Questionnaires on Family Satisfaction in the Adult ICU: A Systematic Review Including Psychometric Properties*

Janneke M. van den Broek, MD1,2; Anja H. Brunsveld-Reinders, MSc1; Aglaia M. E. E. Zedlitz, PhD3; Armand R. J. Girbes, MD, PhD2; Evert de Jonge, MD, PhD1; M. Sesmu Arbous, MD, PhD1

**Objectives:** To perform a systematic review of the literature to determine which questionnaires are currently available to measure family satisfaction with care on the ICU and to provide an overview of their quality by evaluating their psychometric properties.

**Data Sources:** We searched PubMed, Embase, The Cochrane Library, Web of Science, PsycINFO, and CINAHL from inception to October 30, 2013.

**Study Selection:** Experimental and observational research articles reporting on questionnaires on family satisfaction and/or needs in the ICU were included. Two reviewers determined eligibility.

**Data Extraction:** Design, application mode, language, and the number of studies of the tools were registered. With this information, the tools were globally categorized according to validity and reliability: level I (well-established quality), II (approaching well-established quality), III (promising quality), or IV (unconfirmed quality). The quality of the highest level (I) tools was assessed by further examination of the psychometric properties and sample size of the studies.

**Data Synthesis:** The search detected 3,655 references, from which 135 articles were included. We found 27 different tools that assessed overall or circumscribed aspects of family satisfaction with ICU care. Only four questionnaires were categorized as level I: the Critical Care Family Needs Inventory, the Society of Critical Care Medicine Family Needs Assessment, the Critical Care Family Satisfaction Survey, and the Family Satisfaction in the Intensive Care Unit. Studies on these questionnaires were of good sample size (n ≥ 100) and showed adequate data on face/content validity and internal consistency. Studies on the Critical Care Family Needs Inventory, the Family Satisfaction in the Intensive Care Unit also contained sufficient data on inter-rater/test-retest reliability, responsiveness, and feasibility. In general, data on measures of central tendency and sensitivity to change were scarce.

**Conclusions:** Of all the questionnaires found, the Critical Care Family Needs Inventory and the Family Satisfaction in the Intensive Care Unit were the most reliable and valid in relation to their psychometric properties. However, a universal “best questionnaire” is indefinable because it depends on the specific goal, context, and population used in the inquiry. (Crit Care Med 2015; 43:1731–1744)

**Key Words:** family satisfaction; intensive care; psychometric properties; quality of care; questionnaire

In recent years, quality of care has become a central issue in healthcare systems worldwide. Particularly, the quality as perceived by patients and their family is a current focus of interest. It is generally accepted that improvement in the quality of care involves a wide range of strategies including the use of evidence-based health care, guidelines and protocols, quality improvement cycles, and changes in safety and risk management (1). Essential in each of these strategies is the monitoring and evaluation of delivered care. In the ICU, satisfaction with the care provided is considered just one of the many quality of care indicators and an important tool for improving care (2–4). Because most ICU patients cannot make decisions themselves, family members are actively involved in the care process as surrogate decision-makers and are, therefore, judges of care quality. However, family satisfaction with care is complex and is not clearly defined.

In the current body of literature, different aspects of family satisfaction are considered important for family members but...
no gold standard currently exists to assess this concept. One line of reasoning is that satisfaction is the fulfillment of family needs or requirements which, if fulfilled, relieve or diminish the distress of the family members or improve their sense of well-being (5). However, Heyland et al (6) remark that although satisfaction reflects the amount of fulfillment of needs and expectations, meeting needs does not guarantee satisfaction. In general, expectations of care, information provided, communication, hospital infrastructure, and patient- and family-related factors all play a role in family satisfaction with ICU care (1). Family satisfaction is also related to the family being provided with clear information because this enables them to actively participate in the decision-making process (6–8).

At present, there are several tools available, mostly questionnaires, that measure family satisfaction with ICU care. Because family satisfaction can be influenced by multiple factors, and the acquired data must be accurate, good validation is obligatory for the adequate use of the questionnaires. Psychometric properties, such as reliability and validity, are essential elements of questionnaires because these describe the quality of the measurement. Questionnaires lacking good psychometric values may not measure the construct they intend to assess, or the values that arise from the questionnaire may not represent the “true” value. This may not only hamper research but also misguide the clinician working with the tool. Thus, the quality of a questionnaire is determined by its psychometric properties.

Therefore, the aim of this review is to determine which questionnaires assessing family satisfaction with ICU care are currently available and to provide an overview of their quality by determining their psychometric properties.

**METHODS**

**Search Strategy and Selection Criteria**

We searched PubMed, Embase, The Cochrane Library, Web of Science, PsycINFO, and CINAHL from inception to October 30, 2013. The databases were searched for medical literature with the following terms: “questionnaires,” “family satisfaction,” “family needs,” and “intensive care.” The complete electronic search strategy can be found in Appendix I (Supplemental Digital Content 1, http://links.lww.com/CCM/B257).

Reference lists of review articles and eligible primary studies were checked to identify cited articles not captured by electronic searches.

**Study Selection**

Included were studies that specifically used a questionnaire to measure family satisfaction and/or family needs in the adult (>18 years) ICU, published in peer-reviewed journals. The language of the articles was restricted to English.

Excluded were studies that did not use a questionnaire to measure family satisfaction. Also excluded were reviews, editorials, and letters to the editor. Furthermore, studies on instruments for medical staff satisfaction and patient satisfaction were excluded as were studies on parent satisfaction in pediatric or neonatal ICU. The latter was done because the specific parent-patient relationship in children less than 18 years old differs from the family-patient relationship in adults (9). Family was defined as next of kin or other persons with a close relationship to an ICU patient.

Two reviewers (J.B. and A.B.) scrutinized the titles and abstracts of all references on possible inclusion. Second, final inclusion/exclusion decisions were made after independent examination of the full articles. All studies that on full-text examination failed to meet the inclusion criteria were excluded. Disagreement between reviewers was resolved by consensus, and if necessary, judgment of a third author was decisive. Reference manager 12.0 (Thomson ISI ResearchSoft, Philadelphia, PA) was used to manage all search results.

**Extracted Data**

The following data were systematically extracted from the studies: author/research group, year of publication, timeframe and means of collecting information, name and version of the tool used, language of the tool, number of questions, and domains (subscales) in the tool. And furthermore, information on sample size and psychometric properties was extracted (see below).

**Quality Assessment**

A two-step model was used to assess the quality of the tools and the psychometric properties.

**Assessment of General Quality and Global Psychometric Properties.** To establish the general quality and global psychometric properties (ie, validity and reliability) of the tool, first all available data for each tool were grouped. Subsequently, the classification model adapted from Cohen et al (10) was applied. This model is an analogue to the well-accepted criteria used to establish effectiveness of treatment in systematic reviews (11). At the highest quality level (level 1), what is taken into account is whether (A) a tool is presented by different research groups in different peer-reviewed articles, (B) sufficient detail of the tool is available to allow evaluation and replication (eg, complete item list and means must be published), and (C) substantial data are available regarding validity and reliability (Table 1).

A tool had to fulfill all the criteria of a specific level to be assigned the quality of that level. When the combined research of a tool met all three criteria defined above (A, B, and C) for level I, it was considered “well-established quality” (++). When one of these criteria was not met, but a tool did meet the standards for level II quality described in Table 1, it was classified as “approaching well-established quality” (+). When one or more of these level II standards were not met, the tool was evaluated with respect to the criteria of level III, “promising quality” (+/−). Finally, when the tool did not meet one or more of the criteria of level III, it was considered level IV, of “unconfirmed quality” (−).

In category C, “++” was scored when validity and reliability were named precisely and when values presented showed good validity (ie, the values were proven to assess the intended construct, or Cronbach α was > 0.70 for all factors) and good
### TABLE 1. Categories for Classification of Instruments Based on Cohen and Modified by Authors

<table>
<thead>
<tr>
<th>Level of Quality</th>
<th>Criteria for Categories</th>
<th>Quality Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A. The measure must have been presented in at least two peer-reviewed articles by different investigators or investigatory teams (+++)</td>
<td>Well-established quality</td>
</tr>
<tr>
<td></td>
<td>B. Sufficient detail about the measure to allow critical evaluation and replication, e.g., complete description of the items and scoring of the tool (+++)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. Detailed information indicating good validity and reliability in at least one peer-reviewed article (+++)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>A. The measure must have been presented in at least two peer-reviewed articles, which might be by the same investigator or investigatory team (+)</td>
<td>Approaching well-established quality</td>
</tr>
<tr>
<td></td>
<td>B. Sufficient detail about the measure to allow critical evaluation and replication, e.g., the domains and subscales of the tool have been described (+)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. Validity and reliability information either presented in vague terms or only moderate values presented (+)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>A. The measure must have been presented in at least one peer-reviewed article (+/-)</td>
<td>Promising quality</td>
</tr>
<tr>
<td></td>
<td>B. Sufficient detail about the measure to allow evaluation, e.g., the questionnaire and its purpose have been described, or the questionnaire was presented in another article. (+/-)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. Validity and reliability information presented in vague terms (e.g., no statistics) or low values presented (+/-)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Negative sore in A, B, and/or C (-)</td>
<td>Unconfirmed quality</td>
</tr>
</tbody>
</table>

Validity and reliability were assessed and scored as follows:

- **“++”** in category C was scored when both validity (either face validity, content validity, or construct validity) and reliability (either internal consistency, inter-rater reliability, or test-retest reliability) were named precisely and when values presented showed good validity and reliability (e.g., the values were proven to assess the intended construct, or Cronbach’s α > 0.70 for all factors), and good reliability (Spearman Brown or Split half > 0.8 of scale and subscales both, or ρ < 0.061 or Pearson’s r > 0.8).

- **“+”** in category C was scored when both validity (either face-, content-, or construct-) and reliability (either internal consistency, inter-rater reliability, or test-retest reliability) were named but not precisely defined or when values presented showed moderate validity and reliability (authors suggested that the tool assesses the intended construct, or Cronbach’s α > 0.70 but not for all factors), and reliability (Spearman Brown > 0.8 for either the scale or the subscales, but not both).

- **“+/-”** in category C was scored when either validity (either face-, content-, or construct-) or reliability (either internal consistency, inter-rater reliability, or test-retest reliability) were named but not precisely defined, or when no values were presented, or when low values were presented (Cronbach’s α < 0.70 for all factors), or reliability (Spearman Brown < 0.8).

- **“-”** in category C was scored when validity or reliability were not mentioned or when no data on validity or reliability was reported.

reliability (Spearman Brown or Split half > 0.8 of scale and subscales both, or ρ < 0.061 or Pearson’s r > 0.8). In category C, a “+” was scored when both validity (either face validity, content validity, or construct validity) and reliability (either internal consistency, inter-rater reliability, or test-retest reliability) were named but not precisely defined, or when values presented showed moderate validity (Table 1). In category C, a “+/-” was scored when either validity or reliability were named, but not precisely defined, or when no values were presented or when low values were presented. Finally, in category C, a “-” was scored when validity and reliability were not mentioned or when no data on validity or reliability were reported.

**Assessment of Psychometric Properties.** All studies describing tools that were considered to be of “well-established quality” were entered in the second step of the analysis. The sample size of the studies and the following psychometric properties of the tools were systematically assessed: face-, content-, and construct-validity, reliability, measures of central tendency, sensitivity, responsiveness, and feasibility (12). This was achieved by grouping the data for each version of the tools (e.g., language, reduced, or extended version) and coding each psychometric property as 1) good, 2) mediocre, 3) poor, or 4) having insufficient data to judge the quality of the psychometric properties. Psychometric properties were defined as follows:

- **Sample Size.** An adequate sample size is needed to detect reliable psychometric data, we used an arbitrary n > 100 per (sub)group cutoff as published by Friberg et al (13).

- **Validity.** Validity refers to the extent to which a tool actually measures family satisfaction. Three types of validity were distinguished: face validity, content validity, and construct validity. Face validity refers to the extent to which a tool is subjectively viewed as covering the concept it purports to assess. Interviews with experts and focus groups are often used to determine this. Furthermore, to fulfill this criterion, the purpose of the tool must be explicitly stated because omission might lead to a discrepancy between an intended and actually assessed target (13). Content validity differs from face validity in that it does not refer to what is subjectively measured but to whether the items of a tool indeed include the appropriate information and content (12). Open-ended questions in a tool can increase its content validity by exploring not mentioned information. As the literature on content validity in family satisfaction is still
scarce and both face validity and content validity involve the relationship of questions and their intended content, they were grouped together.

Construct validity is determined by the validity of abstract variables that cannot be directly observed (latent variables). These constructs are assessed by their relationships with other variables (12, 14). Factor analysis or comparisons with other scales that are supposed to assess the same construct are used to investigate the internal structure and validity of domains. Without good construct validity, it is hard to determine what the tool exactly measures. In the area of family satisfaction, this could involve questions regarding the atmosphere of the waiting room, which does not necessarily reflect satisfaction with ICU care. Tools were considered adequate in this domain when they either exhibited clear, defined factors that in turn showed good internal consistency (Cronbach $\alpha > 0.70$) or when their concurrent validity was high. The latter means that a questionnaire shows a high correspondence with another questionnaire when assessing the same construct (Pearson’s $r > 0.70$ or high Cronbach $\alpha$) (13). Construct validity also covers the aspect of correct questionnaire translation into a different language (1). Adequate translation of a questionnaire is an important and time-consuming procedure that aims for “equivalence” with the original (12). Because research of family satisfaction is performed in many different countries, results of the data obtained need to be comparable.

Reliability. Reliability refers to the overall consistency of a tool’s data across time, settings, and people. This is important because without sufficient reliability the scores obtained may not reflect the “true” scores. For example, the questions may refer to interpersonal conduct of the nurses at a given moment. This may be different from nurse to nurse and subsequently from shift to shift. Therefore, this question score may change daily and is dependent on family members’ personal preferences. The following aspects of reliability were investigated: internal consistency, inter-rater reliability, and test-retest reliability. Internal consistency is the extent to which all items of a tool measure the same content. Cronbach $\alpha$, which is a measure of the average correlation of scores from a measure with the scores of all of its items, is the most commonly used unit of internal consistency (12). In general, acceptable Cronbach coefficients for research and clinical purposes are 0.70 and 0.90, respectively (12, 15). Other internal consistency units include Spearman-Brown and split-half reliability. In this study, we predetermined a degree of greater than 0.80 for both units to represent adequate internal consistency.

Inter-rater (interobserver) and test-retest reliability are both concerned with the robustness of the outcomes of a tool when applied by another person (inter-rater) or at another moment (test-retest). A good agreement of a measure between different raters/observers or by the same raters at different moments is typically represented by $\kappa$ statistics ($> 0.60$) (12) or by a high correlation between the two outcomes (Pearson’s $r > 0.80$).

Measures of central tendency such as the mean and $sd$ of the scale and subscales need to be known as they form the basis for comparison (13) and interpretation of scores. Information about the presence or absence of floor and ceiling effects is needed too in this regard. When these effects are present, non-parametric test should be applied. In these cases, the interpretability of high or low scores is limited substantially.

Sensitivity is a related concept. It is the ability of a tool to detect a “true problem case” (resulting in the percentage of dissatisfied family members who are correctly identified as feeling dissatisfied). Specificity, on the other hand, measures the proportion of negatives that are correctly classified as such (satisfied family members correctly identified as such). Floor and ceiling effects greatly compromise sensitivity and specificity because the scores of true problem cases and true negatives then tend to lie close to each other or are even indistinguishable. True sensitivity cannot be determined in the field of family satisfaction because a gold standard is unobtainable.

Responsiveness is the ability of a scale to detect (meaningful) changes over time (16, 17). This is a particularly important asset when a tool is used to measure the effect of an intervention, for example, a hospitality workshop for healthcare workers. To demonstrate this ability, the tool must first have good test-retest reliability because otherwise the changes could be attributed to mere chance. Also in this psychometric domain, ceiling and floor effects have detrimental influences.

Feasibility relates to the ease and timeframe needed to administer and process an instrument (14, 18). In other words, whether it is acceptable and practical in clinical use and scientific practice. In this study, we focused on the mode of administration (eg, interview and questionnaire) and the amount of time needed to apply the tool.

RESULTS

Selected Studies

The search detected 3,655 references, of which 2,354 references were excluded because they were duplicates. Thus, 1,301 records were screened based on title and abstract. Of these 1,301 records, 1,153 articles did not meet the inclusion criteria (ie, the abstract originated from a poster, it was not a peer-reviewed article, the article did not study adult patients or did not report on family satisfaction). Subsequently, 148 full-text articles were assessed for eligibility and 13 more articles were excluded (19–31). Reasons for article exclusion were as follows: studies in which family satisfaction was combined with patient satisfaction (19, 28, 31), studies that measured hospital staff satisfaction (22–27, 29, 30), studies in which satisfaction or needs were not measured (20), and a study on the implementation of a quality indicator bundle (21). In total, we selected 135 studies for this review (4–8, 32–170). A flow diagram of the study is depicted in Figure 1.

Definition

No uniformly used definition of family satisfaction was found. Two main domains were identified; these were “needs met” and “satisfaction with care.” Within these domains, several subdomains were studied.
Description of the Tools

In these 135 studies, 27 different questionnaires were described. Twenty-one were self-reported questionnaires, six were applied by structured interview (Table 2). Nineteen tools were classified as level IV, “unconfirmed quality,” three as level III, “promising quality,” and one as level II, “approaching well-established quality” (10). Four questionnaires were classified as level I, “well-established quality”: the Critical Care Family Needs Inventory (CCFNI), the Society of Critical Care Medicine Family Needs Assessment (SCCMFNA), the Critical Care Family Satisfaction Survey (CCFSS), and the Family Satisfaction in the Intensive Care Unit (FS-ICU). A detailed overview of the quality of each study can be found in Supplemental Table 1 (Supplemental Digital Content 2, http://links.lww.com/CCM/B258).

Analysis of High-Quality (Level I) Questionnaires

The four level I questionnaires found were described in 109 studies (k). The psychometric data most reported were as follows: sample size, face/content validity, and internal consistency. In approximately two thirds of these studies, means and SD were reported. Only few studies reported findings on construct validity (k = 17) (4, 8, 35, 43, 56, 60, 71, 83, 86, 97–99, 102, 111, 115, 120, 141), inter-rater or test-retest reliability (k = 9) (44, 59, 73, 99, 106, 133, 141, 143, 168), measures of central tendency (k = 1) (125), responsiveness (k = 11) (36, 100, 102, 119, 120, 125, 153, 155, 157, 169, 171), and sensitivity (k = 1) (168) (for a detailed overview, see Supplemental Table 2, Supplemental Digital Content 3, http://links.lww.com/CCM/B259) (Table 3).

CCFNI

The CCFNI, developed by Molter (69) and adapted by Leske (5), was the first questionnaire on family satisfaction with ICU care. It consisted of 45 items and measured what the needs of the family were in relation to five domains: 1) information, 2) comfort, 3) proximity, 4) assurance, and 5) support. Questions on these domains had to be answered on a four-point Likert scale. Warren (52) in 1993 added the Needs Met Inventory (NMI), to assess the extent to which the needs were met. The NMI consists of an additional 45 items on a four-point Likert scale.

In total, 60 studies of the CCFNI were identified; describing 18 different versions, in eight different languages (English, French, Swedish, Greek, Dutch, Chinese, Arabic, and Portuguese). Furthermore, 10 varieties of the CCFNI with a total number of questions varying between 14 and 90 items were reported (for a detailed overview, see Supplemental Table 2, Supplemental Digital content 3, http://links.lww.com/CCM/B259). About half of the studies were of adequate sample size (k = 29; n > 100) (32–35, 38, 40, 43, 46, 49, 50, 54, 57, 58, 60, 62, 64–66, 70, 74, 78, 80–82, 92, 109, 114, 119, 125). With regard to the psychometric data, face/content validity was found to be “good” for most versions with 45 or 46 items, and lower for versions with 30 items or less. Internal consistency was reported for 11 CCFNI versions of which eight demonstrated good internal consistency, whereas it was poor for the three remaining ones. Mean and SD were reported for most versions. Last but not least, responsiveness was studied in three versions of which one study (36) reported positive outcomes (Chinese 45-item version). Responsiveness was not substantiated by other studies or in other versions of the CCFNI. The time needed to complete the questionnaire varied from 20 to 60 minutes (for a detailed overview, see Supplemental Table 2, Supplemental Digital Content 3, http://links.lww.com/CCM/B259).

SCCMFNA

The SCCMFNA, first described in 1998 by Johnson et al (141), consisted of 14 items and measures the needs of family members with respect to 1) attitude, 2) communication, 3) comforting skill, and 4) isolation. The response scale is a four-point Likert scale.

Six studies (7, 133, 138, 141, 146, 166) on the SCCMFNA have been published, including three different language versions:
English, French, and Arabic. Five of these studies met the pre-defined sample size criterion (7, 133, 138, 141, 146). In general, face/content validity was found to be “good.” However, poor results were reported for construct validity and internal consistency. No information was found on other psychometric data such as measures of central tendency. Mean and SD of the items, as well as Table 2. General Quality and Global Psychometric Properties of Tools to Assess Family Satisfaction with Intensive Care

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Year</th>
<th>A. Literature</th>
<th>B. Sufficient Details</th>
<th>C. Validity/Reliability</th>
<th>Overall Quality (I, II, III, and IV)</th>
<th>Mode of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical Care Family Needs Inventory (5, 32–82, 90, 92, 109, 114, 116, 119, 124, 125)</td>
<td>1979–2013</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>I</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Critical Care Family Satisfaction Survey (8, 83, 85, 86, 95, 97, 98, 111, 115, 120)</td>
<td>2001–2013</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>I</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Quality of Death and Dying-communication (103, 22)</td>
<td>2004–2007</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>II</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Family members perception of nurses roles (117)</td>
<td>2005</td>
<td>+/-</td>
<td>+</td>
<td>+</td>
<td>III</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Quality of Communication (101)</td>
<td>2006</td>
<td>+/-</td>
<td>+</td>
<td>+/-</td>
<td>III</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Liddle et al (158)</td>
<td>1988</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Dockert et al (108)</td>
<td>1988</td>
<td>+/-</td>
<td>–</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Dixon et al (112)</td>
<td>1997</td>
<td>+/-</td>
<td>+</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Malacrida et al (140)</td>
<td>1998</td>
<td>+/-</td>
<td>+</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Keenan et al (139)</td>
<td>2000</td>
<td>+/-</td>
<td>–</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Roland et al (151)</td>
<td>2001</td>
<td>+/-</td>
<td>–</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Del trout et al (118)</td>
<td>2005</td>
<td>+/-</td>
<td>+</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Kjerulf et al (134)</td>
<td>2005</td>
<td>+/-</td>
<td>–</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Humble et al (144)</td>
<td>2009</td>
<td>+/-</td>
<td>+</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Whitcomb et al (156)</td>
<td>2010</td>
<td>+/-</td>
<td>+</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Cheung et al (89)</td>
<td>2010</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Family needs questionnaire (123)</td>
<td>2010</td>
<td>+/-</td>
<td>–</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Sundararajan et al (126)</td>
<td>2012</td>
<td>+/-</td>
<td>+</td>
<td>–</td>
<td>IV</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Cuthbertson et al (88–107)</td>
<td>2000–2010</td>
<td>++</td>
<td>+</td>
<td>–</td>
<td>IV</td>
<td>Interview*</td>
</tr>
<tr>
<td>Kirchhoff et al (137)</td>
<td>2002</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>IV</td>
<td>Interview*</td>
</tr>
<tr>
<td>Kutash et al (149)</td>
<td>2007</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>IV</td>
<td>Interview*</td>
</tr>
<tr>
<td>Sacco et al (164)</td>
<td>2009</td>
<td>+/-</td>
<td>–</td>
<td>++</td>
<td>IV</td>
<td>Interview*</td>
</tr>
<tr>
<td>Nelson et al (132)</td>
<td>2010</td>
<td>+/-</td>
<td>++</td>
<td>–</td>
<td>IV</td>
<td>Interview*</td>
</tr>
<tr>
<td>Siddiqui et al (128)</td>
<td>2011</td>
<td>+/-</td>
<td>+</td>
<td>–</td>
<td>IV</td>
<td>Interview*</td>
</tr>
</tbody>
</table>

Mode of assessment: *Assessed by structured interview other questionnaires were self-reported.
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Version</th>
<th>Sample Size</th>
<th>Content/Face</th>
<th>Construct Validity, of Scales/Subscales</th>
<th>Internal Consistency</th>
<th>Inter-Rater/Test-Retest</th>
<th>Measures of Central Tendency</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical Care Family Needs Inventory</td>
<td>48-Item French (65, 66)</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>46-Item French and English (58)</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>46-Item English (66, 67, 77, 78)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>45-Item English (5, 32, 45, 47, 51, 53, 55, 57, 59–61, 68, 69, 75, 81, 90, 114, 116, 124)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>90-Item English (+NMI) (39, 52)</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>45-Item Dutch (40, 43, 74)</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>−</td>
<td>0</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>45-Item Chinese (35, 36, 41, 42, 44, 76)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>45-Item Arabic (33, 38, 119)</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>45-Item French (80, 109)</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>45-Item Swedish (82)</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td>0</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>45-Item Greek (125)</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>43-Item English (46)</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>43-Item Portuguese (34, 70, 73, 92)</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>−</td>
<td>−</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>34-Item Spanish (48)</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>30-Item English (49, 50, 62, 63, 72, 79)</td>
<td>+</td>
<td>+/-</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>60-Item English (+NMI) (71)</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>15-Item English (54, 64)</td>
<td>+</td>
<td>+/-</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>14-Item English (37)</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+/-</td>
<td>0</td>
</tr>
<tr>
<td>Society of Critical Care Medicine Family Needs Assessment</td>
<td>14-Item English (140)</td>
<td>+</td>
<td>0</td>
<td>−</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>14-Item French (7, 137, 143)</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>14-Item Arabic (157, 133)</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>−</td>
<td>−</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Critical Care Family Satisfaction Survey</td>
<td>20-Item English (8, 83, 86, 97, 98, 111, 115, 120)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>20-Item Arabic (95)</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>20-Item Swedish (85)</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(Continued)
van den Broek et al

### TABLE 3. (Continued). Sample Size and Psychometric Properties of Well-Established Assessment Tools

<table>
<thead>
<tr>
<th>Instrument Version</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample Size</td>
<td>Construct Validity of Scales/Subscales</td>
</tr>
<tr>
<td>Family Satisfaction in the Intensive Care Unit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37-Item English (102, 110)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>34-Item English (6, 93, 94, 104–106, 113, 130, 146)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34-Item German (91, 96, 99, 131)</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>34-Item Dutch (51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26-Item Modified English (160, 161)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24-Item English (4, 84, 87, 100, 127, 141, 147–149)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>24-Item Hebrew (153)</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>24-Item German (159)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24-Item Greek (150)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24-Item Filipino (158)</td>
<td>−</td>
<td>+</td>
</tr>
</tbody>
</table>

Sample size and psychometric properties of well-established assessment tools. To improve legibility, sensitivity was not included in the table because scarcely and results indicating sensitivity was reported.

Sample size: “+” = n > 100, “−” = n < 100.

Content/face validity: “+” = assesses intended construct, “+/−” = does not match the original tool, but assesses intended construct, and “−” = does not match the original tool and does not assess intended construct.

Construct validity: “+” = after factor analysis Cronbach α > 0.70 for all factors, “+/−” = after factor analysis, most factors Cronbach α > 0.70, but not all, “−” = factor analysis shows poor quality with most factors Cronbach α < 0.70.

Internal consistency: “+” = Cronbach α, Spearman Brown or Split Hal < 0.80 of the scale and subscales both, “+/−” = Cronbach α is > 0.80 for either the scale or the subscales, but not both, “−” = Cronbach α < 0.80 for the scale and the subscales.

Reliability (inter-rater/test-retest): “+” = κ > 0.61 or Pearson’s r > 0.80, “−” = κ < 0.61; Pearson’s r < 0.80, or proven change of scores when filled in by junior versus senior staff.

Measures of central tendency: “+” = mean and SD for all subscores, no ceiling or floor effect, “+/−” = Only mean and so described or no floor and ceiling effects described, “−” = no mean and so described.

Responsiveness: “+” = scale shows to be able to detect differences over time or between before and after measurements, “+/−” = differences in before and after intervention scores were found on some items, but no differences were demonstrated on the whole scale or on a domain, “−” = scale is not able to show differences over time, or between before and after measurements.

“0” = no data available.

as completion time of the questionnaire, were not reported (for a detailed overview, see the Supplemental Table 2, Supplemental Digital Content 3, http://links.lww.com/CCM/B259).

### CCFSS

The CCFSS is a questionnaire specifically designed to measure family satisfaction with intensive care. It was developed in 2001 by Wasser et al (8) and consists of 20 items within five domains: 1) assurance, 2) information, 3) proximity, 4) support, and 5) comfort, answered on a five-point Likert scale.

The CCFSS has been published in 10 studies (8, 83, 85, 86, 95, 97, 98, 111, 115, 120) and in three different languages: English, Arabic, and Swedish. Only studies on the English version were of good sample size (k = 6; n > 100) (8, 83, 97, 98, 111, 115). This version shows “good” validity (face/content and construct). Five studies (8, 86, 95, 98, 120) reported adequate internal consistency, whereas four other studies (83, 97, 111, 115) found it to be poor. The mean and SD have been reported once for the English version only (86), and this version shows mediocre responsiveness. Finally, data on other psychometric data are lacking. Completion time of the questionnaire was not reported (for a detailed overview, see the Supplemental Table 2, Supplemental Digital Content 3, http://links.lww.com/CCM/B259).
**FS-ICU**

The FS-ICU was developed in 2001 by Heyland and Tranmer (106) and assesses two conceptual domains: 1) satisfaction with care and 2) satisfaction with decision making. The items in the questionnaire were derived from the existing literature on patient satisfaction, quality of care near the end of life, the needs of families of critically ill patients and family satisfaction with decision making (106).

Eleven different versions of the FS-ICU have been published in 32 studies (4, 6, 8, 8, 87, 91, 93, 94, 96, 99, 100, 102, 104–106, 110, 113, 121, 127, 130, 131, 143, 152–157, 161, 167–169, 171). These versions contain a different number of questions: initially the questionnaire consisted of 34 multiple choice and three open-ended questions. Dowling et al (102) in 2005 modified the FS-ICU 34 into a version with 37 questions as part of a critical care family assistance improvement programme. Later in 2007 (4), a more concise version with 24 multiple choice questions was developed. All versions have a five-point Likert response scale. Furthermore, the questionnaire was published in the following languages: English, German, Dutch, Hebrew, Greek, and Filipino (for a detailed overview, see the Supplemental Table 2, Supplemental Digital Content 3, http://links.lww.com/CCM/B259).

The majority of the studies had good sample size \((k = 27; n > 100)\) (4, 6, 8, 8, 8, 87, 91, 93, 94, 96, 99, 100, 102, 104, 105, 110, 113, 127, 130, 131, 143, 152–157, 168, 169, 171), and most versions of the FS-ICU questionnaire showed good psychometric quality. Face/content validity was found to be “good.” Only scarce data were found on construct validity \((k = 3)\) (4, 99, 102), showing mediocre quality for the 34-item German version (99) and the 24-item English version (4). Internal consistency was found to be good for most versions, except for the 37-item modified English version where poor construct validity and internal consistency was reported \((k = 1)\) (102). Twelve studies reported on mean and \(\sigma\) (4, 102, 106, 110, 127, 131, 153, 156, 161, 167, 169, 171). In six studies, information on responsiveness was found (100, 153, 155, 157, 169, 171). This was reported mainly for individual items that showed differences in measurements taken before and after the event. The time needed to complete the questionnaire varied from 20 to 30 minutes (for a detailed overview, see the Supplemental Table 2, Supplemental Digital Content 3, http://links.lww.com/CCM/B259).

On the basis of summaries of psychometric properties (Table 3), with focus on sample size, validity and measures of central tendency, we concluded that of the four questionnaires, the CCFNI and the FS-ICU displayed the most extensively researched and best psychometric properties.

**DISCUSSION**

The aim of this review was to determine which questionnaires assessing family satisfaction with ICU care are currently available and to provide an overview of their quality by determining their psychometric properties. Therefore, we critically examined the quality of all known versions of family satisfaction assessment tools in a two-step model. First, we determined the general quality and psychometric properties of the questionnaires. Second, we evaluated the questionnaires with the highest quality with respect to their psychometric properties.

Only four questionnaires could be classified as being of “well-established quality”: the CCFNI, the SCCMFNA, the CCFSS, and the FS-ICU. However, these high-quality instruments consisted of 35 different versions, each with large disparities in psychometric qualities. Of the four, the CCFNI and the FS-ICU displayed the most extensively researched and best psychometric properties; hence, we would recommend these for further use and study. The CCFNI and the FS-ICU differ in many ways. The CCFNI is primarily designed to measure family needs, whereas the FS-ICU focuses on family satisfaction. Although the definition of “family satisfaction with ICU care” is not clearly defined and overlaps with “family needs,” they are not the same. Meeting needs does not necessarily reflect satisfaction (6) Despite this potential drawback of focus on needs, studies on the CCFNI, especially in combination with the NMI, have been of great value for increased understanding of the needs contributing to overall satisfaction with ICU care. These studies also contributed to an increase in (content) validity of other questionnaires, such as the FS-ICU (106).

The FS-ICU assesses satisfaction with decision making, besides satisfaction with care. These two domains are central to overall family satisfaction with ICU care (106). First, satisfaction with care provides data on how families experience general aspects of care. Second, family satisfaction with decision making is a major component because the family is a substitute decision maker for their critically ill family member in a complex healthcare environment. The FS-ICU is available in many languages, but some language versions have not yet been published in peer-reviewed journals (172). Although a lot of data exist on the 10 different versions of the FS-ICU, it should be noted that not all these versions display an overall high quality.

In general, limitations of the tools include insufficient data regarding (1) construct and content validity (2) inter-rater reliability (3), test-retest reliability (4), measures of central tendency (5), responsiveness, and especially sensitivity (6). Because construct validity is the extent to which a tool actually measures what it claims to measure, and content validity refers to whether the questionnaire includes the appropriate information, they both are of great importance, especially in a subjective outcome such as satisfaction. However, many different language versions of the originally high-quality questionnaires are available in which construct and content validity data are lacking. Therefore, these versions cannot be necessarily called “equivalent.” Differences may arise due to inherent semantic differences and cultural differences. For example, the degree of family participation in the decision-making process differs across the world (7).

An example of importance of inter-rater reliability is Damghi’s study (133), using the SCCMFNA. It was found that when the questionnaire was self-completed by highly educated family members, they were significantly less satisfied with the provided care compared with members of less educated families for whom the questionnaire was filled out by the investigator in
In conclusion, at present four well-established questionnaires are available to measure family satisfaction with ICU care. When using these questionnaires in clinical practice or for research activities, it is of importance to be aware of the limitations of each tool. Of these four tools, CCFNI and FS-ICU have the best psychometric properties. The CCFNI measures needs and the FS-ICU measures satisfaction. Finally, in the evaluation of family satisfaction with intensive care, the use of valid instruments is essential to gain proper and high-quality information. This information is necessary as an outcome quality indicator and to better target improvement initiatives in the ICU.

REFERENCES


60. Leske JS: Internal psychometric properties of the Critical Care Family Needs Inventory. *Heart Lung* 1991; 20:236–244


67. Lynn-McHale DJ, Bellinger A: Need satisfaction levels of family members of critically ill patients admitted to a neurosciences critical care unit. *Intensive Care Nurs* 1989; 5:25–37


158. Liddle K: Reaching out to meet the needs of relatives in intensive care units. *Intensive Care Nurs* 1988; 4:146–159


