Journal of Mental Health

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ijmh20

# Latent structure and factor reliability of the National Health Service Community Mental Health Service User Questionnaire

Mauricio Scopel Hoffmann, Katia Bones Rocha, Sara Evans-Lacko, Natan Pereira Gosmann, Natalia Becker, Pedro Vieira da Silva Magalhães, Denise Razzouk, Lucas Spanemberg, Marcelo Pio de Almeida Fleck, Jair de Jesus Mari, Graham Thornicroft & Giovanni Abrahão Salum

**To cite this article:** Mauricio Scopel Hoffmann, Katia Bones Rocha, Sara Evans-Lacko, Natan Pereira Gosmann, Natalia Becker, Pedro Vieira da Silva Magalhães, Denise Razzouk, Lucas Spanemberg, Marcelo Pio de Almeida Fleck, Jair de Jesus Mari, Graham Thornicroft & Giovanni Abrahão Salum (2021): Latent structure and factor reliability of the National Health Service Community Mental Health Service User Questionnaire, Journal of Mental Health, DOI: 10.1080/09638237.2021.1922655

**To link to this article:** https://doi.org/10.1080/09638237.2021.1922655

JOURNAL OF MENTAL HEALTH
https://doi.org/10.1080/09638237.2021.1922655

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE



# Latent structure and factor reliability of the National Health Service Community Mental Health Service User Questionnaire

Mauricio Scopel Hoffmann[a,b,c,d], Katia Bones Rocha[a,e], Sara Evans-Lacko[d], Natan Pereira Gosmann[a], Natalia Becker[a], Pedro Vieira da Silva Magalhães[a], Denise Razzouk[f], Lucas Spanemberg[a,e], Marcelo Pio de Almeida Fleck[a], Jair de Jesus Mari[c,f], Graham Thornicroft[g] and Giovanni Abrahão Salum[a,c]

[a]Graduate Program in Psychiatry and Behavioral Sciences, Centro de Pesquisa Clínica, Faculty of Medicine, Universidade Federal do Rio Grande do Sul, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; [b]Departamento de Neuropsiquiatria, Universidade Federal de Santa Maria, Santa Maria, Brazil; [c]National Institute of Developmental Psychiatry for Children and Adolescents (INCT-CNPq), São Paulo, Brazil; [d]Care Policy and Evaluation Centre, London School of Economics and Political Science, London, United Kingdom; [e]Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil; [f]Department of Psychiatry, Univerisdade Federal de São Paulo, São Paulo, Brazil; [g]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

**ABSTRACT**

**Background:** National Health Service use the Community Mental Health Service User Questionnaire (NHS-CMH) to assess care quality. However, its reliability and internal validity is uncertain.
**Aims:** To test the NHS-CMH structure, reliability and item-level characteristics.
**Methods:** We used data from 11,373 participants who answered the 2017 NHS-CMH survey. First, we estimated the NHS-CMH structure using Exploratory Factor Analysis (EFA) in half of the dataset. Second, we tested the best EFA-derived model with Confirmatory Factor Analysis (CFA). We tested the internal validity, construct reliability (omega – $\omega$), explained common variance of each factor (ECV), and item thresholds.
**Results:** EFA suggested a 4-factor solution. The structure derived from the EFA was confirmed, demonstrating good reliability for the four correlated dimensions: "Relationship with Staff" ($\omega = 0.952$, ECV = 40.1%), "Organizing Care" ($\omega = 0.855$, ECV = 21.4%), "Medication and Treatments" ($\omega = 0.837$, ECV = 13.3%), and "Support and Well-being" ($\omega = 0.928$, ECV = 25.3%). A second-order model with a high-order domain of "Quality of Care" is also supported.
**Conclusions:** The NHS-CMH can be used to reliably assess four user-informed dimensions of mental health care quality. This model offers an alternative for its current use (item-level and untested sum scores analysis).

**ARTICLE HISTORY**
Received 26 February 2021
Revised 7 April 2021
Accepted 11 April 2021
Published online 28 April 2021

**KEYWORDS**
Mental health service; care quality; factor analysis

## Introduction

Mental health conditions are prevalent worldwide and globally responsible for 32.4% of years lived with disability and 13.0% of disability-adjusted life-years (Vigo et al., 2016). In such a scenario, societies must provide high-quality mental health care to reduce this burden of disease (IOM, 2001; Vigo et al., 2016; WHO, 2008) especially in the primary and community setting, where the majority of patients live most of their lives (Demyttenaere et al., 2004; Kessler et al., 2009; Shen & Snowden, 2014). To improve mental health services, an important source of information comes from the assessment of care quality from those who receive it, which is associated with treatment adherence and patterns of service utilization (Miglietta et al., 2018; Ruggeri, 1994; Stamboglis & Jacobs, 2020). Therefore, a fundamental step in reducing burden of mental health conditions is to reliably measure dimensions of user-informed quality of care.

In the last few decades, several instruments have been developed to assess quality of mental health care (Kilbourne et al., 2018; Miglietta et al., 2018; Ruggeri, 1994; Woodward et al., 2017). Those instruments aim to cover if effective interventions are being delivered and the subjective patient experience in receiving health care. This is consistent with the World Health Organization (WHO) definition that "*quality is a measure of whether services increase the likelihood of desired mental health outcomes and are consistent with current evidence-based practice*" (WHO, 2003). In that sense, the Crossing the Quality Chasm report emphasizes the measurement of six care quality dimensions, namely patient safety, care effectiveness, timeliness, care efficiency, equity and patient-centeredness (IOM, 2001).

According to a recent review, 25 scales are available to assess the quality of mental health care up until 2018 (Miglietta et al., 2018). A total of 19 dimensions of care are covered and overall satisfaction, perceived care outcomes

CONTACT Mauricio Scopel Hoffmann mauricio.hoffmann@ufsm.br Graduate Program in Psychiatry and Behavioral Sciences, Centro de Pesquisa Clínica, Faculty of Medicine, Universidade Federal do Rio Grande do Sul, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

Supplemental data for this article can be accessed here.

and relationship with staff are consistent dimensions across these instruments. Despite designed to be multidimensional, structural validity analysis demonstrated that most of these scales present one to four dimensions (Miglietta et al., 2018; Schröder et al., 2010). Except for two, these instruments are usually structured as pure or mixed Likert type scales (using two to 10 rating points). Another study reviewed the psychometric properties of 22 instruments and found that most of them presented low reliability, internal consistency and structural validity (Schröder et al., 2010). Lack of psychometric evaluation of these instruments hinder the progress of the user experience evaluation in relation to the quality of mental health services.

To address community mental health care quality based on user experience, the National Health Service (NHS) of the United Kingdom has used the Community Mental Health Service User Questionnaire (NHS-CMH) since 2004. The purpose of the survey is to understand, monitor and improve user experience of NHS community mental health services. It is run by the Survey Coordination Centre for Existing Methods on behalf of the Care Quality Commission. In 1991, the Department of Health in England issued the Care Programme Approach, which requested the NHS trust to "*provide a systematic assessment of the health and social care needs of the patient*" (Webb et al., 2000). Additionally, the NHS patient experience framework established that "*information, communication, and education on clinical status, progress, prognosis, and processes of care*" are necessary "*in order to facilitate autonomy, self-care and health promotion*" (NQB, 2012). These are the guiding components of the NHS-CMH.

Following this recommendation, the "Your Treatment and Care" assessment tool was developed using four dimensions (user care plan, relationship to key care worker, relationship with psychiatrist and if sufficient information was provided). They presented a Cronbach's alpha reliability above 0.8 (Webb et al., 2000). This inspired the first version of the NHS-CMH (Osborn et al., 2004). Questions were included (scored) if they provided actionable information. The 2017 NHS-CMH is the latest survey publicly available data. It was formulated to have ten dimensions, namely (1) Care and treatment, (2) Health and social care workers, (3) Organizing care, (4) Planning care, (5) Reviewing care, (6) Changes in who people see, (7) Crisis care, (8) Treatments, (9) Support and wellbeing and (10) Overall. The instrument was developed to generate weighted information and time-trends to inform the NHS on which areas of care must be improved.

So far, NHS-CMH data have been used as a summed score of different sets of items, averaging all domains as part of a single construct, or by comparing trusts by weighted item-level analysis (Care Quality Commission [CQC], 2017a; Stamboglis & Jacobs, 2020). The problem with a sum score without checking their covariance matrix is that the items might not fit in the desired construct and, therefore, the interpretation of a sum score is compromised due to mixing information of distinct dimensions. Another argument of modern measurement theory is that items are indicators of a latent construct that are used to triangulate and capture the information of interest (Kline, 2015; Reise & Waller, 2009; Stevens, 1946). Therefore, item-level analysis (comparing trusts based on their responses in each item of the questionnaire) is problematic because single item analysis might reflect measurement error, rather than changes in the underlying latent constructs of interest (Brown, 2015; Williams, 2019). To improve the understanding of what NHS-CMH is measuring, factor analysis can be used (Brown, 2015). It can provide information on what these quality factors/dimension would be, how reliable they are, and which items are more sensitive to the variations of the latent quality factors (i.e. factor loading), as well as how much of a dimension of care quality one should perceive to endorse higher scores (i.e, thresholds or difficulty). This is relevant to understanding if and how multiple dimensions of care are perceived by the user, which in turn can help to improve dimensions of care.

Therefore, despite containing useful information for policy and improvement of mental health care, the internal validity, factor reliability, factor loadings and item threshold of the NHS-CMH have never been analysed and those are our aims in the present study. Hence, we carried out an exploratory factor analysis (EFA) to derive the factor structure of NHS-CMH and confirm it using confirmatory factor analysis (CFA) using the 2017 survey database which was the most updated publicly available. Our aims are to estimate the structure, reliability, explained common variance of each factor, factor loadings and how much quality one should perceive to endorse higher response categories (i.e. item threshold). We hypothesized that, similar to "Your Treatment and Care" assessment tool, the factors would be highly correlated and reliable for capturing dimensions of care quality as perceived by the user.

## Materials and methods

We used the most updated free available NHS-CMH survey data set (CQC, 2018). Between February and June 2017, NHS-CMH questionnaires were sent to 47,600 UK citizens. To select participants, each NHS trust ($N = 56$) drew a random sample from their record, to reach 850 subjects aged 18 years or more, who had been seen at the trust for specialist care or treatment for a mental health condition during the sampling period (1 September to 30 November 2016). Participants were excluded before providers drew the random sample if the person: (a) was a current inpatient, (b) was seen only once for an assessment, (c) was seen for assessment only through a liaison service, (d) was primarily receiving services for drug and alcohol, learning disability, forensic, psychological treatments from improving access to psychological therapies, chronic fatigue, psychosexual medicine (sexual dysfunction) or gender identity, or (e) if the person had only been in contact by telephone or email, and had not been seen in person at all.

Ethical approval has been granted by the East of England - Cambridge East Research Ethics Committee (reference 15/EE/0064). It was also granted that participant's information

was processed without consent by the NHS Health Research Authority considering the recommendation from the Confidentiality Advisory Group (reference 16/CAG/0157) because it was understood that "*taking consent would be impracticable for health care staff, would remove the cost and time benefits of employing a contractor to mail out surveys, and would introduce bias into the study due to variable response rates from patients*" (CQC, 2017b).

By the end of June 2017, 12,139 adults responded to the survey (26% response rate). Of the respondents, 57% were females, 14% have less than 35 years of age (18–35), 46% have 36–65 years of age and 40% have 66 years or more (CQC, 2018). We have included only subjects who have seen someone from the NHS for their mental health needs in the previous 12 months to be aligned with the general phrasing of the questions ($N = 11{,}373$). The dataset was randomly divided in half so each part could be used in the exploratory ($n = 5687$) and confirmatory ($n = 5671$) factor analysis.

The NHS-CMH questionnaire uses scored questions (0, 5 or 10) which were further included in the analysis. However, the instrument contains 16 questions that are conditioned upon 9 skip questions. These dependencies between items can cause violations in factor analysis and therefore, we used a parcelling strategy for the scored questions of the above-mentioned dimensions. Item parcels were built in the following manner: (1) If a skip question was coded "0" ("No" or "Not sure"), the parcel was scored as zero. (2) Missing was considered if all questions involved in the parcel were missing or if the skip question was not scored and coded to skip the next question. (3) If the skip question was coded in a way that allows the endorsement of the dependent items (usually "Yes"), the scored items were summed to indicate the highest value of quality in the specific domain/parcelled item. (4) Furthermore, there were some cases in which the skip question was coded as missing but there was an endorsement in any of the dependent question. In such cases we imputed the minimum scored value which would have allowed the dependent question to be endorsed. The following NHS-CMH domains were parcelled: "Organising care" (Q7, Q9 and Q10), "Planning care" (Q11, Q12 and Q13), "Reviewing care" (Q14, Q15 and Q16), "Changes in who people see" (Q17, Q18, Q19 and Q20), "Crisis care" (Q21, Q22 and Q23) and "Treatments" (Q24 and Q25 on involvement in medication decisions; Q24, Q26 and Q27 on information about new medicines; Q24, Q28 and Q29 on medication revisions; Q30, Q31 and Q32 on explanation and involvement about treatments and therapies). In addition, one question (Q10) contained the score options 3.3 and 6.7, which were rounded between 0 and 10 respectively. The question regarding overall satisfaction (Q40) ranged from 0 to 10 with unit intervals. For the purpose of further analysis, we categorized the Q40 into a three-level scored item (0–3 = 0; 4–7 = 5; 8–10 = 10). Therefore, the number of items used in the analysis is 22. The procedure can be verified in cross-tabulation of the raw scored variables and the parcelled domains (Tables S1 to S9) as well as the supplementary R code (10.17605/OSF.IO/SBUP5).

### Statistical analysis

We performed factor analysis using two steps aiming to estimate the structure of the questionnaire and, by confirming it, estimate the properties of the items and the reliability of the NHS-CMH constructs. First, we performed exploratory factor analysis (EFA) to test how many and which factors emerge from the dataset (from one to 10 factors) using geomin rotation. This approach allows free covariance among the items and derives the domains from the data. The best model was selected on the basis of model convergence, fit, Eigenvalue, $\chi^2$ test comparison between models, dimensions with at least three items with factor loadings $\geq 0.3$ (Hayton et al., 2004) and minimum cross-loadings (i.e. items with factor loadings $\geq 0.3$ by more than one factor). Factor loadings are the root-squared of the item standardized variance predicted by the latent factor and informs how much ($-1$ to 1) of the item variability are explained by the latent factor.

Next, we used confirmatory factor analysis (CFA to test the EFA-derived best model. Aside from applying the structural constraints to confirm the model, we used CFA to assess item threshold and factor reliability. Thresholds (or item difficulty) indicate the 50% probability of endorsing a given (scored) category or higher as a function of the latent trait (e.g., from "Yes, to some extent" to "Yes, definitely"). In other words, is the standardized score beyond which participants respond a higher category (Brown, 2015). Factor reliability was assessed using Lucke's omega ($\omega$), a model-based reliability estimate being analogous to alpha coefficient, but appropriate for congeneric tests (varying factor loadings) (Lucke, 2005; Raykov, 2001), and the percent of explained common variance index (ECV), defined as the ratio of variance explained by a given factor divided by the variance explained by all the factors and used as an index of unidimensionality (Bentler, 2009; Reise, 2012). Different from EFA, this approach allows to constraint the items to belong to a theoretically or empirically derived specified domain.

CFA and EFA were estimated using delta parameterization and weighted least square using a diagonal weight matrix with standard errors and mean- and variance-adjusted $\chi^2$ test statistics (WLSMV) estimators using MPLUS 8.4 software (Muthén and Muthén, Los Angeles, California, USA), which apply full information maximum likelihood (FIML) to estimate models using all information available in a dataset with missing data. Item variance were set to 1 to all factors. Model fit parameters were $\chi^2$ test of model fit, root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker Lewis Index (TLI) and Standardized Root Mean Square Residual (SRMR). Values of RMSEA or SRMR near or below 0.08 represent acceptable model fit, and values lower than 0.06 represent good-to-excellent model fit (Brown, 2015; Hu & Bentler, 1999). CFI and TLI values near or above 0.90 represent acceptable model fit, while values higher than 0.95 represent a good-to-excellent model fit (Brown, 2015; Hu & Bentler, 1999). All analysis code can be found at 10.17605/OSF.IO/SBUP5 and supplemental material.

## Results

### Data description

Table S1 described the proportion of scored items endorsement in the dataset used for the factor analysis. Cross tabulation of raw scored questions and parcelled items can be found in Tables S2 to S10. After the parcelling procedure,

### Exploratory factor analysis

EFA was set to estimate solutions with one to 10 factors. Model fit was acceptable for all models (Table S11). Eigenvalues were higher than 1 for solutions with one to three factors (Table S11). The four-factor solution presented a better fit if compared with the three-factor solution (Table S12 for EFA model comparisons) and all factors of this model presented factor loadings ≥0.3 in at least three items, without any item left unexplained by the model. Solutions with five factors or more presented factors with less than three items per factor and items not loaded on any factor (please see Supplementary Tables S13 to S20 for complete description of all EFA models). Therefore, an interpretable four-factor solution was selected. Table 1 demonstrates factor loadings and correlation of the extracted factors. In addition, polychoric correlation matrix used in the EFA can be seen in Table S21.

Item content suggests it is possible to define these factors as "Relationship with staff" (EFA Factor 1), "Organizing care" (EFA Factor 2), "Support and wellbeing" (EFA Factor 3) and "Medication and Treatments" (EFA Factor 4). Item Q37 presented cross-loading in Factor 1 and 3. This was not observed in solutions with fewer and higher number of factors, which was significantly loaded by Factor 1 only. Added to the shared content with other items in Factor 1 (i.e. understanding of needs and preferences by the mental health service) we considered Q37 to belong to Factor 1 in the CFA.

### Confirmatory factor analysis

From the EFA, it is possible to observe a high correlation among the factors and a strong drop in Eigenvalues from one to two-factor EFA solution (Table S11). Thus, EFA results also suggest that is possible to confirm this four-factor solution via a four correlated factor model or a hierarchical second-order model, which estimates the correlation among first-order model as an overarching general factor. Therefore, we estimated these two models in the CFA.

The four-factor correlated model converged normally, with good model fit (RMSEA = 0.041; 90%CI = 0.039–0.042; CFI = 0.987; TLI = 0.985; SRMR = 0.036). Factor loadings and item thresholds are described in Table 2 and the polychoric correlation matrix used in the CFA can be seen in Table S22. ECV was higher for the "Relationship with Staff", being the factor that explains most of NHS-CMH variance (40.1%) and presents the highest internal consistency ($\omega = 0.952$).

**Table 1.** Geomin-rotated factor loadings and correlation from the fitted EFA four-factor model of the 2017 NHS-CMH ($n = 5.687$).

| Items | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| Q3 | **0.597** | 0.132 | 0.19 | −0.072 |
| Q4 | **0.998** | 0.001 | −0.096 | −0.032 |
| Q5 | **0.914** | 0.027 | −0.03 | −0.035 |
| Q6 | **0.769** | 0.061 | 0.021 | 0.054 |
| Q7 + Q9 + Q10 | 0.028 | **0.820** | −0.012 | −0.121 |
| Q11 + Q12 + Q13 | 0.075 | **0.723** | −0.011 | 0.164 |
| Q14 + Q15 + Q16 | 0.044 | **0.578** | 0.098 | 0.183 |
| Q17 + Q18 + Q19 + Q20 | 0.187 | **0.509** | 0.075 | 0.022 |
| Q21 + Q22 + Q23 | −0.238 | **0.504** | 0.117 | 0.040 |
| Q24 + Q25 | 0.081 | 0.026 | 0.01 | **0.746** |
| Q24 + Q26 + Q27 | −0.013 | 0.203 | 0.016 | **0.685** |
| Q24 + Q28 + Q29 | 0.025 | **0.364** | 0.224 | 0.130 |
| Q30 + Q31 + Q32 | 0.218 | 0.233 | −0.001 | **0.410** |
| Q33 | 0.082 | 0.118 | **0.722** | 0.027 |
| Q34 | −0.036 | 0.070 | **0.880** | −0.063 |
| Q35 | −0.034 | 0.109 | **0.845** | −0.004 |
| Q36 | 0.043 | 0.009 | **0.807** | −0.050 |
| Q37 | **0.466** | −0.093 | **0.328** | 0.116 |
| Q38 | 0.053 | −0.141 | **0.823** | 0.095 |
| Q39 | **0.442** | 0.178 | 0.240 | 0.142 |
| Q40 | **0.662** | 0.120 | 0.092 | 0.098 |
| Q41 | **0.733** | −0.053 | 0.056 | 0.145 |
| Correlations | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
| Factor 1 | 1.000 | | | |
| Factor 2 | 0.704 | 1.000 | | |
| Factor 3 | 0.641 | 0.737 | 1.000 | |
| Factor 4 | 0.653 | 0.544 | 0.533 | 1.000 |

*Note*: Bold numbers are factor loadings ≥0.3. EFA: Exploratory factor analysis; NHS-CMH: National Health Service Community Mental Health Service User Questionnaire. Summed items were parcelles of skip questions and their scored items.

The four-factor second-order model converged normally and presented good model fit (RMSEA = 0.041; 90%CI = 0.040, 0.043; CFI = 0.986; TLI = 0.985; SRMR = 0.038). Factor loadings of the overarching "Quality of care" into "Relationship with staff", "Organizing care", "Medication and Treatments" and "Support and Wellbeing" are 0.917 ($R^2$ = 0.841), 0.925 ($R^2$ = 0.856), 0.874 ($R^2$ = 0.764) and 0.855 ($R^2$ = 0.731) respectively. Therefore, this instrument can also be conceptualized as having a hierarchical structure, for which there is an overarching general factor strongly related with all care quality dimensions.

The highest threshold is observed for the parcelled questions regarding crisis care (Q21 + Q22 + Q23). In this item, a service probably has 1.242 standardized care quality latent score (interpreted as z-scores) so a subject can have 50% probability to score from 15 (meaning "Yes" for Q21 and Q22, and "Yes, to some extent" for Q23) to 20 (meaning "Yes" for Q21 and Q22, and "Yes, definitely" for Q23) in this domain. In other words, these are the most difficult items to be endorsed and, when they are, indicate higher quality. "Relationship with Staff" and "Medication and Treatment" dimensions mostly have negative thresholds, which means that they are informative at the lower end of the care quality latent trait (Table 2, thresholds) and most of users will respond positively in those items.

## Discussion

As stated by The Care Quality Commission in the UK, "*to improve the quality of services, it is important to understand what people think about their care and treatment*" (CQC,

**Table 2.** Standardized factor loadings and thresholds of item endorsment levels from the fitted 2017 NHS-CMH confirmatory factor analysis ($n = 5671$).

| | | | Threshold[a] | | | | | |
|---|---|---|---|---|---|---|---|---|
| Factor | Item | Factor loading | 1 | 2 | 3 | 4 | 5 | 6 |
| Relationship with staff | Q3 | 0.809 | −0.758 | 0.065 | | | | |
| | Q4 | 0.885 | −1.506 | −0.541 | | | | |
| | Q5 | 0.860 | −1.223 | −0.339 | | | | |
| | Q6 | 0.860 | −1.138 | −0.166 | | | | |
| | Q37 | 0.734 | −0.864 | −0.172 | | | | |
| | Q39 | 0.891 | −0.959 | 0.082 | | | | |
| | Q40 | 0.886 | −1.186 | −0.131 | | | | |
| | Q41 | 0.829 | −1.472 | −0.650 | | | | |
| Organizing care | Q7 + Q9 + Q10 | 0.695 | −0.720 | −0.291 | −0.073 | | | |
| | Q11 + Q12 + Q13 | 0.882 | −0.739 | −0.582 | −0.390 | −0.051 | 0.229 | 0.515 |
| | Q14 + Q15 + Q16 | 0.839 | −0.642 | −0.478 | −0.317 | 0.074 | 0.347 | |
| | Q17 + Q18 + Q19 + Q20 | 0.729 | −1.137 | −0.655 | −0.225 | 0.039 | 0.528 | 0.712 |
| | Q21 + Q22 + Q23 | 0.355 | −0.578 | 0.864 | 1.242 | | | |
| | Q24 + Q28 + Q29 | 0.665 | −0.751 | | | | | |
| Medication and treatments | Q24 + Q25 | 0.788 | −1.247 | −0.067 | | | | |
| | Q24 + Q26 + Q27 | 0.757 | −1.092 | −0.108 | | | | |
| | Q30 + Q31 + Q32 | 0.836 | −1.861 | −1.143 | −0.434 | 0.047 | | |
| Support and wellbeing | Q33 | 0.924 | −0.388 | 0.335 | | | | |
| | Q34 | 0.833 | −0.154 | 0.454 | | | | |
| | Q35 | 0.884 | −0.232 | 0.578 | | | | |
| | Q36 | 0.790 | −0.213 | 0.508 | | | | |
| | Q38 | 0.810 | −0.052 | 0.686 | | | | |
| Factor correlation | | | | | | | | |
| | | Organizing care | Medication and treatments | Support and wellbeing | | | | |
| | Relationship with staff | 0.842 | 0.824 | 0.780 | | | | |
| | Organizing care | | 0.792 | 0.816 | | | | |
| Medication and treatments | | | | 0.695 | | | | |
| | Support and wellbeing | | | | | | | |
| Factor reliability | | $\omega$ | ECV | | | | | |
| | Relationship with staff | 0.952 | 40.1% | | | | | |
| | Organizing care | 0.855 | 21.4% | | | | | |
| Medication and treatments | | 0.837 | 13.3% | | | | | |
| Support and wellbeing | | 0.928 | 25.3% | | | | | |

*Note:* [a]Threshold are amount of standardized factor score that is necessary for a subject to have 50% probability to endorse from a previous response category to the next; The threshold numbers 1–6 indicate how many category changes are possible in a given item or parcelled item. $\omega$: reliability coefficient omega; ECV: explained common variance; NHS-CMH: National Health Service Community Mental Health Service User Questionnaire.

2018). In this perspective, we describe the latent structure, reliability and item characteristics of the 2017 NHS-CMH. We estimated that it is composed by four dimensions, describing relationship with staff, organizing care, medication and treatments and support and well-being. Factor reliability was good to excellent and corresponded to between 13.3 and 40.1% of the NHS-CMH factor-explained variance. Item threshold could also be assessed to inform the 50% probability of endorsing a higher response category based on the continuous level of the user satisfaction with the provided dimension of care.

## *Internal validity and reliability*

More than 30 instruments aim to measure the user's perspective on mental health care quality. A few are designed for community mental health care and even fewer evaluate what constructs they are measuring and how reliable they are (Miglietta et al., 2018). The 2017 NHS-CMH questionnaire intended to cover 10 dimensions of care. However, six of these dimensions included conditional questions that made them function as an item instead of a dimension. Therefore, the proposed dimensions: "Organising care", "Planning care", "Reviewing care", "Changes in who people see" and "Crisis care", were highly correlated and were informed by one dimension if taken as an item (parcelled). We kept the name "Organizing care" because the content of the items relates to how well, how much information, and how autonomous a user was regarding all these organizational aspects. This construct also informed one item of the "Treatment" dimension, which regards on information about new medicines. Due to the nature of the skip questions, the "Treatment" dimension could be broken into four parcelled items, in which three loaded into the "Medication and Treatments" dimension.

The NHS-CMH constructs are covered in most studies (Miglietta et al., 2018). Relationship with staff is present in all instruments evaluated by a recent review (Miglietta et al., 2018) and covers feeling listened to, comfortable and able to talk with staff. Most of 2017 NHS-CMH items load into this dimension as revealed by EFA and CFA, including all items from the "Your Health and Social Care Worker", two items from the "Support and Well-being" and the two items from the "Overall" evaluation. We estimated that 40.1% of the explained variance due to latent factors correspond to this construct. Taken together, it reveals that this is a consistent and representative dimension of mental health care quality.

Most NHS-CMH items are phrased to capture if the user is well informed, involved with treatment, checked and reviewed on their needs. In this sense, it is well related with patient-centeredness approach to increase care quality (IOM, 2001; Kilbourne et al., 2018). The common phrasing among dimensions is a plausible reason why they are highly

correlated, and a second-order model fit the data well. Nonetheless, a one-factor solution does not describe the questionnaire and some room for specificity remains. Therefore, the structure of the NHS-CMH can be interpreted as a four first-order correlated factor model, with a general factor that explains the correlation among the first-order factors.

### How much care quality one should perceive to respond higher scores?

Previous studies have not shown how much quality a user should perceive to make them endorse higher response category in the questionnaires (Miglietta et al., 2018; Sanchez-Balcells et al., 2018). Here we have estimated the NHS-CMH item threshold, which informs how much quality a user must perceive in the continuous latent trait so the probability to respond "Yes, to some extent" and "Yes, definitely" is equally 50%. With that in mind, we can observe that the question 41, which asks about if the user was treated with respect and dignity, have very low thresholds to respond from "Yes, sometimes" to "Yes, always" (−0.650), which means that a service does not need to be high in quality so a user can respond to the extreme positive of this question (72% responded "Yes, always"). NHS-CMH four dimensions of care captures different levels in the care quality. The "Support and Well-being" dimension presents the higher thresholds if single questions are considered, meaning that for a user endorse "Yes, definitely" to these questions, it is more likely that a service is beyond mean level of support and well-being quality.

"Relationship with Staff" presents high rates of higher response categories, which can be observed from the raw percentages of response in a given category, or by the low thresholds of its items. This construct is present in most instruments measuring care quality and satisfaction (Miglietta et al., 2018). Here we demonstrated that despite its importance and detection by the user, it measures quality in the lower level of the quality trait and does not help to discriminate average from excellent quality. Future studies on this dimension may benefit from evaluating item threshold to improve this dimension so it can capture more information in the quality latent trait.

### Limitations

This study has limitations that may require additional scrutiny. First, the psychometric properties of the NHS-CMH may conceivably vary according to a number of participant characteristics. Unfortunately, the data set does not provide individual demographics, so the instrument cannot be probed for measurement equivalence among these groups. Second, response rate was low and the extrapolation of these results to the general public must be cautionary. However, the data collection procedure is carefully planned so the subjects are representative of all subjects that seek NHS community mental health services across the country. Third, due to exclusion criteria at the sampling period, subjects with specific characteristics are not represented in this survey and analysis (i.e. subjects that were receiving inpatient care at that moment, were seen for assessment only once or through a liaison service, were seen primarily for specific conditions such as drug and alcohol, or if the subjects were not seen in person). Forth, due to missing data and dependency among questions, some original dimensions were impossible to be estimated. We tried to overcome this by using an item parcelling strategy to account for dependencies caused by conditional questions, as well as the utilization of WLSMV estimator using FIML to handle missing data, which is the state of the art for dealing with such data (Kline, 2015)

## Conclusions

Psychometric evaluation of instruments measuring the user experience of community mental health care quality can aid in the understanding of what constructs they are measuring, how the items aggregate and can be reliably summed to provide useful scores, and which items reveal higher care quality if endorsed. The NHS-CMH is particularly structured with conditional questions which limited the estimation and reliability of what was originally intended to be measured. To overcome this, it is possible to modify items in the next surveys, so they are not interdependent, or analyse them as parcelled items. Another possibility is optimization, which can be achieved by excluding theorized constructs with dependent questions and include other constructs from other instruments (Miglietta et al., 2018), such as accessibility and impressions on service environment. Nonetheless, we estimate that the NHS-CMH presents dimensions with good internal reliability. Moreover, item thresholds demonstrate that response levels from questions of the "Support and Wellbeing" and the "Organizing Care" dimension are useful to capture mean to higher care quality levels.

The NHS-CMH is a survey questionnaire that covers four correlated dimensions of care quality as perceived by the user. It suits the patient-centeredness approach by asking if the users are well informed, involved with treatment, checked and reviewed on their needs. The questions are scored through "Relationship with Staff", "Organizing Care", "Medication and Treatments" and "Support and Wellbeing" dimension, which could be reliably used as separate scales, depending on the measurement purpose. Regardless of the purpose, if one dimension is measured, the other three can also be inferred due to its high correlation. Nonetheless, the single-item analysis is informative of other items via its common dimension, as demonstrated by its factor structure. Therefore, this questionnaire should be applied and interpreted as a sum of its items in each of its four dimensions of mental health care quality.

## Disclosure statement

All authors declare no known conflict of interests.

## Funding

This Project is supported by the Brazilian Ministry of Health under the Termo De Execução Descentralizada – [TED 12/2019].

## ORCID

Mauricio Scopel Hoffmann http://orcid.org/0000-0003-4232-3169
Jair de Jesus Mari http://orcid.org/0000-0002-5403-0112
Giovanni Abrahão Salum http://orcid.org/0000-0002-7537-7289

## Data availability statement

The datasets were derived from sources in the public domain at https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8312#!/details. The code underlying the processing and analysis of the data included in this article are available in the online supplemental material and at 10.17605/OSF.IO/SBUP5.

## References

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*(1), 137–143. https://doi.org/10.1007/s11336-008-9100-1

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press. https://www.guilford.com/books/Confirmatory-Factor-Analysis-for-Applied-Research/Timothy-Brown/9781462515363/reviews

Care Quality Commission. (2017a). *Community mental health survey 2017*. NHS Surveys. https://nhssurveys.org/surveys/survey/

Care Quality Commission. (2017b). *2017 Community mental health survey—Statistical release*. 66. NHS

Care Quality Commission. (2018). *National Health Service patient survey programme, 2003-community mental health service user survey, 2017*. PIE. UK Data Service; UK Data Service. SN: 8312. https://doi.org/10.5255/UKDA-SN-8312-1

Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V., Lepine, J. P., Angermeyer, M. C., Bernert, S., de Girolamo, G., Morosini, P., Polidori, G., Kikkawa, T., Kawakami, N., Ono, Y., Takeshima, T., Uda, H., Karam, E. G., Fayyad, J. A., Karam, A. N., … Chatterji, S. (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA*, *291*(21), 2581–2590. https://doi.org/10.1001/jama.291.21.2581

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205. https://doi.org/10.1177/1094428104263675

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

IOM. (2001). *Crossing the quality chasm: A new health system for the 21st century*. National Academies Press (US). http://www.ncbi.nlm.nih.gov/books/NBK222274/

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., Üstün, T. B., & Wang, P. S. (2009). The global burden of mental disorders: An update from the WHO World Mental Health (WMH) Surveys. *Epidemiologia e Psichiatria Sociale*, *18*(1), 23–33. https://doi.org/10.1017/s1121189x00001421

Kilbourne, A. M., Beck, K., Spaeth-Rublee, B., Ramanuj, P., O'Brien, R. W., Tomoyasu, N., & Pincus, H. A. (2018). Measuring and improving the quality of mental health care: A global perspective. *World Psychiatry*, *17*(1), 30–38. https://doi.org/10.1002/wps.20482

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Publications.

Lucke, J. F. (2005). The alpha and the omega of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement*, *29*(1), 65–81. https://doi.org/10.1177/0146621604270882

Miglietta, E., Belessiotis-Richards, C., Ruggeri, M., & Priebe, S. (2018). Scales for assessing patient satisfaction with mental health care: A systematic review. *Journal of Psychiatric Research*, *100*, 33–46. https://doi.org/10.1016/j.jpsychires.2018.02.014

NQB. (2012). *NHS patient experience framework*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215159/dh_132788.pdf

Osborn, C., Reeves, R., Howell, E., & Magee, H. (2004). *Development and pilot testing of the questionnaire for use in NHS Trust-based mental health service user survey*. Picker Institute Europe.

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, *54*(2), 315–323. https://doi.org/10.1348/000711001159582

Reise, S. P. (2012). Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*(1), 27–48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

Ruggeri, M. (1994). Patients' and relatives' satisfaction with psychiatric services: The state of the art of its measurement. *Social Psychiatry and Psychiatric Epidemiology*, *29*(5), 212–227. https://doi.org/10.1007/BF00796380

Sanchez-Balcells, S., Callarisa Roca, M., Rodriguez-Zunino, N., Puig-Llobet, M., Lluch-Canut, M.-T., & Roldan-Merino, J. F. (2018). Psychometric properties of instruments measuring quality and satisfaction in mental health: A systematic review. *Journal of Advanced Nursing*, *74*(11), 2497–2510. https://doi.org/10.1111/jan.13813

Schröder, A., Larsson, B., Ahlström, G., & Lundqvist, L. (2010). Psychometric properties of the instrument quality in psychiatric care and descriptions of quality of care among in-patients. *International Journal of Health Care Quality Assurance*, *23*(6), 554–570. https://doi.org/10.1108/09526861011060924

Shen, G. C., & Snowden, L. R. (2014). Institutionalization of deinstitutionalization: A cross-national analysis of mental health system reform. *International Journal of Mental Health Systems*, *8*(1), 47https://doi.org/https://doi.org/10.1186/1752-4458-8-47

Stamboglis, N., & Jacobs, R. (2020). Factors Associated with patient satisfaction of community mental health services: A multilevel approach. *Community Mental Health Journal*, *56*(1), 50–64. https://doi.org/10.1007/s10597-019-00449-x

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680. https://doi.org/10.1126/science.103.2684.677

Vigo, D., Thornicroft, G., & Atun, R. (2016). Estimating the true global burden of mental illness. *The Lancet. Psychiatry*, *3*(2), 171–178. https://doi.org/10.1016/S2215-0366(15)00505-2

Webb, Y., Clifford, P., Fowler, V., Morgan, C., & Hanson, M. (2000). Comparing patients' experience of mental health services in England: A five-trust survey. *International Journal of Health Care Quality Assurance Incorporating Leadership in Health Services*, *13*(6–7), 273–281. https://doi.org/10.1108/09526860010373253

WHO. (2003). *Quality improvement for mental health*. World Health Organization. https://www.who.int/mental_health/policy/services/essentialpackage1v8/en/

WHO. (2008). *mhGAP: Mental Health Gap Action Programme: Scaling up care for mental, neurological and substance use disorders*. World Health Organization. https://apps.who.int/iris/handle/10665/43809

Williams, M. (2019). *Levels of measurement and statistical analyses* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/c5278

Woodward, S., Berry, K., & Bucci, S. (2017). A systematic review of factors associated with service user satisfaction with psychiatric inpatient services. *Journal of Psychiatric Research*, *92*, 81–93. https://doi.org/10.1016/j.jpsychires.2017.03.020