BMJ Open

WHO standards-based tools to measure service providers' and service users' views on the quality of hospital child care: development and validation in Italy

Marzia Lazzerini ,[1] Ilaria Mariani ,[1] Tereza Rebecca de Melo e Lima,[2] Enrico Felici,[3] Stefano Martelossi,[4] Riccardo Lubrano,[5] Annunziata Lucarelli,[6] Gian Luca Trobia,[7] Paola Cogo,[8] Francesca Peri,[9] Daniela Nisticò,[9] Wilson Milton Were,[10] Valentina Baltag,[10] Moise Muzigaba,[10] Egidio Barbi,[9,11] on behalf of the CHOICE Study Group

For numbered affiliations see end of article.

Correspondence to
Dr Marzia Lazzerini;
marzia.lazzerini@burlo.trieste.it

ABSTRACT

Objectives Evidence showed that, even in high-income countries, children and adolescents may not receive high quality of care (QOC). We describe the development and initial validation, in Italy, of two WHO standards-based questionnaires to conduct an assessment of QOC for children and young adolescents at inpatient level, based on the provider and user perspectives.

Design Multiphase, mixed-methods study.

Setting, participants and methods The two questionnaires were developed in four phases equally conducted for each tool. Phase 1 which included the prioritisation of the WHO Quality Measures according to predefined criteria and the development of the draft questionnaires. In phase 2 content face validation of the draft questionnaires was assessed among both experts and end-users. In phase 3 the optimised questionnaires were field tested to assess acceptability, perceived utility and comprehensiveness (N=163 end-users). In phase 4 intrarater reliability and internal consistency were evaluated (N=170 and N=301 end-users, respectively).

Results The final questionnaires included 150 WHO Quality Measures. Observed face validity was excellent (kappa value of 1). The field test resulted in response rates of 98% and 76% for service users and health providers, respectively. Among respondents, 96.9% service users and 90.4% providers rated the questionnaires as useful, and 86.9% and 93.9%, respectively rated them as comprehensive. Intrarater reliability was good, with Cohen's kappa values exceeding 0.70. Cronbach alpha values ranged from 0.83 to 0.95, indicating excellent internal consistency.

Conclusions Study findings suggest these tools developed have good content and face validity, high acceptability and perceived utility, and good intrarater reliability and internal consistency, and therefore could be used in health facilities in Italy and similar contexts. Priority areas for future research include how tools measuring paediatric QOC can be more effectively used to help health professionals provide the best possible care.

Strengths and limitations of this study

► This study describes the development and validation of tools to assess perceived quality of care from the perspective of service providers and users, based on the 'WHO Standards to Improve the Quality of Care for Children and Young Adolescents at Facility Level'.

► The major strength of the tools is the multiphase approach used for their development, which aimed at assessing different properties of the questionnaires, including: content validity—assessed with the contribution of both experts and end users—face validity, acceptability, perceived utility and comprehensiveness, reliability and internal consistency assessed in volunteers.

► The tools shall be further validated for use in countries other than Italy.

BACKGROUND

Despite reductions in child and adolescent mortality over the last 30 years, the global burden of disease remains immense. In 2019 alone, 7.4 (95% CI 7.2 to 7.9) million children and adolescents died mostly due to preventable or treatable causes.[1] Europe, Norther America and Australia are the regions with the lowest child mortality.[1] Notwithstanding low child mortality, even in high-income country's quality of care (QOC) for children and adolescents is still a challenge in many settings.[2–11]

Evidence suggest that key gaps in the quality of inpatient child healthcare in high-income and upper middle-income countries include inappropriate hospitalisations, medical errors, drugs over-use, inadequate pain management and unsatisfactory patient

experience of care.[2–10] For example, a recent report from the WHO highlights extreme variations in paediatric hospitalisation rates across Europe, ranging from 150 to 550 per thousand population, suggesting inequity in healthcare.[2] Multicountry surveys and systematic reviews[4 5] report antibiotic prescription rates of up to 60%–75% for common paediatric conditions such as fever, upper respiratory tract infections and diarrhoea, driving high healthcare costs and increasing the risk of antibiotic resistance.[6 7] On the other hand, pain prevention and treatment for children continues to be suboptimal, with a need for wider implementation of both pharmacological and non-pharmacological interventions.[8 9] Finally, patient experience of care has been reported as unsatisfactory in several high-income countries.[10]

For adolescents, evidence from high-income, middle-income and low-income countries shows that adolescents experience many barriers to receiving quality healthcare, including related to factors such as low agency, restrictive laws and policies regarding informed consent, judgmental attitude of healthcare providers, unequal access to resource's, health services fragmentation and poor coordination.[11]

Poor QOC impacts individual health outcomes and increases risks and costs for the entire community. The WHO Global strategy for women's, children's and adolescents' health (2016–2030) recognises QOC as a priority for improving the health of children.[12] To operationalise this vision, a framework for paediatric QOC and standards of care were developed between 2015 and 2018.[13] The WHO Framework[13] identified eight domains of QOC grouped under three key dimensions: (1) provision of care; (2) experience of care; and (3) availability of resources (figure 1). In 2018 WHO defined, through an extensive consultation, eight standards for improving the

quality of paediatric and young adolescent (0–15 years) care, articulated in 40 WHO Quality Statements and 520 WHO Quality Measures.[13]

These WHO Standards and Measures have been developed in the best interests of children and young adolescents, to ensure that their particular needs and rights (eg, to family-friendly health facilities and services; child-specific and young adolescent-specific appropriate equipment, etc) are met and their risks for harm are minimised during health service delivery.[13] The WHO Standards should be implemented in healthcare facilities following the 'Plan Do Study Act' cycle, which implies, as a first step, a baseline assessment.[13]

Nevertheless, there is a lack of documented experience on how best to collect data on the Quality Measures defined by WHO,[13] especially in high-income countries. While tools have been developed to collect data on WHO Quality Measures related to maternal and newborn QOC,[14 15] and outpatient and primary care for adolescents,[16] no tool yet exists related to the WHO inpatient paediatric standards. In 2019, drawing on previous research conducted on the WHO Standards,[17–21] we started a multicentre project called CHOICE (Child HOspItal CarE) aiming at implementing the WHO Standards[13] to improve QOC for children and young adolescents in health facilities in Italy and Brazil. This paper describes the development of two WHO-Standards-based questionnaires[13] and their validation in Italy, which were the initial steps of the CHOICE project. These two questionnaires aim at collecting data on priority WHO paediatric Quality Measures from service users and service providers. The process of validation in other countries, as well as the development of a third tool aiming at collecting key measures on 'provision of care' from official hospital records, will be reported separately.
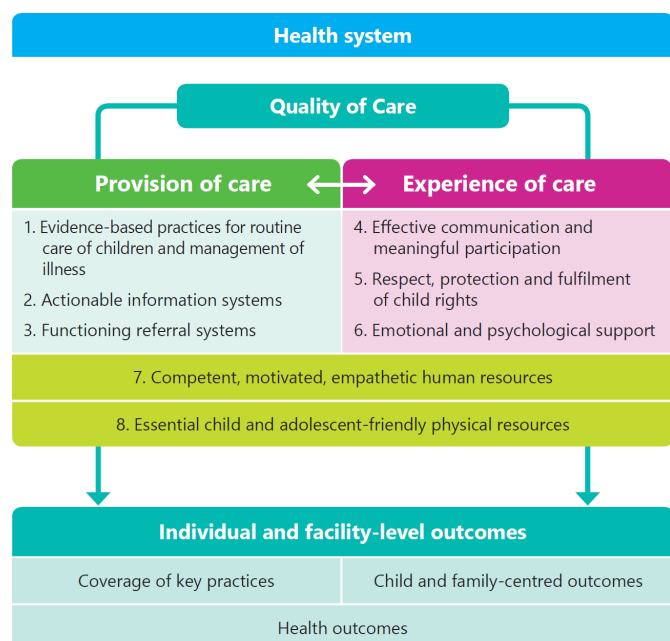

Figure 1 Key phases in questionnaires development.

METHODS

The development of the two CHOICE questionnaires included four subsequent phases, as shown in figure 2, which applied to both questionnaires equally.

The methodology used for the development and validation of the tools was based on existing guidelines,[21–27] examples of questionnaires development reported in literature[14 24 28] and authors' experience in developing similar tools.[16–20]

Table 1 summarises the properties of the questionnaires which were evaluated through the whole process, and the methods used.

Phase 1: development of the draft questionnaire

As a preliminary step, we conducted a literature review to assess whether any other similar tool existed. Relevant experts in the field were consulted. A wide search strategy (online supplemental table 1) was applied to PUBMED, with no language restrictions. A snowballing process was used to identify additional relevant articles for the review using the reference list from primary articles. No other tool was identified therefore the process went on into
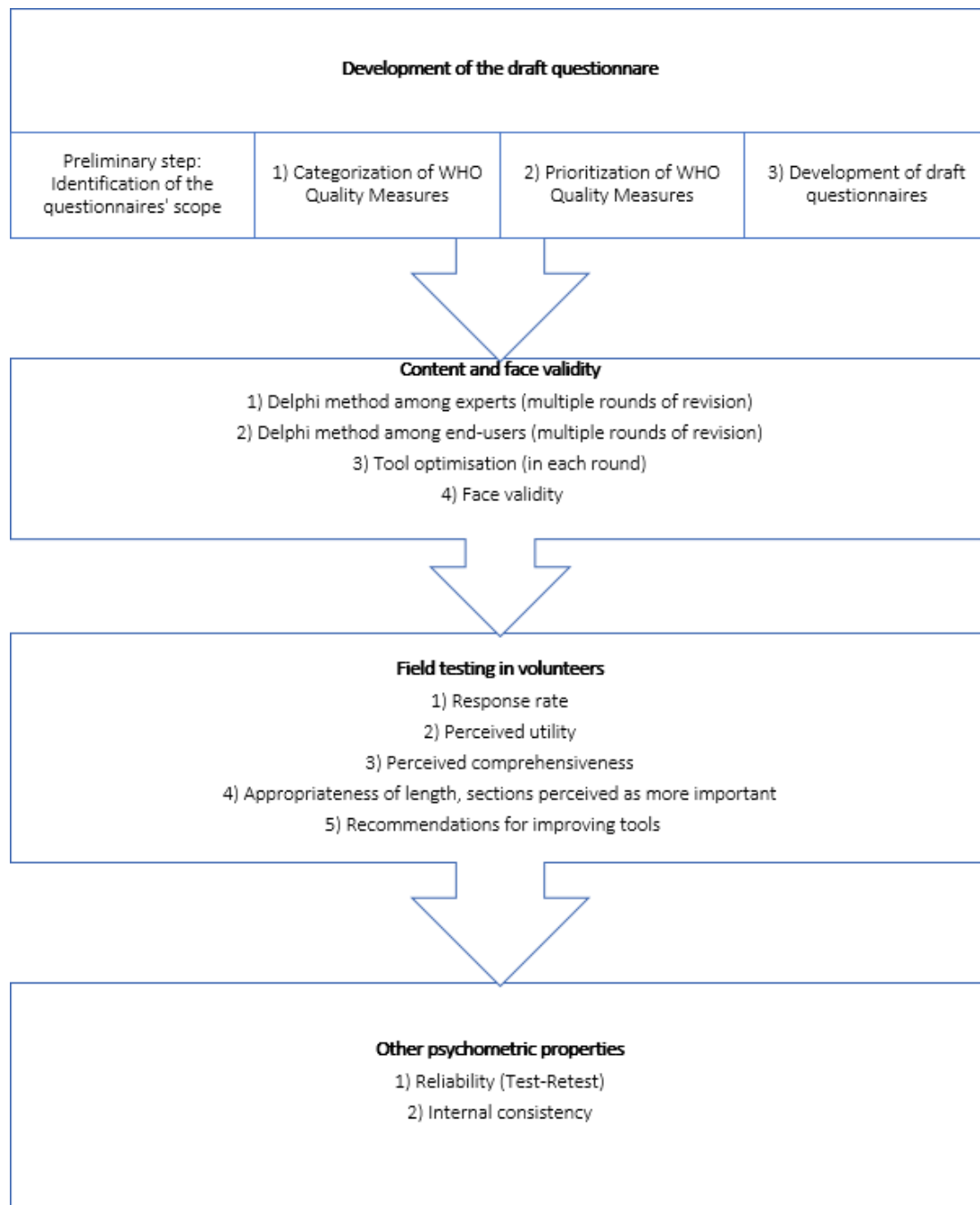
Figure 2 The WHO Framework for improving the quality of paediatric and young adolescent care.[13]

defining the questionnaires' scope and desired characteristics (table 2).

The expected use (table 2) of the two questionnaires was to collect priority indicators useful to improve paediatric QOC as defined by the WHO Quality Measures[13] at facility level in high-income and upper middle-income countries. The focus on this specific setting, as well as the identified data sources (ie, service providers or service users) were considered critical for prioritising the WHO Quality Measures. The two tools were conceived as complementary to a third tool aiming at collecting key measures of provision of care from hospital records. Based on previous experience[16–20] we felt it was important

to include several open questions in the CHOICE questionnaires, allowing for collection of any additional comment on QOC, and questions related to responders' recommendations on how to improve care in their own setting. Criteria for questionnaire structure and wording, were based on existing guidance on how to develop a questionnaire.[22–24]

After these preliminary steps, the following steps which brought to the draft questionnaires were: the categorisation of the WHO Quality Measures, their prioritisation, and their translation into questions in the two draft questionnaires. Specifically, first the WHO Quality Measures for paediatric QOC were categorised based on:

Table 1	Questionnaire property evaluation[24 25]	
Evaluated properties and methods		
Property	**Definition**	**Methods used**
Content validity	The extent to which a questionnaire item includes the most relevant and important aspects of a concept in the context of a given measurement application	Delphi method among experts Delphi method among health service providers and health service users Field test by end-users
Face validity	The ability of an instrument to be understandable and relevant to the targeted population	Formal statistical testing in a sample of volunteers
Acceptability	The degree of acceptability of the tool among responders	Field test by end-users
Reliability over time (intrarater agreement)	Ability of a questionnaire to produce the same results when administered to the same person at two different points in time	Formal statistical testing in a sample of volunteers
Internal consistency	The extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct	Formal statistical testing in a sample of volunteers
Properties not evaluated and reason for exclusion		
Property	**Definition**	**Reason for exclusion**
Diagnostic validity	The accuracy of a questionnaire in diagnosing certain conditions (eg, neuropathic pain)	The questionnaire does not aim to diagnose a specific health condition
Construct validity	The degree to which a tool measures what it claims, or purports, to be measuring	Convergent and divergent validity not possible to assess due to the lack of other validated instruments to measure QOC. Proxy indicators (eg, child mortality) not appropriate as comparison in the Italian setting
Criterion validity	The ability of a questionnaire to predict a final priority outcome (eg, gold standard, reference test to compare with)	Cannot be assessed due to the lack of a final priority outcome or 'gold-standard' to measure QOC
Inter-rater agreement	The degree of agreement among different raters on the QOC	Agreement between different responders is not relevant in a questionnaire which aims at collecting patient individual experience of care

QOC, quality of care.

(1) domain of the WHO Framework[13] they pertained to; (2) most appropriate source of information (ie, health providers, health service users or both).

Second, the WHO Quality Measures for paediatric QOC were prioritised by a team of experts, including paediatricians, adolescent health specialists and researchers involved in developing the WHO paediatric QOC framework. A predefined criteria and scoring system (from a minimum of 0 to a maximum of 5) was used to prioritise Quality Measures: (1) relevance to QOC in the context of high-income to upper middle-income countries in the WHO European Region; (2) feasibility of data collection and expected data reliability and (3) potential utility of the information for use in a quality improvement process. All Quality Measures with a total score of at least 10 points were selected for the first draft of the questionnaire. Sociodemographic items (age, sex, type of diseases for children, type of health professionals, etc) were chosen and designed according to literature and previous

experience.[15 23] Indicators relevant to COVID-19 were extracted from existing WHO guidance and relevant literature available.[29 30]

Finally, the prioritised Quality Measures were translated into questions in the two draft questionnaires, following existing guidance (clear, specific, direct, concise questions) and previous experience[22–24] (online supplemental table 2).

Phase 2: assessment of content and face validity

The two draft questionnaires were submitted to both volunteered experts and end-users to assess content validity (table 1). Opinion of end-users, and not only of experts, was considered particularly important, based on the fact that the two questioners aimed at collecting information on their perceived QOC. Two rounds of revision, including both experts and end-users were conducted.

The team of experts who reviewed both questionnaires included 49 paediatricians involved in the CHOICE

Table 2 Expected used and desired characteristics of the CHOICE questionnaires

Expected use	Collect data useful to improve the QOC for children and young adolescents at facility level in high-income and upper middle-income countries
Phenomena of interest	QOC as perceived by service users and service providers, in line with selected key WHO Quality Measures[13]
Responders	1. Service users, defined as parents or other caretakers of hospitalised children aged between 0 and 15 years, or children if with appropriate age to answer the questionnaire 2. Service providers, defined as any healthcare provider routinely assisting children at facility level
Context	Hospitals in high-income and upper middle-income countries
Administration format	Adaptable (self-administered paper-based or online, or interviews), anonymous and voluntary; informed consent required
Other desired characteristics	1. Collecting information on key WHO Quality Measures for paediatric QOC for which service users' and service providers views are appropriate 2. Multi-item instruments including the following dimensions of QOC: experience, resources, organisation of care, COVID-19 3. Complementary to a third tool collecting information on the provision of care (data source: hospital records) 4. Collecting also data on key indicators relevant to the COVID-19 pandemic (additional section) 5. Content informed by end-users (ie, service providers and service users) 6. Content sufficiently comprehensive, retaining acceptability (ie, good response rate) 7. Structured in logical sequence, by WHO domains of QOC[13] and easy to follow; wording of questions based on existing guidance[24] (eg, clear, specific, concise questions) 8. Including open questions to improve paediatric QOC and collect suggestions and additional feedback 9. Good psychometric properties 10. Allowing scoring of QOC with a single quantitative indicator

CHOICE, Child HOspItal CarE; QOC, quality of care.

project, with experience both at tertiary and secondary care and senior experts from different settings (Italy and Brazil) with long-term experience in developing and/or using WHO indicators and standards.[13 15 17–21 31 32] The WHO Standards[13] were made available to experts. The questionnaires were circulated in Italian, thanks to knowledge of the Brazilian expert of this language.

End-users included health professionals and parents of children hospitalised. Volunteers were selected based on responder characteristics as defined in table 2. In each of the two round of revision 30 health professionals with different backgrounds (senior paediatricians, junior paediatricians, residents in paediatrics, nurses, chief nurses), from different countries (Italy and Brazil) and settings (hospital of different level) reviewed the questionnaire for service providers. Similarly, 30 parents of hospitalised children with different conditions, and presenting different characteristics (ie, age, education, parity, nationality), including a subsample of immigrants living in Italy reviewed the questionnaire for service users.

General Delphi process rules were followed[33]: experts and end-users reviewed the questionnaires and provided written feedback through the two rounds. In each round, specific feedback on the following topics were requested: (1) formulation and wording of questions (ie, whether each question was clear, specific to a single measure and sufficiently concise); (2) importance and relevance of every question, including whether any item should be added or dropped; (3) organisation of domains (ie,

division of items in different sections) and (4) overall content and length of the questionnaires. Recommendations for improvement received were discussed within the team of experts until consensus on a final version was reached. The resulting revised version was then assessed for face validity.

Face validity was assessed by asking end-users (ie, parents of hospitalised children and health workers) to evaluate each question in written form, using a dichotomous scale (Yes/No) in terms of: (1) 'relevance' (defined as the ability of a question to address the extent to which findings, if accurate, apply to the setting of interest) and (2) 'appropriateness' (defined as the ability of items' content to describe the intended characteristic of a construct). Face validity was expressed as absolute frequencies, per cent observed agreement and Cohen's kappa (K) statistics. The minimum predefined acceptable value of K based on existing literature[34–41] was 0.70. Responders were selected at random among the population of health professionals and parents of hospitalised children at the Institute for Child Health Burlo Trieste, a large referral maternal and child hospital caring for paediatric cases from the whole national territory. The sample selection aimed at including responders with different backgrounds (ie, for service providers: senior and junior paediatrician registers in paediatrics, senior and junior nurses; for service users: parents of different ages, sexes, education levels and nationalities of children hospitalised due to different conditions).

The questionnaires were optimised based on the steps above.

Phase 3: field testing

The revised questionnaires were field tested among 163 volunteers (130 parents and 33 health workers) to assess: (1) response rate (calculated as number of respondents out of those asked to complete the questionnaire); (2) perceived utility (yes/no); (3) comprehensiveness (yes/no); (4) length appropriateness (right length/too short/too long; (5) sections perceived as more important (all/A/B/only specific items in each section) and (6) any further recommendations for improving the tool (eg, adding or deleting questions, rephrasing, etc).

Phase 4: assessing reliability and internal consistency

The final questionnaires, optimised based on the steps above (online supplemental annex 1 and 2, table 3) were assessed for intrarater reliability over time, by administering the questionnaire twice (test–retest). It was evaluated on multiple-choice questions—excluding sociodemographic items—using the Cohen's lappa (K) statistic and other indexes of agreement (ie, Gwet's AC1, Bennet and Brennan-Prediger coefficients of agreement).[38–41]

Internal consistency was assessed through Cronbach's alpha correlation (alpha), for sections A and B of each questionnaire (online supplemental table 3), where items were meant to be interrelated. For values of Cronbach's alpha greater than or equal to 0.70 internal consistency was considered good.[22 34] Both reliability and internal consistency were assessed in a sample of volunteers from different regions across Italy.

A simple scoring system (online supplemental table 4) was developed based on examples in literature and previous authors' experience.[15 28] In developing the scoring system, the following key considerations were made. First, it was acknowledged that other recent scoring systems developed to describe the QOC[28 41 42] did not attribute different weights to different Quality Measures; in fact, it is difficult if not impossible to quantify the importance of different aspects of care (eg, antibiotic prescriptions vs respectful care), as all of these aspects are equally linked to human rights.[13] Second, literature suggests that a scoring system with values ranging from 0 to 100 is easier to understand (compared with other ranges).[28] Consequently, in the CHOICE scoring system each WHO Quality was given the same weight, with a total score ranging in each domain from 0 to 100, thus allowing easy comparison across domains (eg, resources, experience, etc). This study did not aim at further testing the scoring system.

Data analysis

For face validity, the required minimum sample size, calculated based on exiting guidance,[21 22 34] resulted in 20 service users and 20 healthcare providers, assuming in the null hypothesis a K value of 0.3 and in the alternative hypothesis a K value of at least 0.7, 80% power and a significance level of 2.5% with a one-tailed test. For reliability, assuming in the null hypothesis a K value of 0.45 and in the alternative hypothesis a K value of at least 0.65 (with a proportion of 0.20, 0.3 and 0.5 in the three categories of the item), 80% power and a significance level of 2.5% with a one-tailed test, the required sample resulted in 74 cases for each questionnaire. For the internal consistency, assuming in the null hypothesis an alpha of 0.55, and in the alternative hypothesis an alpha of at least 0.70, 80% power, a number of items equal to 10 (to be conservative), and a significance level of 2.5% with a one-tailed test, a required sample of 108 service users and 108 health professionals was needed.

Summary statistics were presented as absolute frequencies, percentages, and as K statistic and other indexes of agreement (ie, Gwet's AC1, Bennet and Brennan-Prediger coefficients of agreement) and as Cronbach's alpha correlation value, as appropriate. All tests performed were two-tailed and a p value of <0.05 was considered statistically significant. Statistical analyses were performed using Stata V.14 and R V.3.6.1.

Patient and public involvement statement

Service user, selected on a voluntary basis, were involved in the development and validation of the CHOICE questionnaires. They had the opportunity to provide feedback on the health service user questionnaire, and express freely their preferences. Inputs received from patients were used to revise the content of the questionnaire, including reducing its length, and to improve acceptability.

RESULTS

Phase 1: questionnaire draft development

The process of prioritisation of the WHO Quality Measures resulted in the inclusion of 85 Quality Measures in the service user questionnaire, and 80 Quality Measures in the service provider questionnaire, respectively. Considering additional items (ie, questions to assess sociodemographic characteristics of responders, and open questions), these first versions included 100 and 95 total questions respectively. The draft questionnaires were further assessed and optimised in the following phases described below.

Phase 2: content and face validity

The Delphi process among experts optimised several questions, including questions on management of diarrhoea, respiratory infections, fever, pain and organisation of care. A few items were dropped and substituted by other WHO Quality Measures which were deemed more specific, relevant to the context of high-income and middle-income countries, and potentially actionable (eg, availability of clear criteria for hospitalisation for diarrhoea, constant availability of a minimum set of drugs to treat pain in children, non-pharmacological pain prevention). Specific questions required rewording after feedback from end-users.

Since responders recommended to reduce the length of the questionnaires, the total number of included WHO Quality Measures was slightly decreased. Specifically, 10 measures which were repeated in both questionnaires were dropped from the service user questionnaire; 5 measures deemed less relevant by end-users and experts were dropped from the service provider questionnaire. The revised tools included 75 Quality Measures each, for a total of 150 WHO Quality Measures across the two instruments (online supplemental table 3).

Results of the subsequent face validity test are reported in online supplemental table 5. More responders than expected based on the initial sample size calculation contributed to face validity, resulting in a final sample of 30 parents and 20 health providers. For most questions it was impossible to estimate the Cohen's kappa statistics due to the fact that none of the responder considered any question as not relevant nor appropriate, except for a single question in each questionnaire with a resulting kappa value of 1, indicating excellent face validity. Thus, there was no need to further modify the questionnaires.

The final version of the two questionnaires is reported as online supplemental annex 1, 2. The questionnaire for health workers included the following six sections: (A) physical resources for health workers (40 items); (B) organisation of work (25 items); (C) management of COVID-19 pandemic (12 items); (D) overall satisfaction (two questions); a section to collect sociodemographic characteristics of health workers; a final section to collect feedbacks on the perceived utility and acceptability of the questionnaire. Similarly, the questionnaire for health service users included the following six sections: (A) physical resources for children and their care-givers (25 items); (B) experience of care (40 items); (C) management of COVID-19 pandemic (10 items); (D) overall satisfaction (two questions); a section to collect socio-demographic characteristics of health workers; a final section to collect feedbacks on the perceived utility and acceptability of the questionnaire. In each of the two questionnaires, section A, B, C, D included a final open question to collect suggestions from health workers on how to improve the QOC (online supplemental table 3).

Phase 3: field testing

The field testing of the final version of two questionnaires with 163 volunteers resulted in a high response rate (98% for service users, 76% for service providers), among which 96.9% and 90.4%, respectively rated the questionnaires as useful (online supplemental table 6). Overall, 86.9% and 93.9%, respectively rated the questionnaire as comprehensive, with most responders considering all sections of the questionnaire as important (83.1% and 75.8%, respectively).

In the open field for recommendations for improvement we received several messages of appreciation, and only a minor suggestion for revisions. No other changes were therefore needed after field testing.

Phase 4: reliability and internal consistency

Findings on intrarater agreement are reported in online supplemental table 7. We received more answers than expected, resulting in a final sample of 95 parents and 75 service providers, resulting in a power of 89% and 88%, respectively. The value of Cohen's kappa was at least 0.70 for all questions, with the exception of selected cases were the paradox of Cohen's kappa (ie, low kappa values in presence of a high degree of agreement) was observed, due to substantial imbalance in the table's marginal totals.[37 39] All additional indexes of agreement—Gwet's AC1, Bennet index, Brennan-Prediger coefficient—indicated for all items at least a good agreement (Gwet's AC1 >0.60),[40] with the exception of two question with a value of Gwet's AC1 of 0.55 and 0.60, respectively, which were rephrased by the team of experts to improve clarity.

Internal consistency findings are reported in online supplemental table 8. We received more answers than expected, thus resulting in a final sample of 193 parents providing a power 96.4%. The Cronbach's alpha values were 0.84 and 0.83 for sections A and B of the service user questionnaire, respectively, while for the service provider questionnaire the values were 0.95 and 0.85, indicating very high internal consistency for both questionnaires.[21–23]

DISCUSSION

Collecting service users' and service providers' views on paediatric QOC is critical for improving it. This paper presents the first results of the long process of designing, developing and validating two questionnaires which comprehensively collect data on 150 WHO Quality Measures[13] for measuring QOC for children and young adolescents at hospital level. The ultimate objective of these tools is helping department directors and other policy makers understand what works well and what needs to be improved in facilities where children and adolescent receive healthcare. The availability of a unified comprehensive approach to measuring QOC for children at facility level as defined by the WHO Standards[13] and using validated tools will allow comparisons of data across settings and over time and enhance efforts to improve paediatric QOC.

We believe that the process we adopted had several strengths. It included multiple phases, based on existing recommendations on health questionnaire development,[22–26] and on guidance to evaluate patients' experience of care.[27 28] The initial questionnaires were optimised through a sequence of logical steps, which included several rounds of revisions after feedback from international experts and end-users, field testing, and formal statistical assessment of the relevant psychometric properties of the tools. Other questionnaires previously developed and used in recent large surveys did not go through all these steps.[43] Interestingly, a recent systematic review emphasised the lack of clear, scientifically sound recommendations on methods to validate patient-reported outcomes measures.[44]

As a limitation of this study, we acknowledge that the sample size used for validation only included professional and parents from Italy. The questionnaires and the score system are now undergoing additional validations and field-testing in Brazil and in other countries. Results of these ongoing efforts will be reported separately. Another priority area for future research is documenting how these tools can be better used to drive a quality improvement process. In the future the questionnaire may also be further adapted for use in large 'quick' online surveys.

The two questionnaires were intentionally aiming at assessing perceived inpatient QOC for children and young adolescent from services users and service providers perspectives. As such, they may have the limitations of excluding older adolescents, excluding the outpatient and low-income settings, and capturing only perceptions on QOC. We believe that no single tool can fit all purposes while retaining acceptability. Most tools to measure QOC actually use surveys in service users, since this is an important perspective.[14 15 18 28 42 43] Further research is needed to develop tools that cover the population and settings excluded by the two-questionnaire described in this study.

As anticipated in the introduction, to allow triangulation of data from different data sources, we developed a third, complementary tool aiming at collecting key WHO Quality Measures on provision of paediatric care from official hospital patient records. The three tools have been conceived and developed in parallel, aiming at collecting, when used together, 170 WHO Quality Measures on paediatric QOC.[13] Findings of the development and validation of this third tool will be reported separately.

The scoring system should be intended only as a complementary (not substitutive) way to quantitatively measure paediatric QOC in a synthetic format, and should always be interpreted looking at detailed results of the whole list of indicators collected. Properties of the score system shall be evaluated in future studies.

CONCLUSIONS

This study suggests that the two WHO standards-based tools developed have good content and face validity, high acceptability and perceived utility, and good intrarater reliability and internal consistency, and therefore could be used in health facilities in Italy and similar contexts. Priority areas for future research include how tools measuring paediatric QOC can be more effectively used to help health professionals provide the best possible care.

Author affiliations
[1]WHO Collaborating Centre for Maternal and Child Health, Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste, Italy
[2]Instituto de Medicina Integral Professor Fernando Figueira/IMIP, Recife, Brazil
[3]Pediatric and Pediatric Emergency Unit, The Children Hospital, AO SS Antonio e Biagio e Cesare Arrigo, Alessandria, Italy
[4]Pediatric Unit, Ospedale Santa Maria di Ca Foncello, Treviso, Italy
[5]Department of Pediatrics, "La Sapienza" University -Hospital "Santa Maria Goretti" of Latina, Roma, Italy
[6]Pediatric Emergency Department, Giovanni XXIII Pediatric Hospital, University of Bari, Bari, Italy
[7]Pediatric and Pediatric Emergency Room Unit, "Cannizzaro" Emergency Hospital, Catania, Italy
[8]Division of Paediatrics, Department of Medicine DAME, Academic Hospital Santa Maria della Misericordia, University of Udine, Udine, Italy
[9]University of Trieste, Trieste, Italy
[10]Department of Maternal, Newborn, Child and Adolescent Health and Ageing, World Health Organization, Geneva, Switzerland
[11]Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste, Italy

MD (Pediatric and Neonatology Unit, Maternal and Child Health Department, 'La Sapienza' University of Roma – Hospital 'Santa Maria Goretti' of Latina, Roma, Italy), Annunziata Lucarelli MD (Giovanni XXIII Pediatric Hospital, Pediatric Emergency Department, University of Bari, Bari, Italy), Eleonora Canzio MD (Giovanni XXIII Pediatric Hospital, Department of Pediatrics, University of Bari, Bari, Italy), Roberta Parrino MD (Pediatric Emergency Unit, Maternal and Child Department, Arnas Civico, Palermo, Italy), Salvatore Gambino (Pediatric Maternal and Child Department, Arnas Civico, Palermo, Italy), Melania Guardino MD (Department of Neonatology and NICU, University Hospital Policlinico P. Giaccone, Palermo, Italy), Luca Lagalla MD (Department of Sciences for Health Promotion and Mother and Child Care 'G. D'Alessandro', University of Palermo, Via A. Giordano 3, 90127, Palermo, Italy), Beatrice Vaccaro (Pediatric Maternal and Child Department, Arnas Civico, Palermo, Italy), Giuseppina de Rosa (Pediatric Maternal and Child Department, Arnas Civico, Palermo, Italy), Vita Antonella Di Stefano MD (Pediatric and Pediatric Emergency Room Unit, 'Cannizzaro' Emergency Hospital – Catania, Italy), Francesca Patané MD (Pediatric Postgraduate Residency Programme, Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy).

ORCID iDs
Marzia Lazzerini http://orcid.org/0000-0001-8608-2198
Ilaria Mariani http://orcid.org/0000-0001-8260-4788

REFERENCES

1 UNICEF, WHO, World Bank, UN-DESA Population Division. *Levels & trends in child mortality report 2020. Estimates developed by the U Inter-Agency Group for Child Mortality Estimation*. Geneva: United Nations Children's Fund, 2020. https://www.unicef.org/reports/levels-and-trends-child-mortality-report-2020

2 World Health Organization. *Situation of child and adolescent health in Europe*. World Health Organization, 2018.

3 World Health Organization, OECD & International Bank for Reconstruction and Development. *Delivering quality health services: a global imperative for universal health coverage*. World Health Organization, 2018. https://apps.who.int/iris/handle/10665/272465

4 Van de Maat J, van de Voort E, Mintegi S, *et al*. Research in European pediatric emergency medicine Study Group. antibiotic prescription for febrile children in European emergency departments: a cross-sectional, observational study. *Lancet Infect Dis* 2019;19:382–39.

5 Clavenna A, Bonati M. Differences in antibiotic prescribing in paediatric outpatients. *Arch Dis Child* 2011;96:590–5.

6 Piovani D, Clavenna A, Sequi M, *et al*. Reducing the costs of paediatric antibiotic prescribing in the community by implementing guideline recommendations. *J Clin Pharm Ther* 2013;38:373–8.

7 European Centre for Disease Prevention and Control. Antimicrobial resistance in the EU/EEA (EARS-Net) annual epidemiological report for 2019. Available: https://www.ecdc.europa.eu/en/publications-data/surveillance-antimicrobial-resistance-europe-2019 [Accessed 5 Mar 2020].

8 Williams S, Keogh S, Douglas C. Improving paediatric pain management in the emergency department: an integrative literature review. *Int J Nurs Stud* 2019;94:9–20.

9 Krauss BS, Calligaris L, Green SM, *et al*. Current concepts in management of pain in children in the emergency department. *Lancet* 2016;387:83–92.

10 National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Health Care Services; Board on Global Health; Committee on Improving the Quality of Health Care Globally. *Crossing the global quality chasm: improving health care worldwide*. Washington, DC: National Academies Press, 2018.

11 World Health Organization. *Global accelerated action for the health of adolescents (AA-HA!): guidance to support country implementation*. Geneva: World Health Organization, 2017. http://apps.who.int/iris/bitstream/10665/255415/1/9789241512343-eng.pdf?ua=1

12 World Health Organization. *Every woman every child. Global strategy for women's, children's and adolescents health 2016–2030*. Geneva: World Health Organization, 2016. http://www.who.int/life-course/partners/global-strategy/global-strategy-2016–2030/en/

13 World Health Organization. *WHO standards to improve the quality of care for children and young adolescents at facility level*. Geneva: World Health Organization, 2018.

14 Bohren MA, Vogel JP, Fawole B, *et al*. Methodological development of tools to measure how women are treated during facility-based childbirth in four countries: labor observation and community survey. *BMC Med Res Methodol* 2018;18:132.

15 Lazzerini M, Argentini G, Mariani I, *et al*. A WHO standards-based tool to measure women's views on the quality of care around the time of childbirth at facility level in the WHO European Region: development and validation in Italy. *BMJ Open* 2022;12:e048195.

16 WHO/UNAIDS. Global standards for quality health-care services for adolescents. *Tools to conduct quality and coverage measurement surveys to collect data aboutcompliance with the global standards*. Geneva: World Health Organization, 2015. https://apps.who.int/iris/bitstream/handle/10665/183935/9789241549332_vol3_eng.pdf?sequence=5

17 Lazzerini M, Shukurova V, Davletbaeva M, *et al*. Improving the quality of hospital care for children by supportive supervision: a cluster randomized trial, Kyrgyzstan. *Bull World Health Organ* 2017;95:397–407.

18 Lazzerini M, Valente EP, Covi B, *et al*. Use of WHO standards to improve quality of maternal and newborn hospital care: a study collecting both mothers' and staff perspective in a tertiary care hospital in Italy. *BMJ Open Qual* 2019;8:e000525.

19 Lazzerini M, Mariani I, Semenzato C, *et al*. Association between maternal satisfaction and other indicators of quality of care at childbirth: a cross-sectional study based on the WHO standards. *BMJ Open* 2020;10:e037063.

20 Lazzerini M, Semenzato C, Kaur J, *et al*. Women's suggestions on how to improve the quality of maternal and newborn hospital care: a qualitative study in Italy using the WHO standards as framework for the analysis. *BMC Pregnancy Childbirth* 2020;20:200.

21 Mohamed R, Fahmy FF, Senanayake H. Correlation among experience of person-centered maternity care, provision of care and women's satisfaction: cross sectional study in Colombo, Sri Lanka. *PLoS One* 2020;16:e0249265.

22 Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. 5th edn. Oxford University Press, 2014.

23 Terwee CB, Bot SDM, de Boer MR, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.

24 Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J Anaesth* 2017;11:80–9.

25 Taherdoost H. Validity and reliability of the research instrument; how to test the validation of a Questionnaire/Survey in a research. *Int J Acad Res Manag* 2016;5:hal-02546799.

26 Oluwatayo JA. Validity and reliability issues in educational research. *J Educat Soc Res* 2012;2.

27 Larson E, Sharma J, Bohren MA, *et al*. When the patient is the expert: measuring patient experience and satisfaction with care. *Bull World Health Organ* 2019;97:563–9.

28 Afulani PA, Phillips B, Aborigo RA, *et al*. Person-centred maternity care in low-income and middle-income countries: analysis of data from Kenya, Ghana, and India. *Lancet Glob Health* 2019;7:e96–109.

29 World Health Organization. *Maintaining essential health services: operational guidance for the COVID-19 context*. Geneva: World Health Organization, 2020.

30 World Health Organization. *COVID-19 strategic preparedness and response plan operational planning guidelines to support country preparedness and response*. Geneva: World Health Organization, 2020.

31 World Health Organisation Regional Office for Europe. *Hospital care for children: quality assessment and improvement tool*. Copenhagen: World Health Organisation Regional Office for Europe, 2015.

32 Pessa Valente E, Barbone F, de Melo E Lima TR, *et al*. Quality of maternal and newborn hospital care in Brazil: a quality improvement cycle using the who assessment and quality tool. *Int J Qual Health Care* 2021;33:mzab028.

33 McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm* 2016;38:655–62.

34 Anthoine E, Moret L, Regnault A, *et al*. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes* 2014;12:176.

35 Taber KS. The use of cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ* 2018;48:1273–96.

36 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159.

37 Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543–9.

38 Randolph JJ. Free-Marginal Multirater Kappa (multirater K[free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa [Internet]. Online Submission 2005;14.

39 Zec S, Soriani N, Comoretto R, *et al*. High agreement and high prevalence: the paradox of Cohen's kappa. *Open Nurs J* 2017;11:211–8.

40 Wongpakaran N, Wongpakaran T, Wedding D, *et al*. A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol* 2013;13:61.

41 Vedam S, Stoll K, Rubashkin N, *et al*. The mothers on respect (MOR) index: measuring quality, safety, and human rights in childbirth. *SSM Popul Health* 2017;3:201–10.

42 Vedam S, Stoll K, Martin K, *et al*. The mother's autonomy in decision making (MADM) scale: Patient-led development and psychometric testing of a new instrument to evaluate experience of maternity care. *PLoS One* 2017;12:e0171804.

43 Bohren MA, Vogel JP, Fawole B, *et al*. Methodological development of tools to measure how women are treated during facility-based childbirth in four countries: labor observation and community survey. *BMC Med Res Methodol* 2018;18:132.

44 Sawyer A, Ayers S, Abbott J, *et al*. Measures of satisfaction with care during labour and birth: a comparative review. *BMC Pregnancy Childbirth* 2013;13:108.