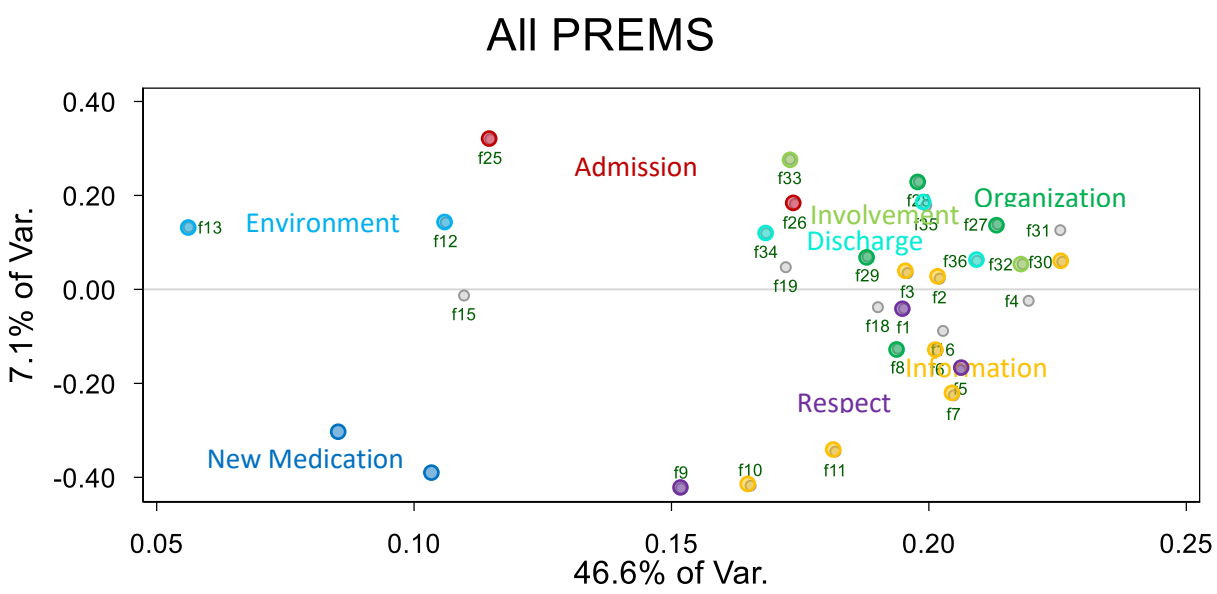
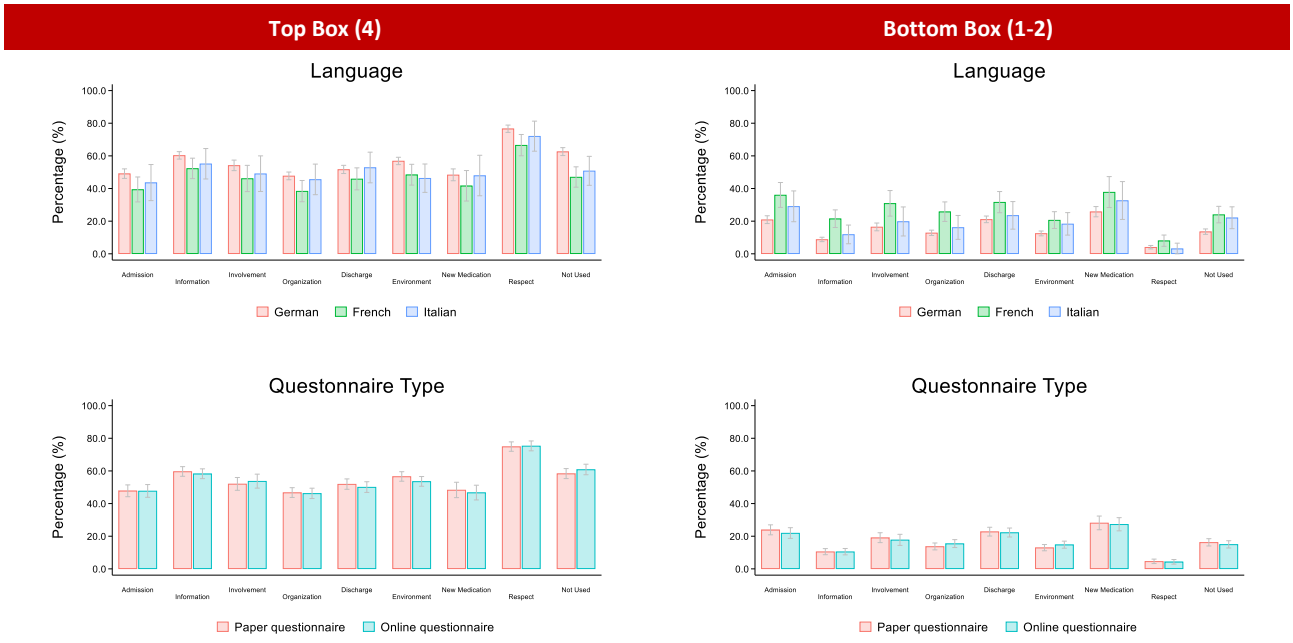


Figure 88 Principal component analysis (PCA) of correlations among 4-item PREMs questions in the rehabilitation pilot sample. Colors represent thematic attributions.

Influence of Data Collection

Results similar to those in acute care were observed in the in psychiatry (Figure 89) and rehabilitation (Figure 90) pilot studies, confirming the robustness of these PREMs scores. The same tendency for French-speaking respondents to rate their experience lower is evident in the psychiatry sample, while in the rehabilitation sample, the Italian-speaking subgroup tends to give higher ratings. Most importantly, the questionnaire format does not appear to influence PREMs scores.



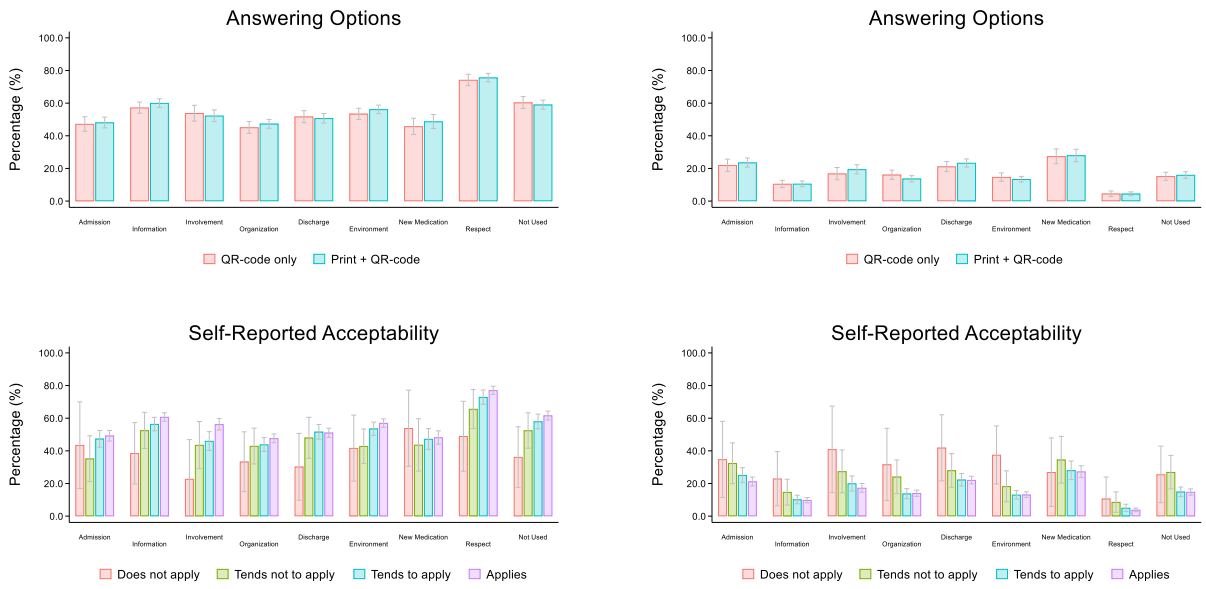


Figure 89 Differences in national averages of top and bottom box proportions of the eight thematic categories according to different questionnaire versions and data collection processes, as measured on the psychiatry pilot sample.

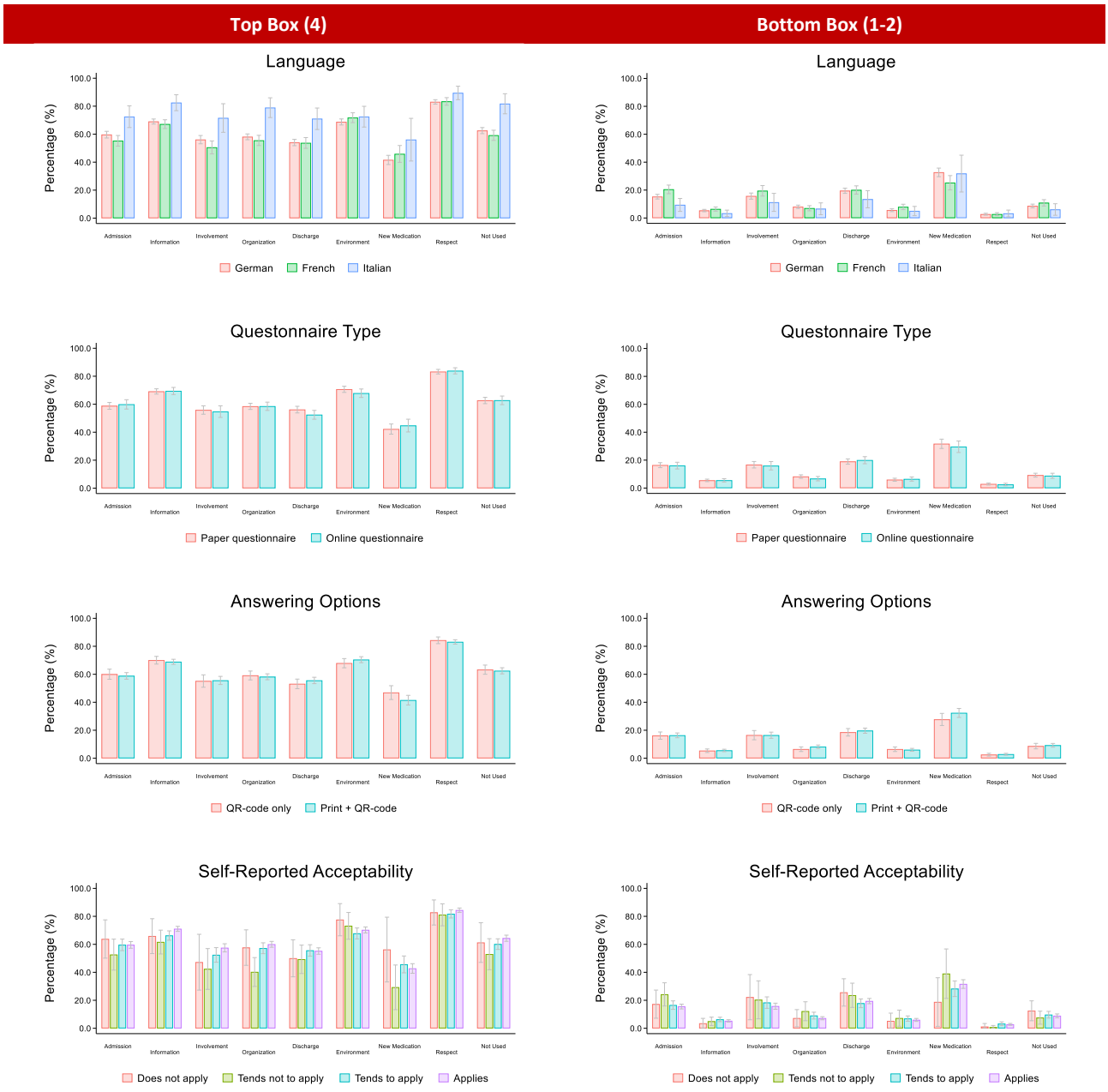


Figure 90 Differences in national averages of top and bottom box proportions of the eight thematic categories according to different questionnaire versions and data collection processes, as measured on the rehabilitation pilot sample.

Influence of Socio-Demographic Factors

Variations across socio-demographic groups are less pronounced in psychiatry (Figure 89) and rehabilitation (Figure 92) compared to those observed in acute care (Figure 68). Self-reported health and insurance status do not exhibit the consistent patterns seen in acute care.

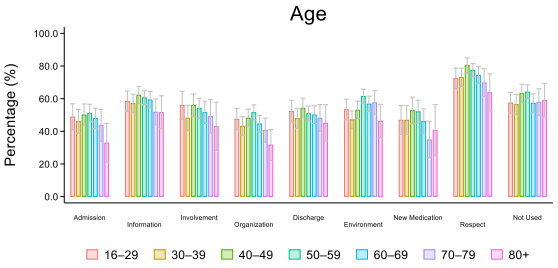
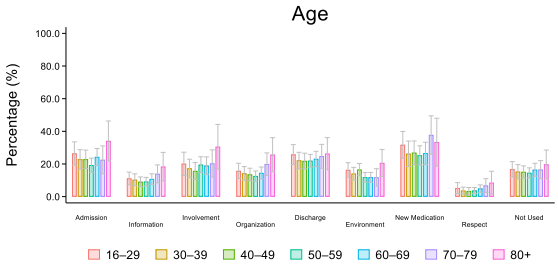
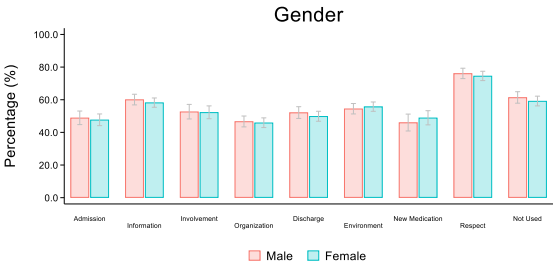
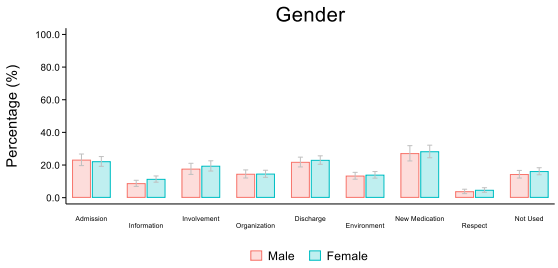
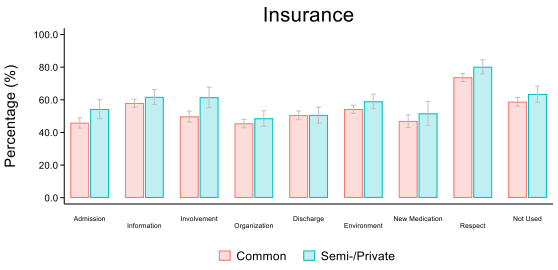
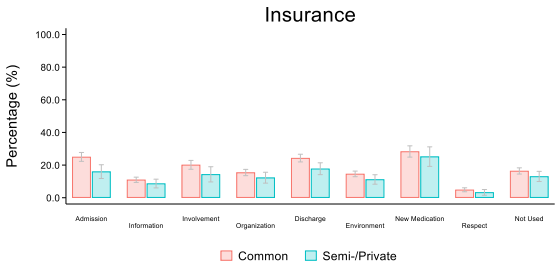
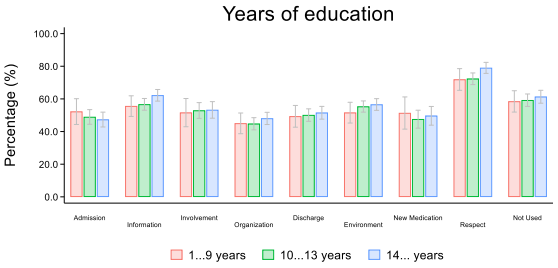
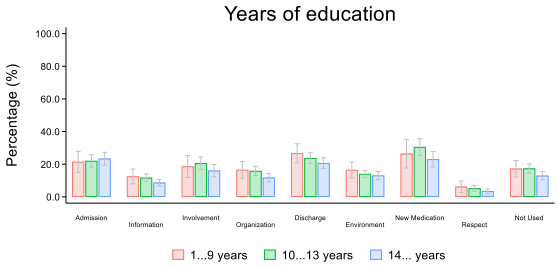
Top Box (4)	Bottom Box (1-2)
	
	
	
	
	

Figure 91 Differences in national averages of top and bottom box proportions of the eight thematic categories according to different socio-demographic factors, as measured on the psychiatry pilot sample.

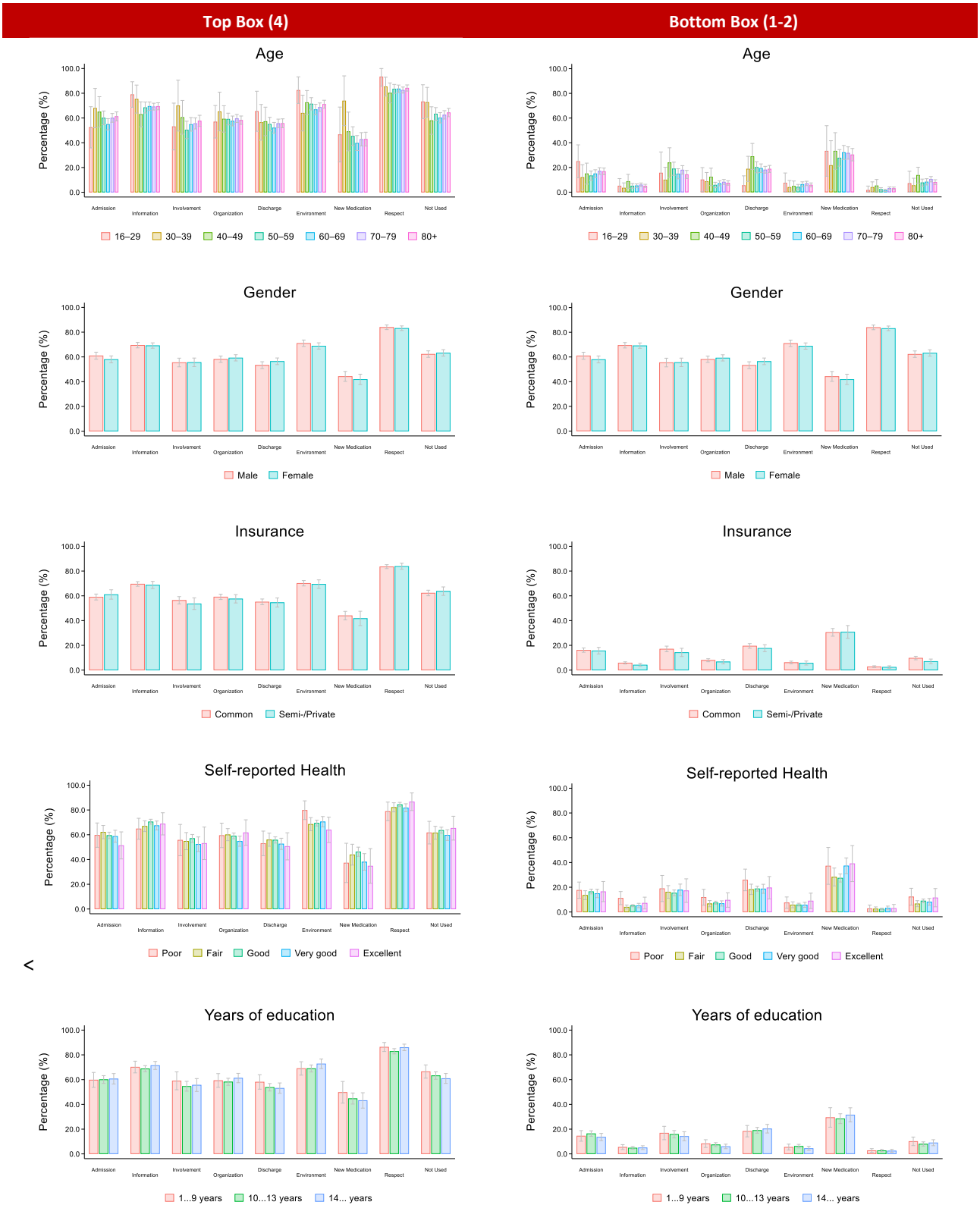


Figure 92 Differences in national averages of top and bottom box proportions of the eight thematic categories according to different socio-demographic factors, as measured on the rehabilitation pilot sample.

5.3. Adjusted PREMs Scores

5.3.1. Differences in Case-Mix

First, do we observe any differences between hospitals in terms of socio-demographic variables, and how do these relate to hospital categories? While patient profiles appear to vary across hospitals—except for self-reported health and education—Table 43 indicates that differences between hospital types occur less frequently. Only the placement variable remains highly significant, with age showing a slight effect.

Table 43 Measures of association between socio-demographic factors and hospitals or hospital types[f] as defined by the Swiss Statistical Federal Office (SFO or BfS in German) categorization.

		Hospital			BfS Category		
Factor	Test	Statistic	Parameter	P-Value	Statistic	Parameter	P-Value
Age	Kruskal-Wallis	101.0	36	0.0%	11.2	4	2.4%
Gender	Chi-square	63.5	36	0.3%	5.5	4	23.6%
Insurance Status	Chi-square	175.5	36	0.0%	7.5	4	11.0%
Self-reported Health	Chi-square	167.8	144	8.5%	14.2	16	58.4%
Education	Chi-square	138.0	108	2.7%	17.0	12	14.9%
Placement	Chi-square	141.3	36	0.0%	30.1	4	0.0%

In the following sections, we take a closer look at the distribution of socio-demographic data across hospitals of different types.

[f] When benchmarking hospitals at the national level, comparisons are often made within the same SFO category, where a certain degree of homogeneity is expected. In this analysis, we also examine inter-category variability in relation to the socio-demographic factors available.

Age

Median age differs significantly from hospital to hospital, as shown in Figure 93.

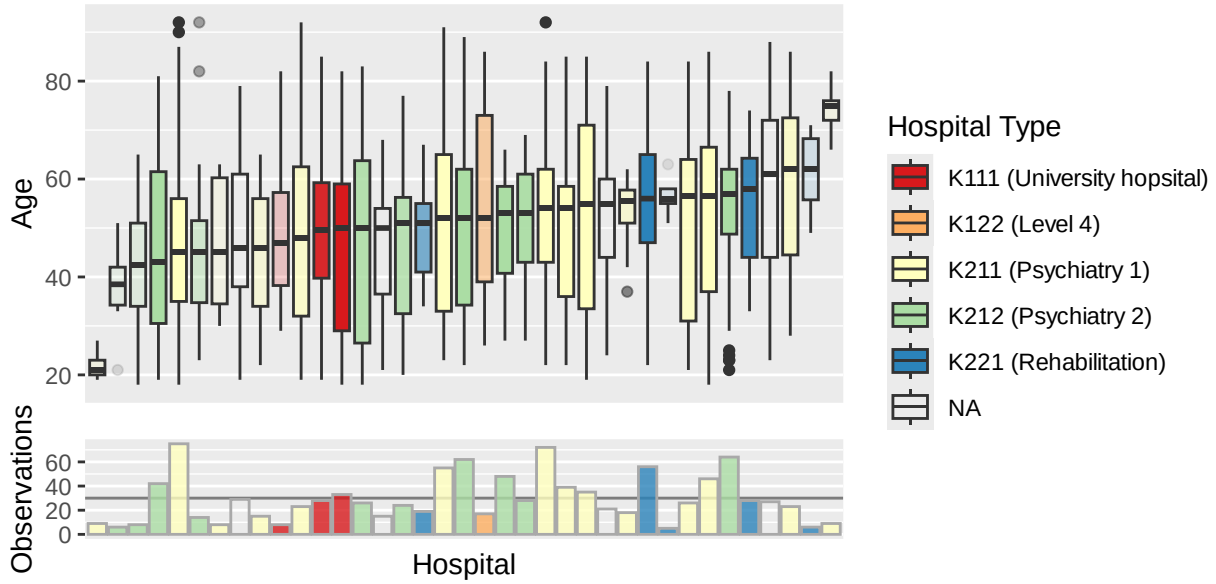


Figure 93 Median age and sample size in each hospital of the pilot sample in psychiatry. Color indicates BfS hospital category.

The very weak grouping of hospital categories, indicated by the 2.4% association level, is not clearly visible in Figure 94 either.

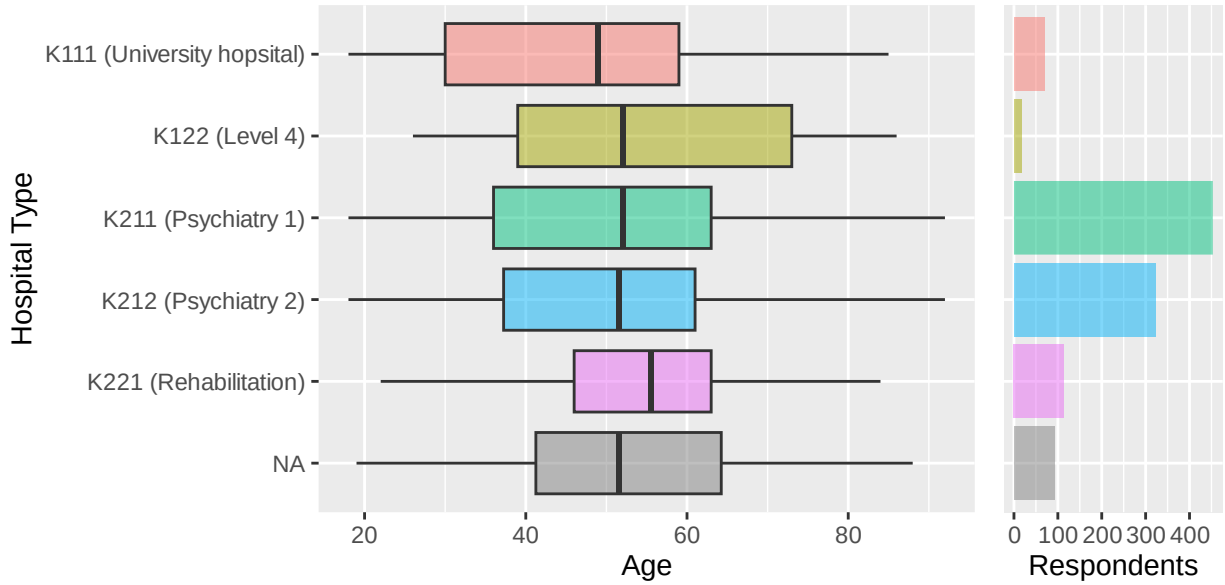


Figure 94 Age distribution across BfS hospital categories.

Overall, age does not appear to be an effective discriminator of hospital type: hospitals of all categories are dispersed across the full age spectrum, with no distinct clustering evident when ordered by patients' age—except for the two university hospitals, which are grouped near the middle.

Gender

How do hospitals differ in terms of the proportion of male and female patients? While differences are observable at the individual hospital level (see Figure 95) they disappear entirely when comparing hospital types (Figure 96).

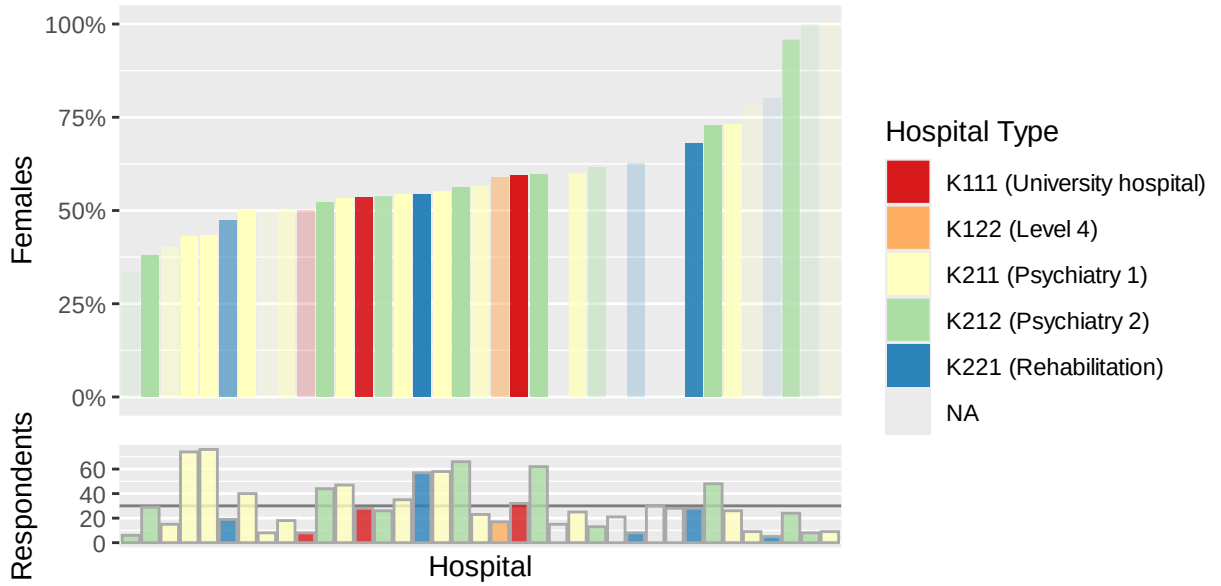


Figure 95 Differences in gender distribution (proportion of female patients) across hospitals. Color indicates BfS hospital category.

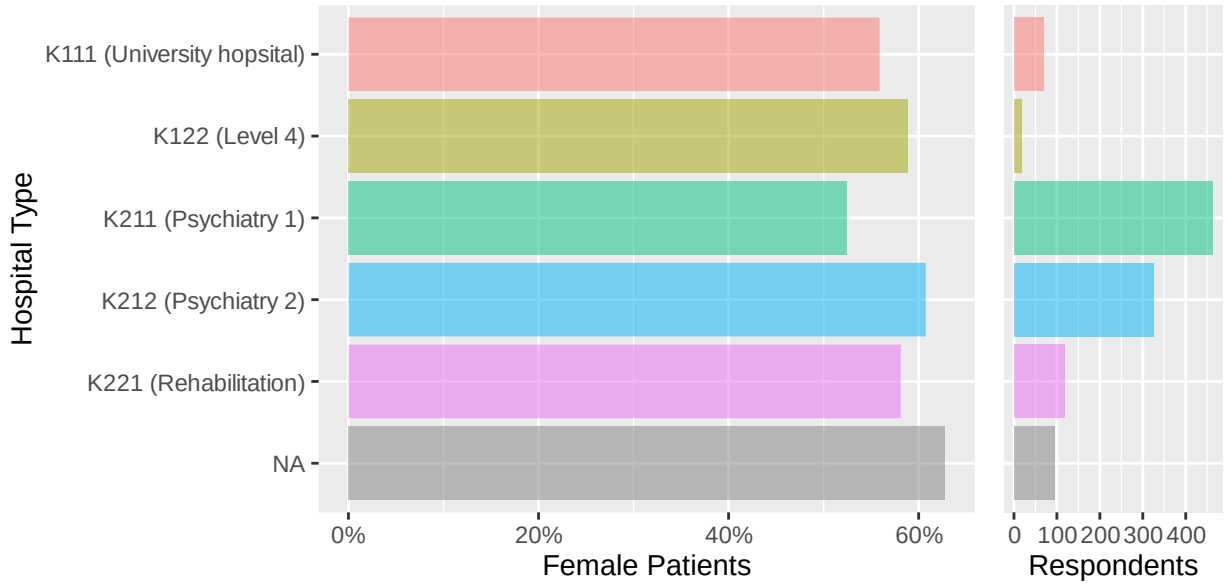


Figure 96 Gender distribution across BfS hospital categories.

Overall, the male-to-female ratio does not appear to effectively distinguish between hospital types: no clear clustering emerges when hospitals are ordered by the proportion of female patients. All hospital categories are distributed across the full spectrum, apart from three hospitals of unknown type.

Insurance Status

How do hospitals differ in terms of proportion of patients with public versus private insurance? While some clinics treat almost exclusively patients covered by private insurance (Figure 97), these variations disappear when hospital categories are compared (Figure 98).

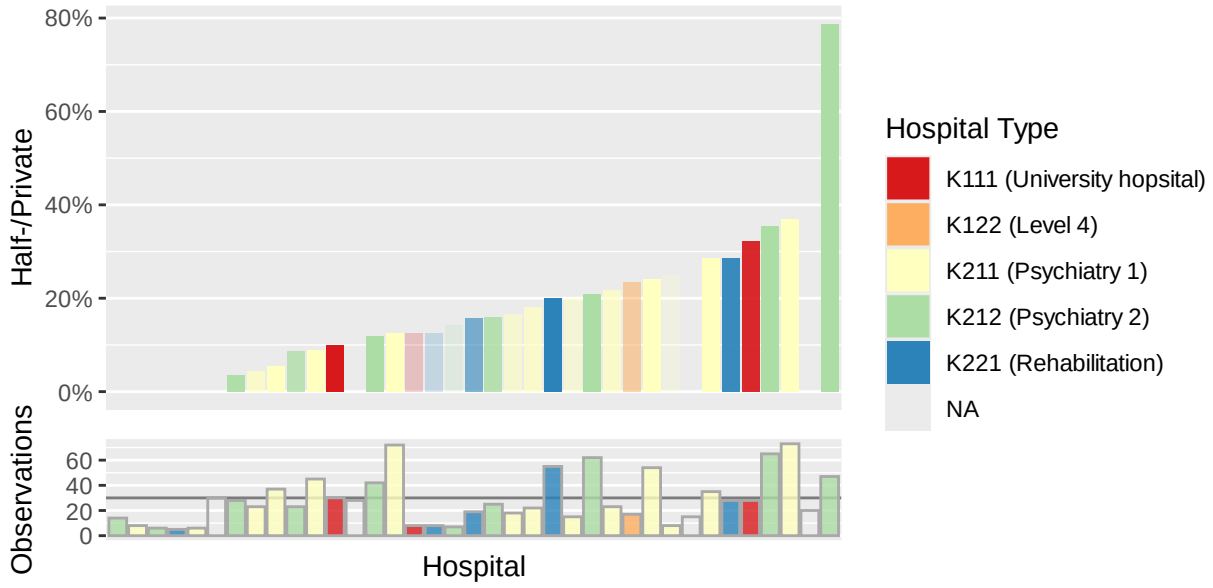


Figure 97 Differences in insurance status (proportion of patients with a semi- or private insurance) across hospitals. Color indicates BfS hospital category.

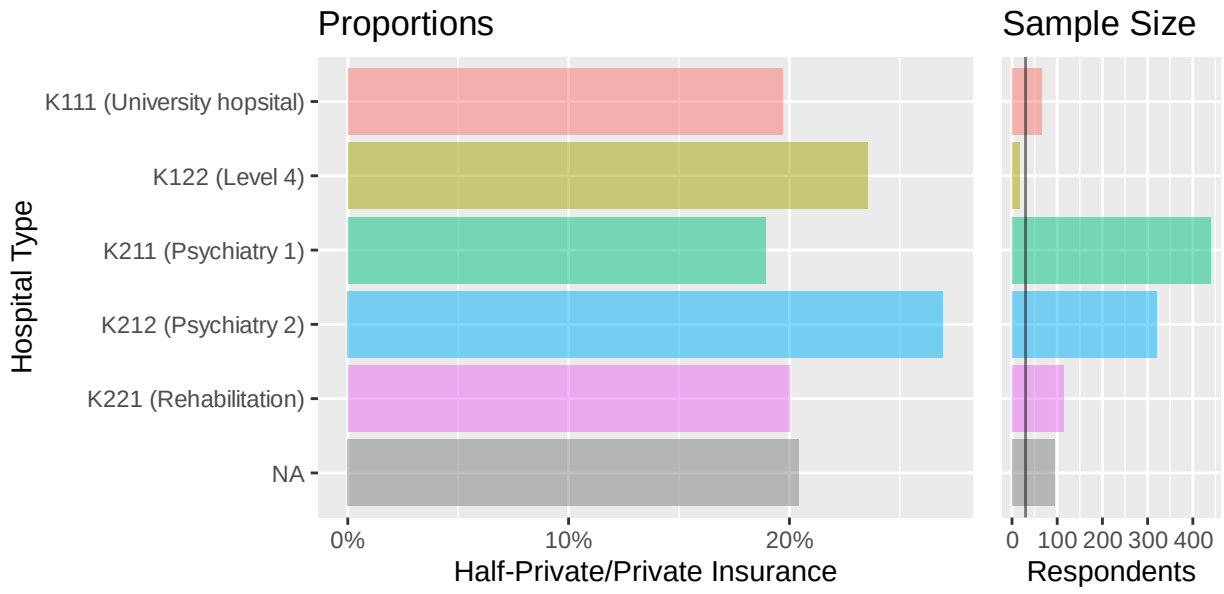


Figure 98 Insurance status distribution across BfS hospital categories.

Once again, the ratio of private to common insurance coverage does not appear to effectively differentiate hospital types. No distinct clustering is observed when hospitals are ordered by this proportion; all hospital categories, including university hospitals, are spread across the full spectrum.

Self-Reported Health

How do hospitals differ in terms of their patients' self-reported physical health? Although this variable is closely linked to PREMs outcomes and thematic scores (see Figure 90 in Section 5.2.3) all hospitals exhibit a similar distribution, both individually (Figure 99), and across hospital categories (Figure 100).

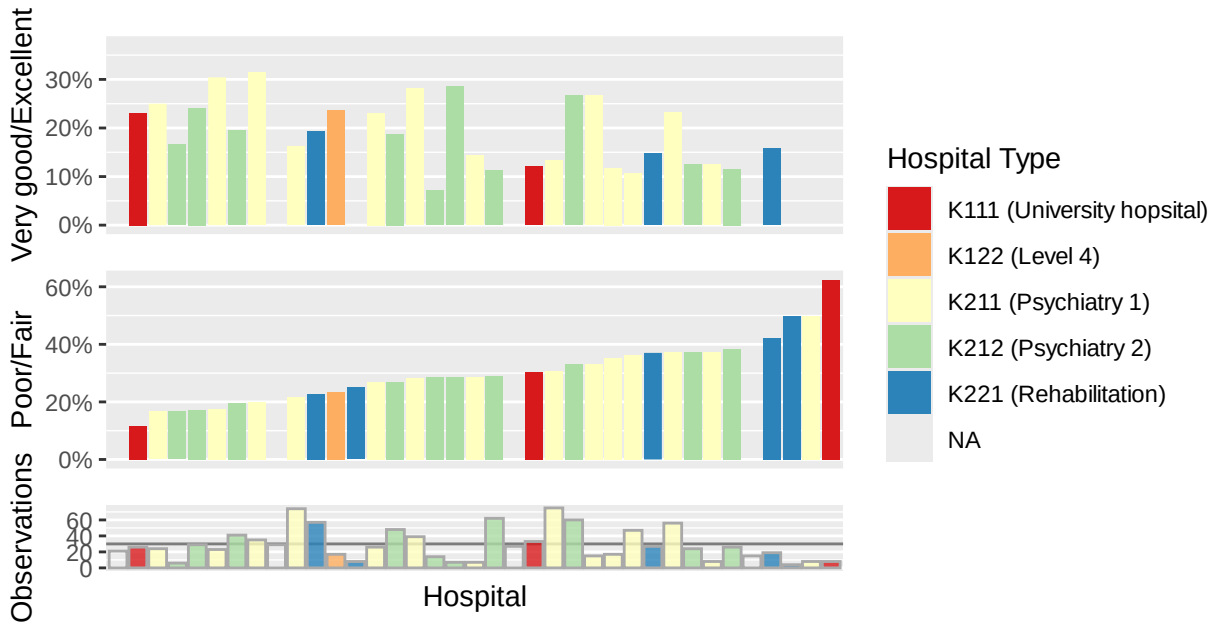


Figure 99 Differences in health condition of patients across hospitals. Proportion of patients declaring a better (top) or worse (bottom) condition. Color indicates BfS hospital category.

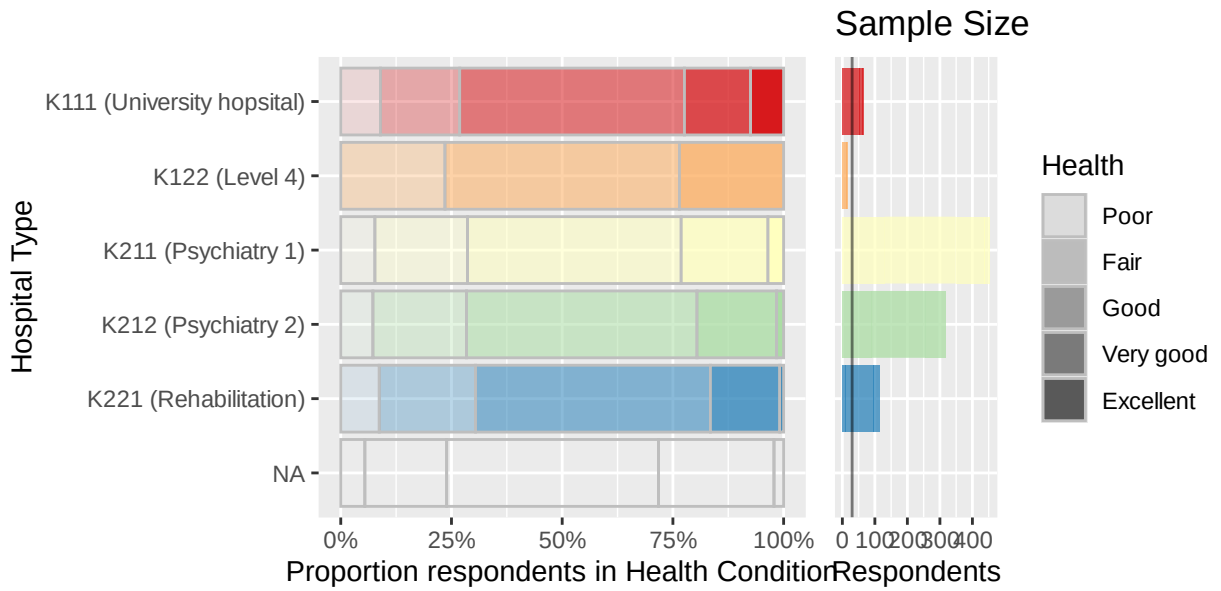


Figure 100 Distribution of patients' self-reported health condition across BfS hospital categories.

No distinct clustering is observed when hospitals are ordered by health condition of their patients; all hospital categories, including university hospitals, are spread across the full spectrum.

Education Level

How do hospitals differ in terms of their patients' education level? Like insurance status, education level (used as a proxy for socio-economic status) is expected to vary slightly between individual hospitals (Figure 101). However, no differences emerge when comparing hospital types (Figure 102).

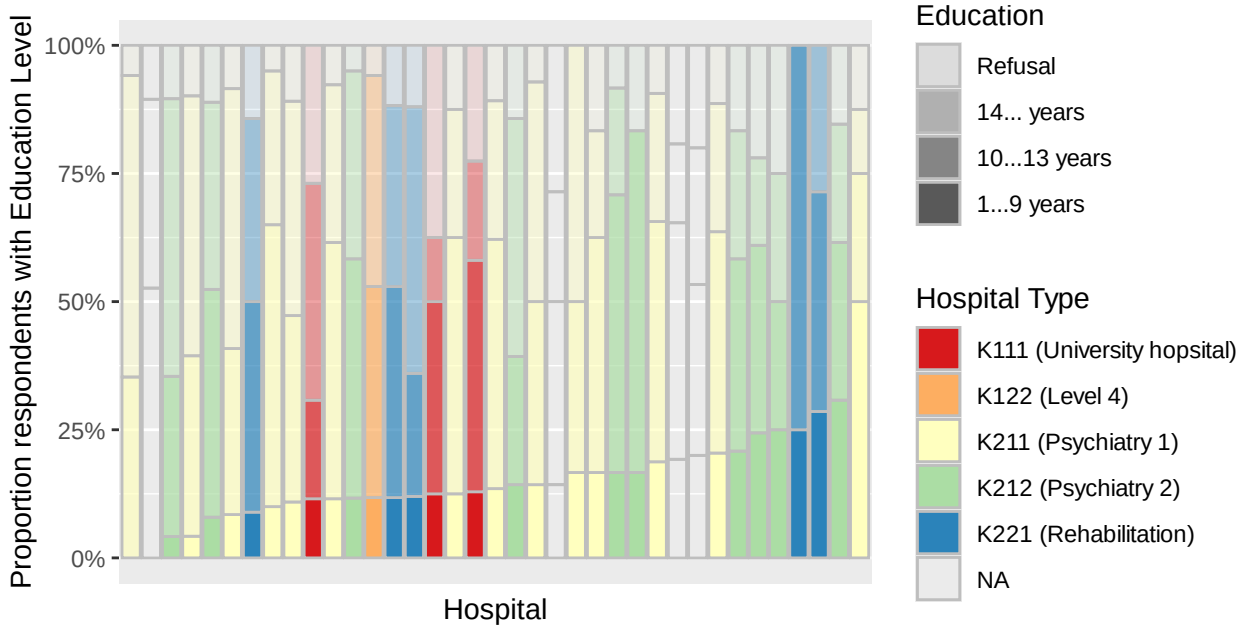


Figure 101 Differences in education level of patients across hospitals. Color indicates BfS hospital category.

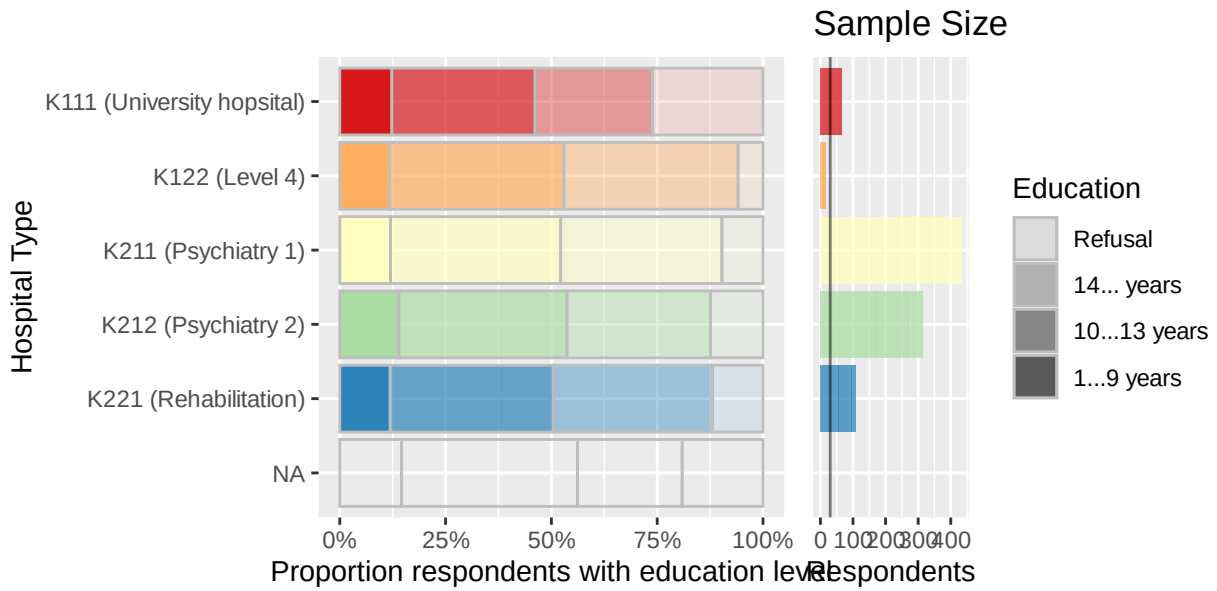


Figure 102 Distribution of patients' education level across BfS hospital categories.

Education level does not appear to significantly influence overall thematic scores (see Figure 68) nor does it effectively characterize hospital case mixes. An additional limitation is the variable's high rate of missing data.

Placement

Patient admission methods in psychiatric clinics have been used in PatZu's national measurements for stratifying metrics since their inception. How does the proportion of involuntary admissions vary between hospitals? As shown in Figure 103, this proportion ranges from zero to nearly half of patients in one university hospital (note the small sample size). A similar gradient is evident across hospital categories, as illustrated in Figure 104.

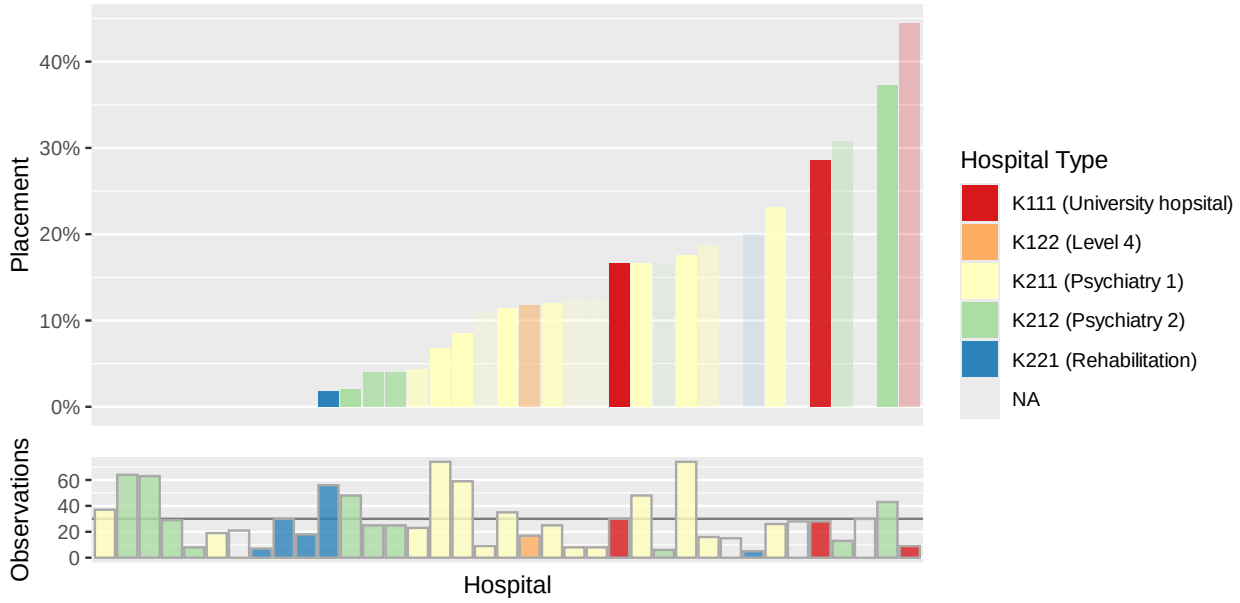


Figure 103 Differences in the proportion of patients admitted against their will across hospitals. Color indicates BfS hospital category.

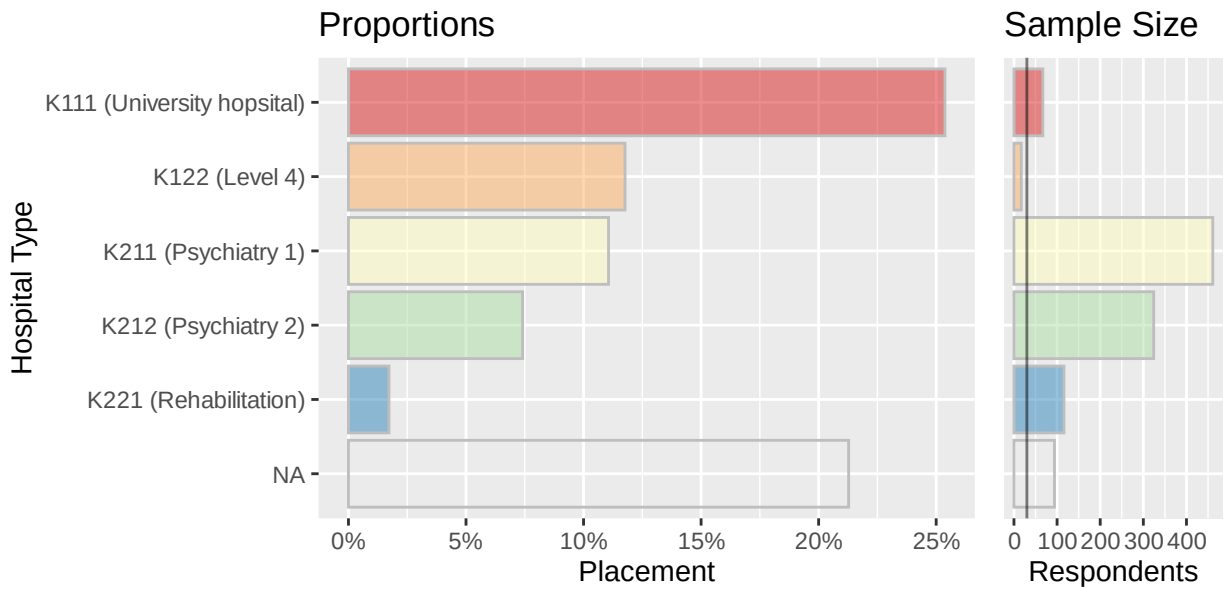


Figure 104 Distribution of the proportion of patients admitted against their will across BfS hospital categories.

Figure 103 and Figure 104 confirm that the mode of patient admission is a meaningful factor for distinguishing patient populations across hospitals. In addition to its impact on patients' experiences during their clinic stay (see Figure 89 in Section 5.2.3), this variable is a strong candidate for adjusting or stratifying results.

Multivariate Description

Except for patient admission, univariate plots of individual socio-demographic variables did not reveal a distinct pattern across hospital types. Does this picture change when all socio-demographic variables are considered simultaneously—for example, by projecting hospitals onto the principal component plane (PCA) derived from these variables?

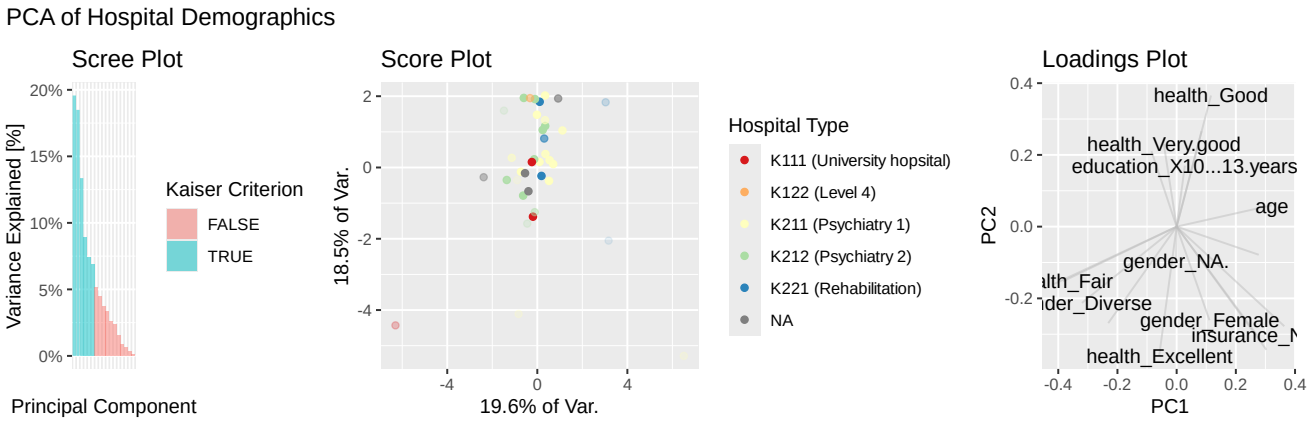


Figure 105 Projection of the six socio-demographic factors onto their principal components analysis (PCA) plane, at the hospital level.

The message conveyed by the PCA analysis reinforces what was already suggested by the univariate results: there is no clear clustering of hospitals by type, as hospitals of all categories are randomly dispersed across the principal component plane.

5.3.2. Influence of Factors on PREMs Scores

We begin by examining whether any bivariate associations emerge between score–factor pairs. Such associating where not clearly visible on the stratified plots in Figure 91.

To support visual interpretation, potential associations between categorical socio-demographic factors and continuous scores (ranging from 0% to 100%) are assessed using the Kruskal–Wallis statistic across all respondents.

Note that the number of tests corresponds to the total number of combinations:

#(all factor levels) * #(themes) * #(boxes).

Because these tests are repeatedly applied to the same data, the issue of appropriate significance levels arises. However, in this context, the tests are used purely descriptively to identify recurring patterns in the data.

Bivariate Influence

At Patient Level

Analyzing the entire population and performing association tests does not reveal a coherent pattern in Figure 106. Almost no factor-score pair shows a significant association, and when one does, it occurs only once. The only pair confirmed in both top and bottom boxes is the association between education level and the *New Medication* theme. However, as previously mentioned, the 5% threshold is overly permissive given the large number of tests conducted. Therefore, we conclude that no meaningful association is present in the overall sample.

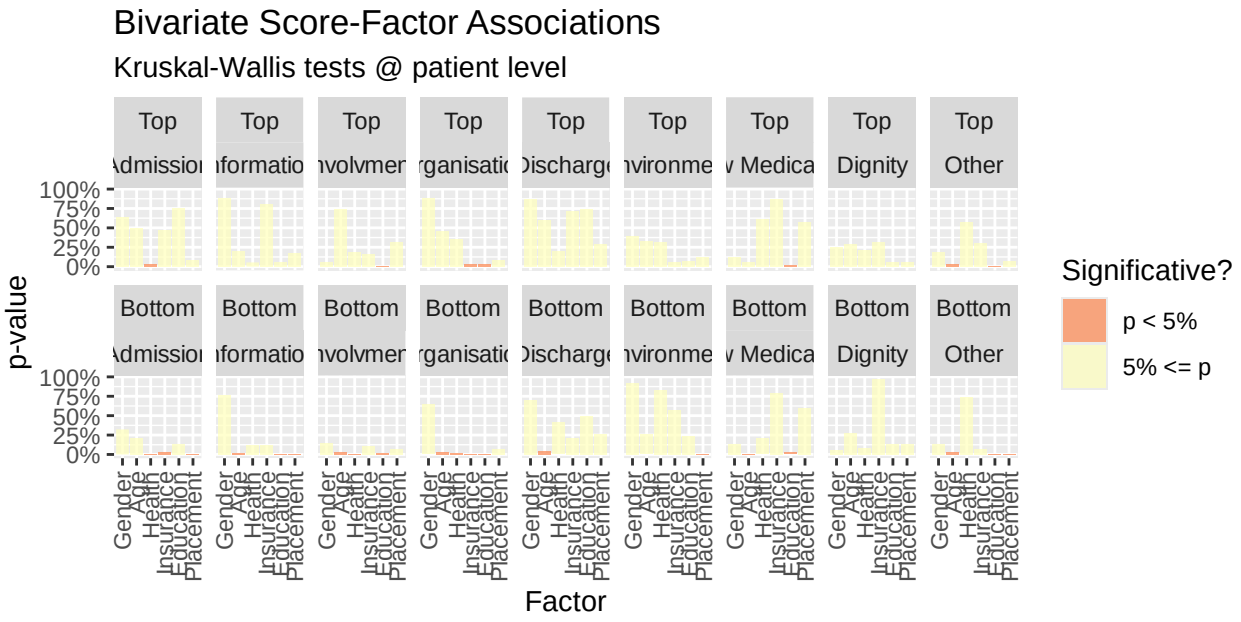


Figure 106 Significance level of association between the six socio-demographic factors and 16 thematic PREMs scores, as measured by a Kruskal-Wallis test computed on the entire pilot sample. Orange bars indicate that a significant association (p-value below 5%) was detected.

At Hospital Level

Admitting relative homogeneity among patients within each hospital, and acknowledging differences between hospitals, the same tests can be applied separately to each hospital's subsample. Figure 107 displays the proportion of hospitals in which a significant association was found for each factor-theme pair. Associations are more apparent than in Figure 106, particularly in the top boxes, which are also more densely populated. The clearest association between the education *Information* theme, confirmed in both top and bottom boxes. Beyond this, no recurrent pattern suggests that any factor consistently influences PREMs scores.

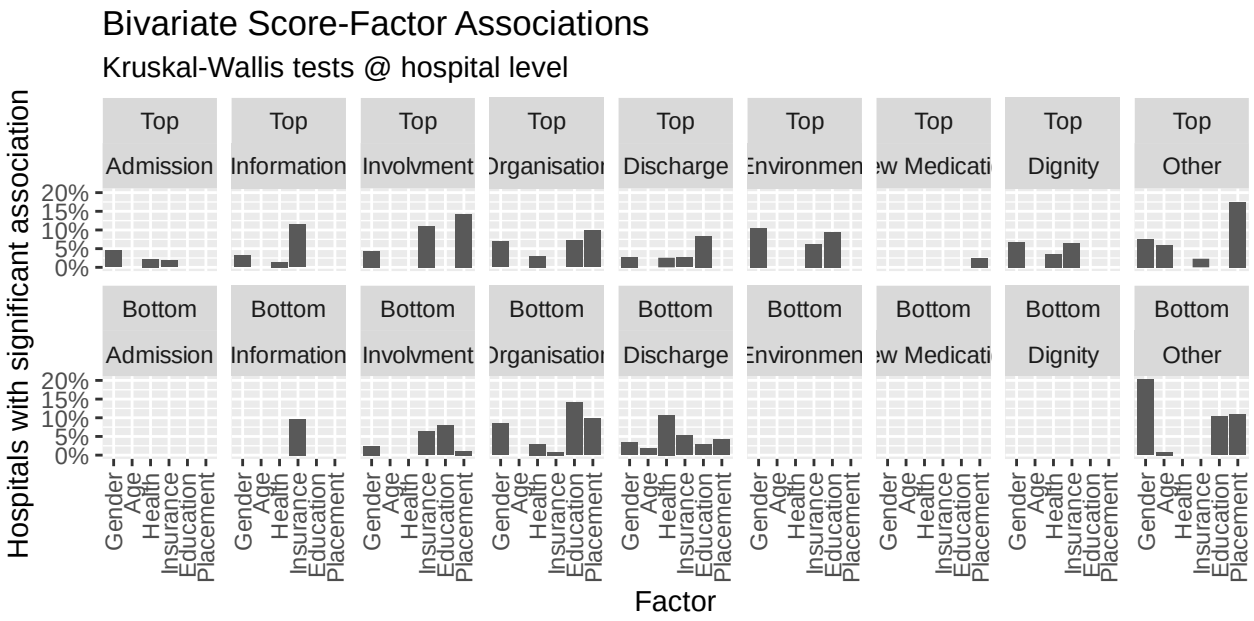


Figure 107 Frequency of significant associations between the six socio-demographic factors and the 16 thematic PREMs scores, as measured by a Kruskal-Wallis repeated on each hospital's subsample. Higher bars indicate that a significative association is detected more often.

Multivariate influence

Next, we incorporate all factors to compute multivariate associations using a binomial regression. The underlying assumption of the model is that each respondent has a consistent probability of selecting either the highest (top) or one of the two lowest (bottom) responses across all answered PREM items within the same composite score. Testing all factors at once reduces the number of tests to conduct to 16 = 2 boxes * 8 themes.

At Patient Level

Figure 108 illustrates how the sixteen predictions of PREMs scores yield consistent results. Depending on the theme, the effects (model coefficients) retain their sign and exhibit stable values, as reflected by the compact boxes in the figure. Note that statistical significance is not shown here. The figure is used purely for descriptive purposes to identify recurrent patterns. Although the impact of factors on PREMs scores is minimal in the psychiatry pilot study, it remains consistent.

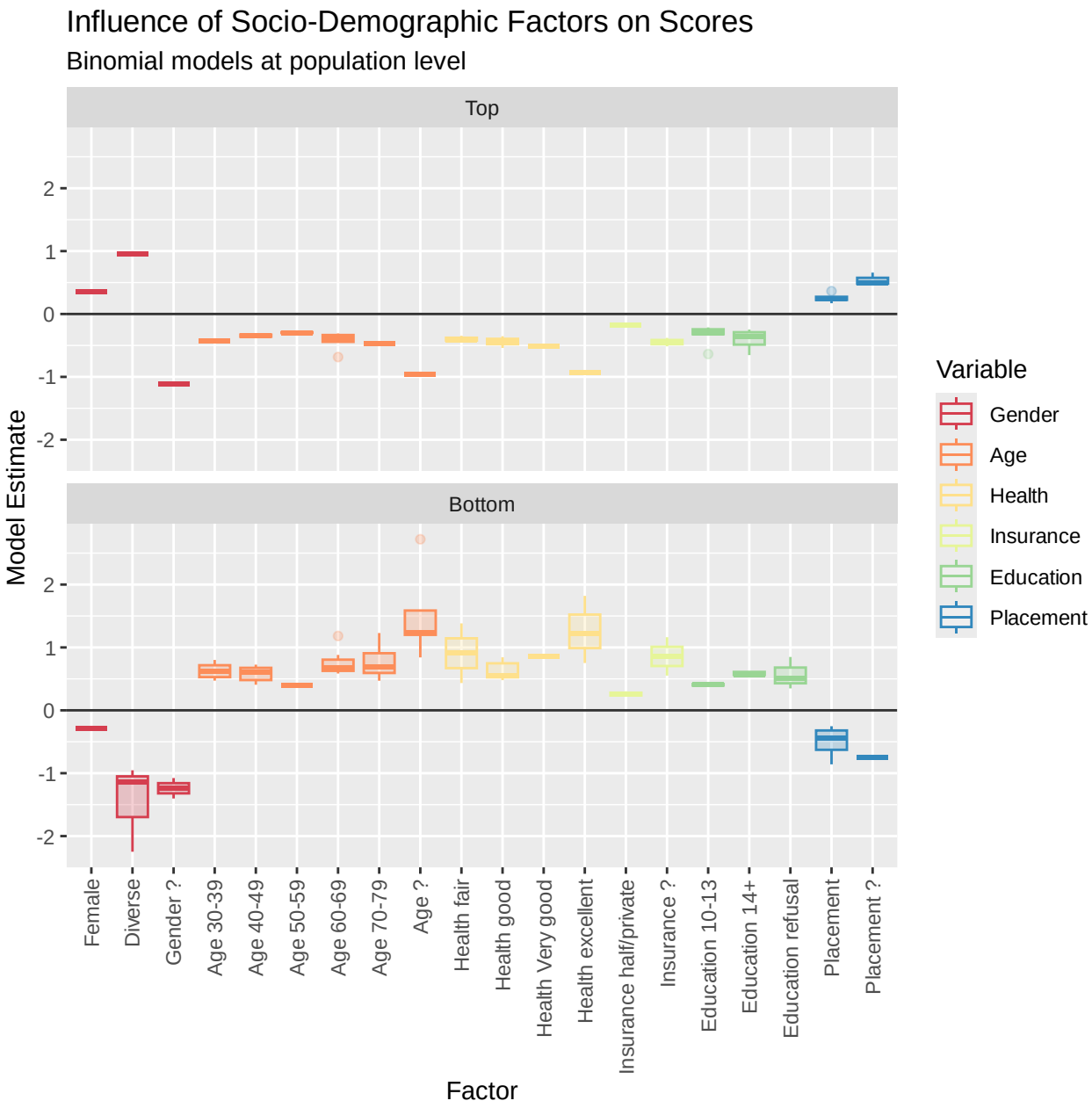


Figure 108 Distribution of binomial regression coefficients from the 16 PREMs score predictions. The reference profile represents the youngest male patients, with lowest education level, poorest self-reported health, basic insurance coverage, and voluntary admission to a psychiatry clinic.

5.3.3.Impact of Adjustment on PREMs Scores

Although various factors seem to influence scores, how significant is the change in a hospital's score after adjustment? We compare adjusted and unadjusted scores and analyze how the differences relate to socio-demographic variables.

Distribution of Adjustment Differences

Beyond the variation of scores across socio-demographic subgroups, the role factor may play in national comparisons occurs indirectly through the adjustment procedure. This section tries to measure to what extent this procedure changes the estimated scores.

Score differences, overall

The adjustment process is expected to yield minor corrections, as illustrated in Figure 109. Adjusted scores closely align with raw scores across all themes, with no major deviations are observed.

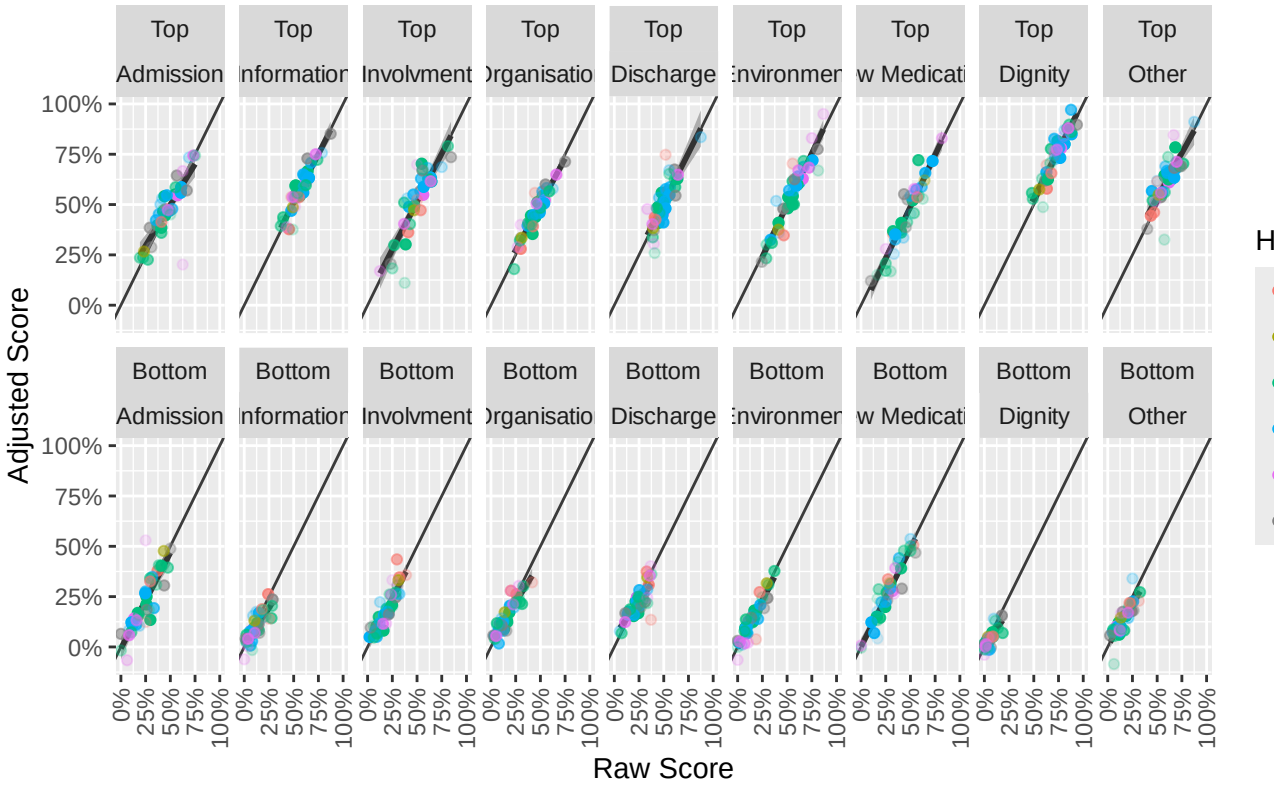


Figure 109 Raw versus adjusted PREMs score in the psychiatry pilot study.

Score differences, by theme and hospital

More precisely, when assessing the extent of deviations, the most pronounced changes occur in hospitals with very small sample sizes, making meaningful comparisons impossible. Among hospitals with more than 20 returned questionnaires, substantial changes, defined as deviations exceeding 10% of the original scores, are rare (highlighted in orange).

Figure 110 Absolute deviations of adjusted scores from their raw values. Each bar represents a hospital, with color indicating the magnitude of change.

Overall, above the 20-questionnaire threshold, minor changes (between 5% and 10%) affect no more than a quarter of the hospital sample, while more pronounced variations (more than 10%) can impact around 10% of the sample, as shown in Figure 111. The effect is more noticeable in top boxes, with the *Admission* and *New Medication* themes being the most affected.

Figure 111 Proportion of hospitals affected by adjustment in computing each PREMs score in the psychiatry pilot study.

5.3.4.Influence of Factors on Adjustment

Using linear regression to investigate which socio-demographic factors influence adjustment-induced differences, Figure 112 presents a nuanced picture. Gender and self-reported health status appear to have the most pronounced effects. Since all terms are binary indicator variables, length of bars can be interpreted directly as a contribution to percentage increase or decrease. While their effect reverses in the *Admission* theme, the results remain consistent across boxes. Regarding the placement variable, the indicator for respondents who did not answer the questions has a stronger effect than the indicator for involuntarily admitted patients.

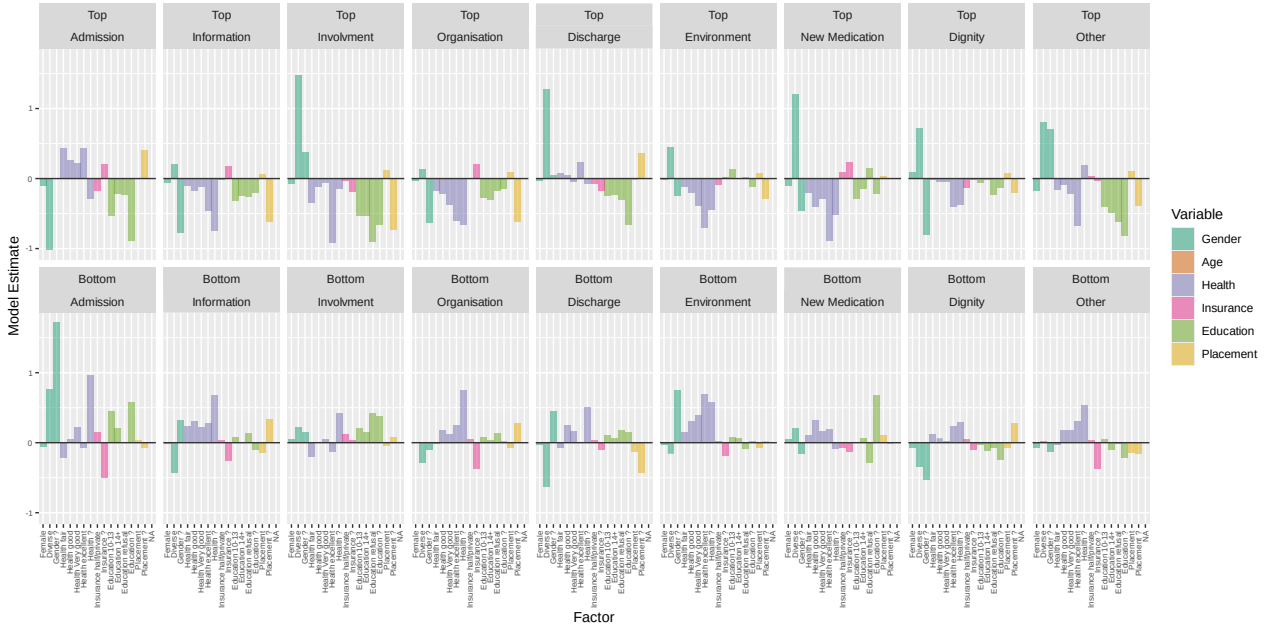


Figure 112 Linear regression coefficients of socio-demographic factors predicting the differences between adjusted and raw PREMs scores in the psychiatry pilot study.

5.3.5.Impact of Adjustment on Funnel Plots

A further investigation into the impact of adjustment on the hospital positioning within funnel plots reveals that nearly 90% of the points representing hospitals do not cross the confidence intervals. As a result, their performance with regard the national average remains unaffected by the adjustment procedure (see Table 44).

Table 44 Number of PREMs measures affected by adjustment in the psychiatry pilot study. A case is a hospital - PREMs score combinations, as one point on a funnel plot represents one hospital in a given measure. A hospital's position improves if it moves from the region located below the lower confidence interval into the funnel, if it exists the funnel to reach the region located above.

Position in funnel plot	Cases	
Improves	33	5.0%
Does not change	586	88.5%
Worsens	43	6.5%

A closer examination of Table 45 reveals the adjustment procedure has greater impact on bottom boxes, likely due to higher variability caused by smaller sample subgroups. Additionally, themes composed of more items tend to exhibit higher stability with respect to adjustment. *Information*, *Organization* and *Discharge* show the highest rates of unaffected hospital scores, while *Environment* and *Admission*, each composed of two items, display more pronounced changes.

Table 45 Proportion of hospitals whose position changes in the funnel plots of each PREMs score in the psychiatry pilot study.

	Top Box			Bottom Box		
Theme	Worsens	Does not change	Improves	Worsens	Does not change	Improves
Admission	8.1%	83.8%	8.1%	8.1%	86.5%	5.4%
Information	0.0%	94.6%	5.4%	8.1%	86.5%	5.4%
Involvement	8.1%	86.5%	5.4%	2.7%	91.9%	5.4%
Organization	0.0%	97.3%	2.7%	13.5%	78.4%	8.1%
Discharge	2.7%	97.3%	0.0%	2.7%	91.9%	5.4%
Environment	13.5%	86.5%	0.0%	8.1%	89.2%	2.7%
New Medication	2.7%	86.5%	5.4%	8.1%	86.5%	0.0%
Dignity	8.1%	81.1%	10.8%	5.4%	94.6%	0.0%

5.4. Structure of Final Questionnaires

Table 46 provides an overview of all questions (whether retained, discarded, or newly added) in the Swiss PREMs questionnaires. Both common and domain-specific items are listed in the order they appear in the questionnaire.

Table 46 Final structure of the Swiss PREMs questionnaires after shortening across the three domains of care. Identical items are aligned in the same row; question numbers from the pilot study are shown in gray, while the definitive numbering for the future Swiss PREMs questionnaires appears in black. Discarded items are highlighted in blue, and empty cells indicate non-relevant items.

		Psychiatry		Rehabilitation		Acute Care	
Question	Type	Pilot	Final	Pilot	Final	Pilot	Final
YOUR CARE FROM DOCTORS							
Courtesy and respect (doc.)	PREM	1	**1**	1	**1**	1	**1**
Careful listening (doc.)	PREM	2	**2**	2	**2**	2	**2**
Understandable explanations (doc.)	PREM	3	**3**	3	**3**	3	**3**
Availability (doc.)	PREM			4	**4**	4	**4**
YOUR CARE FROM THE NURSING STAFF							
Courtesy and respect (nurs.)	PREM	4	**4**	5	**5**	5	**5**
Careful listening (nurs.)	PREM	5	**5**	6	**6**	6	**6**
Understandable explanations (nurs.)	PREM	6	**6**	7	**7**	7	**7**
Availability (nurs.)	PREM	7	**7**	8	**8**	8	**8**
YOUR CARE FROM THERAPEUTIC STAFF							
Courtesy and respect (ther.)	PREM	8	**8**	9	**9**		
Careful listening (ther.)	PREM	9	**9**	10	**10**		
Understandable explanations (ther.)	PREM	10	**10**	11	**11**		
THE HOSPITAL/CLINIC ENVIRONMENT							
Cleanliness	PREM	11	**11**	12	**12**	9	**9**
Calm	PREM	12	**12**	13	**13**	10	**10**
Safe space	PREM	13	**13**				
YOUR EXPERIENCE IN THIS HOSPITAL/CLINIC							
Help for bathroom?	Filter	14		14		11	
Help for bathroom	PREM	15		15		12	**11**
Pain control	PREM	16		16		13	**12**
New Medicine?	Filter	17	**14**	17	**14**	14	**13**
Purpose of new medicine	PREM	18	**15**	18	**15**	15	**14**
Side-effects of new medicine	PREM	19	**16**	19	**16**	16	**15**
AFTER YOU LEFT THE HOSPITAL/CLINIC							
Destination?	Filter	20	**17**	20	**17**	17	**16**
Help after leaving	PREM	21	**18**	21	**18**	18	**17**

Symptom surveillance	PREM	22	**19**	22	**19**	19	**18**
YOUR OVERALL RATING							
Clinic rating	Satisfaction	23		23		20	
YOUR ARRIVAL AT THE HOSPITAL/ CLINIC							
Admission type	Other	25	**20**			22	**19**
Explanations on direction	PREM						**20**
Information on waiting	PREM						**21**
Information about admission	PREM	26	**21**	25	**20**	23	
Organization of admission	PREM	27	**22**	26	**21**	24	
Frequency of restriction	PREM	28	**23**				
Explanations for restriction	PREM	29	**24**				
Information about condition (emerg.)	PREM					25	
Information about unfolding (emerg.)	PREM					26	
Waiting for transfer (emerg)	PREM					27	
Organization of transfer (emerg.)	PREM					28	
DURING YOUR HOSPITAL/CLINIC STAY							
Communication between staff	PREM	30	**25**	27	**22**	29	**22**
Staff up to date	PREM	31	**26**	28	**23**	30	**23**
Punctuality	PREM	32	**27**	29	**24**	31	**24**
Information about health	PREM	33	**28**	30	**25**	32	**25**
Support with worries	PREM	34	**29**	31	**26**	33	**26**
Involvement (patient)	PREM	35	**30**	32	**27**	34	**27**
Involvement (reference)	PREM	36	**31**	33	**28**	35	**28**
YOUR HOSPITAL/CLINIC DISCHARGE							
Understanding of medications	PREM	37	**32**	34	**29**	36	**29**
Information about what to do	PREM	38	**33**	35	**30**	37	**30**
Better understanding of condition	PREM	39	**34**	36	**31**	38	**31**
YOUR OVERALL RATING							
Helpfulness	Satisfaction	40		37		39	
Experience rating	Satisfaction	41	**35**	38	**32**	40	**32**
Recommendation	Satisfaction	24	**36**	24	**33**	21	**33**
ABOUT YOU							
Self-reported health	Factor	42	**37**	39	**34**	41	**34**
Year of birth	Factor	43	**38**	40	**35**	42	**35**
Gender	Factor	44	**39**	41	**36**	43	**36**
Insurance category	Factor	45	**40**	42	**37**	44	**37**
Education	Factor	46		43		45	
Admission type					**38**		
Ease	Factor	47		44		46	

5.5. References

CIHI (2022). Acute Care Patient-Reported Experience Measures — Methodology Notes. Ottawa, Canadian Institute for Health Information**:** 20.
Epstein, J., R. M. Santo and F. Guillemin (2015). "A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus." Journal of Clinical Epidemiology **68**(4): 435-441.
Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. US, American Psychological Association. **76:** 378-382.
Hadibhai, S., J. Lacroix and K. Leeb (2018). "Developing the first pan-Canadian acute care patient experiences survey." Patient Experience Journal **5**(3): 25-33.
Lance, C. E., M. M. Butts and L. C. Michels (2006). "The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say?" Organizational Research Methods **9**(2): 202-220.
Landis, J. R. and G. G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data." Biometrics **33**(1): 159-174.
McHugh, M. L. (2012). "Interrater reliability: the kappa statistic." Biochem Med **22**(3): 276 - 282.
Savidan-Niederer, A., I. Peytremann Bridevaux and C. C. Storari (2022). Auswertungskonzept ANQ: Nationale Patientenzufriedenheitsmessungen. Lausanne, Unisanté
Storari, C., F. Cathieni, T. Carron and I. Peytremann Bridevaux (2022). Measuring Patient Satisfaction: Proposals for adaptations of existing instruments. Lausanne, Unisanté (Groupe ESOPE)**:** 39.
Storari, C., J. Junod, S. Crupi, S. E. Geiser and I. Peytremann Bridevaux (2024). Measuring Patient Satisfaction: CPES-IC translation, adaptation, and content validation. Lausanne, Unisanté**:** 185.

5.6. Pilot Questionnaires

As including the nine pilot questionnaires in the appendix would be cumbersome, readers are invited to contact the ANQ directly should they wish to consult the original survey material from the pilot study.