BMJ Open

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

Marzia Lazzerini ,[1] Giorgia Argentini,[1] Ilaria Mariani ,[1] Benedetta Covi,[1] Chiara Semenzato,[1] Ornella Lincetto,[2] Moise Muzigaba,[2] Emanuelle Pessa Valente [1]

[1]WHO Collaborating Centre for Maternal and Child Health, Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste, Italy
[2]Department of Maternal, Newborn, Child and Adolescent Health and Ageing, World Health Organization, Geneve, Switzerland

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

ABSTRACT

Objectives A comprehensive WHO standards-based tool to measure women's perceived quality of maternal and newborn care (QMNC) in health facilities is needed to allow for comparisons of data across settings and over time. This paper describes the development of such a tool, and its validation in Italy.

Design A multiphase, mixed-methods study involving qualitative and quantitative research methods.

Setting Nine health facilities in Italy.

Methods The questionnaire was developed in six phases: (1) Defining the scope, characteristics and a potential list of measures for the tool; (2) initial content and construct validation; (3) first field testing to assess acceptability and perceived utility for end-users (1244 women, 35 decision makers) and further explore construct validity; (4) content optimisation and score development; (5) assessment of face validity, intrarater reliability and internal consistency and (6) second field testing in nine maternity hospitals (4295 women, 78 decision-makers).

Results The final version of the tool included 116 questions accounting for 99 out of the 350 of the extended lists of WHO Quality Measures. Observed face validity was very good, with 100% agreement for 101 (87%) questions and Kappa exceeding 0.60 for remaining ones. Reliability was good, with either high agreement or Kappa exceeding 0.60 for all items. Cronbach alpha values ranged from 0.84 to 0.88, indicating very good internal consistency. Acceptability across seven hospitals was good (mean response rate: 57.4%, 95% CI 44.4% to 70.5%). The questionnaire proved to be useful, driving the development of actions plan to improve the QMNC in each facility.

Conclusions Study findings suggest that the tool has good content, construct, face validity, intrarater reliability and internal consistency, while being acceptable and useful. Therefore, it could be used in health facilities in Italy and similar context. More research should investigate how effectively use the tool in different countries for improving the QMNC.

Strengths and limitations of this study

► This is the first report on the development and validation of a tool based on WHO Standards to measure women's views of the quality of maternal and newborn care (QMNC) around the childbirth at health facility in the WHO European region.

► The tool collects the perceptions of key service users of maternal and newborn health services (ie, mothers) and allow for comparisons of data across settings and over time.

► The major strength of the tool is the multiphase approach used for its development, including an extensive literature review (total 32 754 papers screened), a Delphi process with international experts and mothers, a formal assessment of psychometric properties, and two rounds of field testing (total 5539 mothers, 113 decision makers) to evaluate acceptability and utility in the real-world setting.

► The scoring system should be considered and used as a complementary (not substitutive) synthetic measure of women's perceptions of the QMNC, and should always be interpreted jointly with detailed results of all Quality Measures collected with the questionnaire.

BACKGROUND

Despite progress in the reduction of maternal and newborn mortality, recent estimates show that about 300 000 women still die every year worldwide during or following pregnancy,[1] nearly 2 million babies are born dead,[2] and 2.4 million die within the first month of life.[3] Most of these deaths are due to preventable causes.[1–3]

In high-income countries, where maternal and newborn mortality remain generally low, gaps still exist in the quality of maternal and newborn care (QMNC). Large national surveys and reports from high-resource

countries such as the UK, USA and Italy,[4–9] have high-lighted gaps in adherence to evidenced-based practices. Overuse of medical interventions such as caesarean sections, episiotomy and induction of labour, low use of appropriate procedures for pain control, and low rates of early and exclusive breastfeeding are just some examples of practices which could be improved.[4–9] Furthermore, as reported by systematic reviews[10 11] and highlighted by a statement of the WHO, women worldwide continue to experience disrespect and abuse during childbirth at health facilities, even in high-income countries.[12] Lack of privacy, ineffective communication, verbal and emotional abuse (eg, use of harsh or blaming language), and lack of informed consent have been documented in studies in high-income countries such as the UK,[13] Canada[14] and Italy.[15 16] These practices violate women's rights, deter women from seeking healthcare and may have major implications for the mental and physical health of mothers and newborns.[11 12 17]

In the past few years, WHO has placed increasing emphasis on the importance of improving the quality of healthcare for all mothers and children worldwide. In 2015, WHO developed a framework for improving the QMNC[17] that identifies eight domains of quality of care grouped under three key dimensions: (1) provision of care (evidence-based practices, efficient information and referral systems); (2) experience of care (effective communication, respect, dignity and emotional support) and (3) availability of resources (motivated human resources and physical resources). In 2016, based on the above described eight domains of QMNC, WHO defined eight standards for improving maternal and newborn care in health facilities, articulated in 31 'Quality Statements' and 350 total 'Quality Measures'.[18] The WHO standards define what healthcare planners, managers and care providers should ensure in order to guarantee high QMNC. These standards should be implemented in healthcare facilities following the 'Plan Do Study Act' cycle, which implies a baseline assessment to identify priorities for action.[19]

So far there is a lack of evidence on how to better collect data on the Quality Measures defined by WHO.[18] A recent review highlighted that existing facility assessment tools do not cover at least 25% of the input and process WHO Quality Measures, lacking many measures of the experience of care.[20] Two WHO standard-based tools[18] to record how women were treated during facility-based childbirth were recently developed,[21 22] however, prioritisation of the WHO Quality Measures for these tools focused specifically on the context of low-resources settings, thus making them not applicable in high-income countries. Other existing tools either focused on specific aspects of QMNC- such as respectful care[23 24] or maternal autonomy in decision making[25]—or were not developed based on the WHO Framework and Standards.[17 18] There is still a need to develop and test instruments and methods to collect women's views on QMNC in high-income and middle-income countries. A unified comprehensive

approach to measure QMNC as defined by the WHO standards, through validated tools, would allow for comparisons of data across settings and over time, and enhance efforts to address gaps in QMNC.[18 20 21]

More than one tool is needed to collect data on all domains of the WHO standards.[18] There is common agreement[21–25] that the WHO Quality Measures[18] related to experience of care should be collected from service users (ie, mothers), while those related to many other aspects—such as availability of physical resources, actual training received, existence of quality improvement mechanisms—may be obtained from care providers. Direct observation of all WHO Quality Measures may have limited sustainability outside research settings.[21]

In 2017, following a pilot project,[16] we started the IMAgiNE (Improving MAternal Newborn CarE) project in Italy, aimed at implementing the WHO Maternal and Newborn Standards[18] in all maternity hospitals of the Friuli Venezia Giulia Region in North-east Italy. The project included the development and validation of two tools to collect data on priority WHO Quality Measures relevant for the WHO European region setting, using two key respondent populations: service users (ie, mothers) and service providers (ie, health workers). This paper describes the process of developing and validating the first of these tools, the IMAgiNE questionnaire for mothers. The description of the development and validation of the second tool, the IMAgiNE questionnaire for health workers, will be the object of a separate report.

METHODS

The development of the IMAgiNE questionnaire for mothers was carried out in six successive and complementary phases (figure 1), using both qualitative and quantitative research methods.

Phase 1: questionnaire's scope, characteristics, indicators (April to July 2016)

To verify whether any other tools based on the WHO Standards[18] existed, and to help define the scope, characteristics, structure, and criteria for prioritising indicators for the research tool, in July 2016, we conducted a review of the literature on existing and similar tools which have been used in high-income to middle-income settings to measure perceived QMNC. We used a wide search strategy (online supplemental table 1) and applied it to PubMed, with no language restrictions. Conference proceedings and references of retrieved articles were also reviewed manually.

Based on this process, a core group of researchers identified the desired characteristics for the IMAgiNE questionnaire for mothers. Existing recommendations on how to develop health-related questionnaires,[26–33] and other relevant literature on how to assess patients' experiences and person-centred measures[21–25 34–37] were also used to identify additional characteristics of the IMAgiNE questionnaire for mothers, as well the possible initial structure.

- Phase 1. Questionnaire's scope, characteristics, indicators
  a) Literature review
  b) Define the questionnaire's desired characteristics
  - c) Prioritize the WHO Quality Measures

Result: Draft questionnaire

-156 questions (three open questions)

Phase 2. Initial content/construct validation
  a) Delphi among experts
  b) Delphi among mothers

Result: Version for large field testing
- 120 questions (three open questions)
- 94 (26.8%) WHO Quality Measures

Phase 3. First field testing and additional construct validation
  a) Acceptability by end-users
  b) Utility (use of data by decision makers)
  c) Multivariate analysis and thematic analyses

Result: Revised questionnaire
- 117 questions (11 open questions)
- 97 (27.7%) WHO Quality Measures

Phase 4. Final content optimisation and score development
  a) Updated literature review
  b) Development of a simple score system

Result: optimised questionnaire with score system
- 116 questions (11 open questions)
- 99 (28.2%) WHO Quality Measures
- Score system by domains, range 0-100

Phase 5. Statistical validations
  a) Face validity
  b) Intra-rater reliability over time
  c) Internal consistency

Result: questionnaire validated

Phase 6. Second field testing in 9 maternity hospitals
  a) Acceptability by end-users
  b) Utility (use of data by decision makers)
  c) Exploratory factor analysis
  d) Properties of the score system
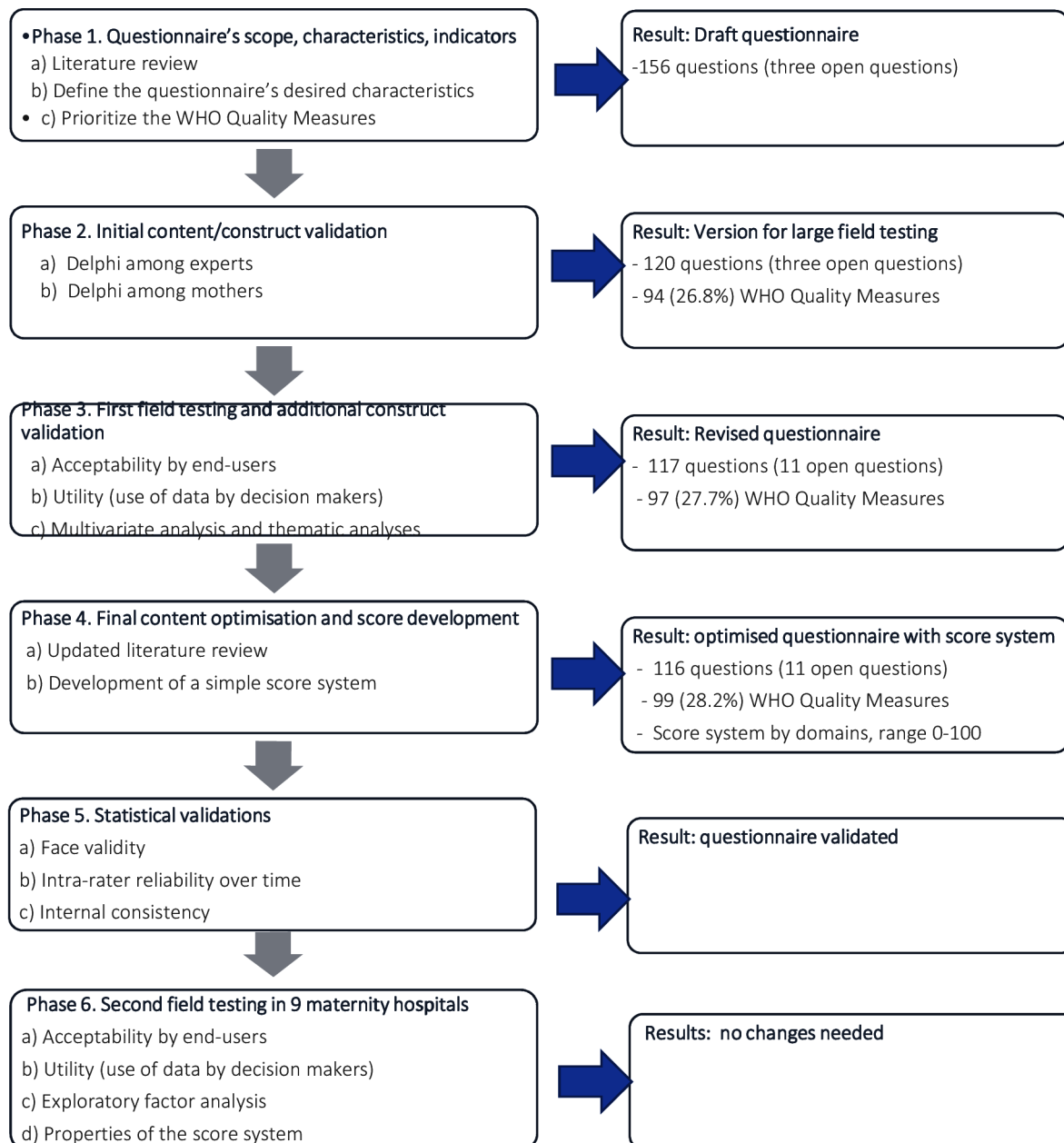
Results: no changes needed

Figure 1 Keys steps in the development of the IMAgiNE questionnaire for mothers. IMAgiNE, Improving MAternal Newborn CarE.

Subsequently, the 350 WHO Quality Measures were tabulated in an Excel file according to the WHO Standard[18] they pertain to. The Quality Measures were then rated according to the following predefined criteria: (1) measures for which it was appropriate to use mothers as a source of information; (2) importance to QMNC in the context of high-income to middle-income countries with low maternal and newborn mortality in the WHO European region; (3) potential utility of the information for use in a quality improvement process; (4) balance among the different domains of the WHO Framework (ie, provision of care, experience, availability of resources).[18] This process was undertaken by a team of 25 people, including both experts with different backgrounds (obstetricians, midwives, epidemiologists, neonatologists, nurses)

and end-users (volunteering mothers). The final list of priority measures (n=99) was agreed through four rounds of discussion. Items collecting sociodemographic characteristics of mothers (eg, age, education, country of birth) were defined based om existing example identified with the systematic review. In translating WHO Quality Measures into questions for mothers, we followed recommendations on questionnaire writing.[26–28 30]

Phase 2: initial content and construct validation (July to November 2016)

This phase aimed at assessing and improving the content and construct validity of the first version of the questionnaire. Content validity and initial construct validity were evaluated through a Delphi process with two groups of

participants. The first group included 26 senior experts in QMNC (obstetricians, midwives, neonatologists, nurses, epidemiologists, psychologists and social scientists) from different settings (Italy, UK, Ireland, Brazil) and with extensive experience in developing tools and conducting assessment of QMNC in many countries of WHO European region.[38–40] The second group consisted of 29 volunteer mothers, with a recent birth, selected among different networks (postpartum groups, private networks), from different settings (Italy, UK, Ireland, Brazil), and with different characteristics (ie, age, education, parity).

General Delphi process rules were followed.[33][34] Initially, experts and mothers first reviewed the questionnaire and provided written feedback. We organised three rounds of feedback revisions; in each round we requested revisions and recommendations for improvement in the following areas: (1) formulation and wording of questions (we asked 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 questionnaire. They then participated in group discussions at the end of which consensus was reached.

Phase 3: first testing and additional construct validation (December 2016 to November 2018)

The version of the questionnaire from phase 2 was field tested from December 2016 to September 2018 at the Institute for Maternal and Child Health IRCCS Burlo Garofolo, a public tertiary level university hospital in Northeast Italy where between 1300 and 1500 births occur every year. Field testing was aimed at exploring: (1) acceptability (response rate) as a self-administered tool and (2) utility of the tool in collecting and promoting the use of data to drive the development of a local action plan to improve the QMNC.

Additional construct validation was also undertaken, by multivariate and thematic analysis. Detailed methods of these analyses have been already reported elsewhere.[41][42] Briefly, we looked through multivariate analyses at the association between overall maternal satisfaction with the care received and another 61 key variables, including measures of 'provision of care', 'experience of care', 'availability of resources' and other maternal characteristics.[41] In addition, a thematic analysis of free-text comments provided by mothers was conducted to identify any emerging theme worth adding to the questionnaire.[42]

Phase 4: final content optimisation and score development (December 2018 to May 2019)

We updated to October 2019 the literature review previously conducted in phase 1, to further refine the content and construct validity of the IMAgiNE questionnaire and to analyse score systems previously used to measure QMNC. We used the same search strategy as in the first literature review (online supplemental table 1). We selected key questionnaires more relevant to the WHO Standards and including a score system, and tabulated their items in an Excel file. We compared each item with our questionnaire, and scored each item of our questionnaire with the same criteria as for Phase 1, using a score from 1 to 5 on each item. We also compared the score systems used.

Phase 5: statistical validation (June 2019 to November 2020)

The methodology that we used for formal statistical validation of the psychometric properties of the tool followed the existing recommendations on testing psychometric properties of health questionnaires.[26–33]

We assessed face validity by asking end-users (ie, mothers) to provide written feed-back on each question using a dichotomous scale (yes/no), on: (1) the 'relevance' (defined as the property of a question to address the extent to which findings, if accurate, apply to the setting of interest) and (2) 'appropriateness' (defined as the property of items' content to describe the intended characteristic of a construct). Face validity was expressed as an absolute frequency, percentage of observed agreement, and Cohen's Kappa (K) statistic. The minimum predefined acceptable value of K based on existing literature[26–32] was 0.60. The required sample size for this evaluation was calculated based on exiting guidance,[26–28] and resulted in 20 mothers, selected as representative of the average population of Italian mothers in terms of age, parity, education and social environment.

Intrarater reliability over time was tested administering the questionnaire twice 1 week apart to volunteer women (test–retest). Intrarater reliability was assessed for all multiple-choice questions, with the exception of those containing only sociodemographic information and, for practicality, in those answered only by a subsample of women (ie, by mode of birth). Intrarater reliability was evaluated with the Cohen's Kappa (K) statistic, and was considered acceptable for K values equal or higher than 0.60.[26–33] The sample size was calculated based on exiting guidance[26–28] assuming in the null hypothesis a K value of 0.50 and in the alternative hypothesis a K value of at least 0.60, 80% power and a significance level of 2.5% with one-tailed tests,[25][39] resulting in a required sample of 80 mothers

Internal consistency was assessed through Cronbach's alpha correlation (alpha). The sample size was calculated based on exiting guidance[26–28] assuming in the null hypothesis an alpha of 0.60, and in the alternative hypothesis an alpha of at least 0.75, 80% power, a number of items equal to 20, and a significance level of 2.5% with one-tailed tests, resulting in 77 women. The alpha was calculated only on sections where items were meant to be interrelated, that is, sections F (experience of care) and G (satisfaction). The cut-off values for Cronbach's alpha significance were predefined as follows: for values greater than or equal to 0.75 internal consistency was considered good; for values below 0.75, changing or dropping

Table 1 Questionnaire's properties evaluated and reasons for not considered some specific properties[29 30]

Property evaluated and methods

Property	Definition	Methods used
Content validity	The extent to which a questionnaire measure includes the most relevant and important aspects of a concept in the context of a given measurement application	Literature review Delphi among experts Delphi among mother
Construct validity	The degree to which items/scores on the questionnaire measure relate to other measures (eg, patient reported or clinical indicators) in a manner that is consistent with theoretically derived a priori hypotheses concerning the concepts that are being measured	Delphi among experts Delphi among mother Multivariate analysis Thematic analysis Factorial analysis
Face validity	The ability of an instrument to be understandable and relevant to the targeted population	Test in volunteer mothers
Reliability over time (intrarater agreement)	Ability of a questionnaire to produce the same results when administered to the same person at two different points of time	Test in volunteer mothers
Internal consistency	The extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct	Test in volunteer mothers
Acceptability	The degree of acceptability by end-user, including both those who respond to the questionnaire (response rate) and those who receive the data (decision makers)	Two field tests
Utility	Use of data by decision makers	Two field tests

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
Criterion validity	The ability of a questionnaire to predict a final priority outcome (gold standard, reference test to compare with)	Cannot be assessed due to the lack of a 'gold-standard' to measure the QMNC
Inter-rater agreement	The degree of agreement among different raters	Agreement between different responders is not relevant in a questionnaire which aims at collecting patient individual experience of care
Responsiveness	The ability of a questionnaire to detect clinically important changes over time	Will be evaluated in future studies

QMNC, quality of maternal and newborn care.

questions and re-evaluating the internal consistency on a new sample of participants would be required.[26–32]

Compared with expected properties of a health status questionnaire,[29] due to the nature of the instrument, some specific aspects of validations were considered not appropriate (table 1).

Phase 6: second field testing in nine maternity hospitals (November 2019 to May 2021)

During this phase, the questionnaire was deployed in nine maternity hospitals in the Friuli Venezia Giulia Region, in North-east Italy. This field test aimed at evaluating again acceptability and utility as well at evaluating properties of the score system.

In addition, we further assessed construct validity by performing factor analysis on data collected from November 2019 to May 2021. Exploratory factor analysis was performed on items which were answered by all women and contributing to the scoring system. We reverse coded negative items and estimated polychoric correlation matrix. To identify the number of factors to extract, we used both Kaiser's rule, that is, all eigenvalues over one should be retained and the 'break' in the screeplot, that is, the plot of eigenvalues. Factor loadings >0.30 were used as a cut-off to identify items to maintain in the questionnaire. Since components of the scoring system are intended to be related, oblique rotation was carried out as it allows correlation among factors.

Participants in the survey and in the statistical validation were informed about the objectives and methods of the study, including their rights in declining participation, and signed an informed consent before responding to the questionnaires. Anonymity in data collection during the survey phase was ensured by not collecting any information that could disclose participants' identity.

Patient and public involvement statement

Mothers, selected on a voluntary basis among women with a recent pregnancy, were involved in the development and validation of the questionnaire. Women had the opportunity to provide feedback on the questionnaire, and express freely their priorities, experiences and preferences on the content of the questionnaire, including their views on its acceptability and specific written feedback on how to improve it. Inputs received from mothers were used to revise the questionnaire.

RESULTS

Phase 1: questionnaire's scope, characteristics, indicators

The literature review resulted in 28 664 articles. Among these, we identified 49 papers (online supplemental table 2) which described the design, validation and use of questionnaires aimed at collecting information on QMNC in high-income or middle-income countries. These studies were carried out in several countries in the WHO European region—including France, Hungary, Germany, Italy, the Netherlands, Norway, Slovenia, Slovakia, Slovenia, Spain, Sweden, Turkey, UK—as well as other countries high-income or middle-income countries. None described a tool to measure women's views on QMNC based on the WHO Standards.[18] Among retrieved reports, we selected

those related to questionnaires used in large projects such as national or multicountry surveys. We reviewed selected questionnaires' characteristics and structures, how items were phrased, and how the questionnaires were administered (eg, interviews vs self-compilation). In addition, for the most relevant questionnaires identified (ie, those used for national surveys), we tabulated each question in an Excel file and conducted a thematic analysis using the WHO Standards[18] and the Respectful Maternity Care Chart[36] as frameworks to assess questionnaire content and structure.

Table 2 reports the identified desired characteristics for the IMAgiNE questionnaire for mothers. The predefined aim of the questionnaire was to collect data useful to improve QMNC, as defined by the WHO Quality Measures,[18] at facility level in high-income and middle-income countries in the WHO European region. The applicability on the context of high-income to middle-income countries in the WHO European region was considered a key criterion for prioritising the WHO Quality Measures for inclusion in the tool, in respect to the other existing questionnaire focusing to low-resources settings.[21 22] The target population was women giving birth at facility level, with very limited exclusion criteria (table 2). The time period of interest was the period

Table 2	Preidentified desired characteristics of the IMAgiNE questionnaire for mothers
Expected use	Collect data useful to improve the QMNC at facility level in high-income and middle-income countries in WHO European region
Phenomena of interest	QMNC as defined by selected key WHO Quality Measures
Target population	Women giving birth at facility level
Exclusion criteria (reason)	1. Mothers under 18 years of age (informed consent regulations). 2. Underlying psychiatric conditions (validity). 3. Maternal death (unable to respond). 4. Refusal to participate.
Context	High-income and middle-income countries in the WHO European region
Time period of interest	Childbirth, from arrival at the hospital to discharge
Administration format	Mixed and adaptable (self-administered paper based or online, face to face or phone interviews), anonymous and voluntary
Other properties	1. Multi-item instrument including different dimensions of QMNC around the time of childbirth. 2. Collecting information of the key WHO Quality Measures for which the users' views is appropriate and important. 3. Complementary to a second tool (HW questionnaire). 4. Content heavily informed by end users (ie, mothers and health professionals). 5. Sufficiently comprehensive without overloading the responder (ie, acceptable response rate). 6. Structure in logical sequence and easy to follow. 7. Wording of questions based on existing guidance[29 30] (eg, clear, specific, concise questions). 8. Containing a mix of closed and open-ended questions. 9. Collecting also women's recommendations to improve the QMNC and additional feedbacks (open questions). 10. Good psychometric properties. 11. Proved utility in initiating a quality improvement process. 12. Allowing scoring of QMNC with a single quantitative indicator.

IMAgiNE, Improving MAternal Newborn CarE; QMNC, quality of maternal and newborn care.

around the time of childbirth. It was critical to recognise that no one questionnaire can be 100% comprehensive in terms of Quality Measures included, without overloading responders, and that an acceptable compromise between comprehensiveness and acceptability was needed. Additional desired characteristics of the questionnaire are reported in table 2.

The ranking of the WHO Quality Measures and two rounds of expert group discussions resulted in a draft questionnaire that included a total of 156 questions (among which three open questions), divided in eight sections.

Phase 2: initial content and construct validation

Several key recommendations emerged from the questionnaire review conducted by experts and mothers. One crucial aspect was the need to reduce the length of the questionnaire in order to improve acceptability to mothers. Furthermore, specific items also required rewording, removal or additions. For example, some questions regarding disrespect and mistreatment required reformulation to improve clarity and directness, while items related to a specific aspect of care (eg, informed consent, non-evidence-based interventions, additional key aspects of case management during delivery) were added on request of either mothers or professionals. Overall, 49 questions were removed in this phase, while 13 new questions were added. The structure of the questionnaire did not change. The resulting questionnaire included 120 questions, with three open-ended questions.

Phase 3: first field testing and additional construct validation

In this first testing, a total of 1244 women answered the questionnaire, with a response rate—measured as a percentage of all women given birth in the hospital—of 52%. The sample included women with a different age, parity and educational, employment, marital status, as described in table 3. A total of 189 (15.2%) mothers were foreign women not born in Italy. No relevant differences between characteristics of responders and non-responders were observed (table 3).

Survey findings were presented to key stakeholders and then in a 1-day final workshop with 35 key decision-makers (ie, hospital general director, executive director, scientific director, chiefs of department, chief nurses) in November 2018. Details of this process have been reported elsewhere.[16] Briefly, an action plan with 55 specific recommendations on how to improve QMNC was agreed as a result of this assessment, proving data utility in driving the initiation of a quality improvement process, as expected with use of the WHO Standards.[18] This field test on data use suggested that the questionnaire had a good balance among quality measures pertaining to different domains (eg, provision of care vs experience of care vs availability of resources). It also suggested that several questions should be revised, as they were not perceived as completely clear, while others could be dropped, since the quality measures they explored were not perceived

Table 3 Characteristics of the sample included in the first field testing (N=1244)

	n	%
Age		
<35 years old	735	59.1
≥35 years old	509	40.9
High education (college or above)		
No	582	46.8
Yes	655	52.7
Born in Italy		
Yes	1051	84.5
No	189	15.2
Citizenship		
Italian	1124	90.3
Not Italian	111	8.9
Occupational status		
Employed	987	79.3
Non employed	251	20.2
Marital status		
Single/other	52	4.2
Married/living with a partner	1184	95.2
Parity		
Primiparous	658	52.9
Multiparous	586	41.1
Multiple pregnancy		
Yes	21	1.7
No	1223	98.3

as essential by professionals. Finally, a synthetic score system—as a complementary (not substitutive) way to quantitatively measure QMNC in a synthetic way- was perceived as necessary by stakeholders, to compare survey results over time and across settings and to simplify reporting.

Detailed results of the multivariate and thematic analyses have been reported separately.[41 42] Briefly, these analyses suggested that many variables are strongly associated with women's satisfaction with QMNC during childbirth, supporting the use of multiple indicators to monitor the QMNC, rather than few selected indicators.[41] Thematic analysis highlighted the critical importance of retaining open questions in the questionnaire as an opportunity to collect additional inputs and recommendation from mothers.[42] Because recommendations from mothers tend to vary substantially by individual[42 43] as well by setting, it was decided not to add any predefined multiple-choice question, but rather including an open question at the end of each section of the questionnaire.

Based on these findings, further steps to optimise the questionnaire included: (1) dropping questions that were not shown to add valuable information (eg, similar topics or items that could be dropped for perceived lack

of relevance); (2) revising questions which were not clear, specific or concise enough and (3) optimising and adding open questions. The resulting questionnaire included 117 questions, 11 of which were open ended.

Phase 4: final content optimisation and score development

The literature update resulted in 4090 new articles retrieved. Out of these, none described the development of a tool to be used in high-income to middle-income settings to collect data on mother's views of QMNC as defined by the WHO Standards.[18] However, we identified 12 new papers on tools used to measure maternal perception of QMNC (online supplemental table 2) and a new tool based on the WHO Standard and conceived for use in low resources settings.[22 23] We selected the key questionnaires more relevant to the WHO Standards and including a score system, tabulated their items in an Excel file, and compared each item with our questionnaire.

Following this revision, few questions were rephrased (eg, transformed into more direct questions), and few were substituted with questions that seemed more relevant and useful. The resulting questionnaire included 116 questions, 11 of which were open ended, for a total of 99 (28.2%) WHO Quality Measures (online supplemental table 4). The WHO Quality Measures included for each WHO standards are further detailed in online supplemental table 5.

The comparison of score systems used across key selected tools (online supplemental table 6) highlighted several important aspects. First, items of previously existing scoring system[21–25] were not weighted by importance. In fact, it is difficult if not impossible to attribute a different quantitative coefficient of importance to different aspects of care (eg, caesarean section rate vs respectful care), as all of these aspects are equally linked to human rights and relevant to achieve positive health outcomes.[11] Second, previous scoring systems[21–25] were not validated against a single priority outcome. This is because a single priority outcome against which to compare measures does not exist. Mortality/morbidity outcomes are not appropriate since the literature clearly shows that mistreatment of women can occur even in settings with low maternal and/or newborn mortality.[10–12] On the other hand, maternal satisfaction, a measure often chosen as a single comprehensive indicator of women's experiences, has shown to poorly reflect many other quality measures of QMNC around childbirth and is heavily affected by the outcome of birth.[35 43] Third, existing literature suggests that a scoring system with values ranging from 0 to 100 is the most acceptable and useful, being easy to understand compared with scoring systems with other ranges.[22 23]

Based on this literature review and on existing guidance on scale development,[26–29 44] we developed a simple scoring system ranging from 0 to 100 in each domain of the WHO Framework:[18] provision of care, experience of care, and availability of resources (online supplemental table 7). The scoring system was intended as a complementary quantitative synthetic measure suggesting to some degree the overall QMNC, which should be interpreted in conjunction with other indicators.

Phase 5: statistical validation

For each of the 116 total questions, the percentage of women identifying the question as both relevant and appropriate was at least 75%, with most questions having values in between 85% and 100%. Specifically, 101 (87%) questions had a 100% level of agreement. For these questions it was impossible to estimate the Cohen's Kappa statistics due to complete agreement. For the remaining questions, the Kappa value always exceeded the predefined cut-off of 0.60, indicating good agreement. Detailed results of the analysis are presented in online supplemental table 8.

The observed intrarater agreement ranged from 70.0% to 100%, with 13 items reaching 100% agreement (online supplemental table 9). The value of Cohen's Kappa was at least 0.60, except in a few cases with high agreement and substantial imbalance in the table's marginal totals either vertically or horizontally, were the paradox of Cohen's Kappa[45] was observed, thus justifying low Kappa values even in case of high agreement.

Internal consistency, assessed through Cronbach's alpha for sections F and G, ranged from 0.84 to 0.88 (online supplemental table 10).

Phase 6: second field testing in nine maternity hospitals

As of May 2021, 4295 women had responded the questionnaire, with a mean response rate across facilities—measured in each facility as the percentage of all women given birth in the hospital—of 50.2% (95% CI 49.2% to 51.3%). The sample included women with a different age, parity and educational, employment, marital status (table 4). A total of 966 (22.4%) mothers were foreign women not born in Italy, and of these 577 (60.5%) were born in other countries of the WHO European region. Missing values ranged from 0% to 2.4% across all sections (online supplemental table 11).

Preliminary survey findings were presented to 79 key decision-makers (chiefs of newborn departments, chief of obstetric departments, chief nurses, chief midwives, key staff and other staff involved in QMNC) in individual workshops during November and December 2020. This process resulted in the development of an action plan to improve QMNC in each facility. Details on this process and on the findings of the survey will be reported in a separate paper. No critical feedback was made on the questionnaire.

The exploratory factor analysis identified four factors that satisfied both Kaiser's rule and screeplot 'break' criterion (online supplemental figure 1) and accounted for 55% of the total variance. The four-factor solution (online supplemental table 12) reflected the nature of the score system with the first factors confirming that the questionnaire explored the domains of the WHO framework with the first factor including the majority of itens theoretically referring to experience of care domain, the

Table 4 Characteristics of the sample of the second field testing in nine maternity hospitals (N=4295)

	n	%
Age (median, range)	33 (18–55)	–
Foreign mothers	966	22.4
Countries of provenience of foreign mothers		
WHO European region	557	60.5
Africa	126	13.7
Asia	65	7.1
South America	48	5.2
North America	41	4.5
Central America	27	2.9
Others	16	1.7
Primiparous	2283	53.2
Multiple pregnancy	64	1.5
Full-time employ	2337	54.4
High education (university degree or above)	1948	45.4
Baby in intensive care unit	223	5.2
Type of delivery		
Spontaneous	3086	71.9
Operative	351	8.2
Caesarean section	858	20.0

second representing availability of resources domain, and the third including only provision of care items. The fourth factor included a mixture of resources and experience items referring to companionship. Items referring to availability of resources loaded reasonably well with items related to experience of care and vice versa confirming interrelation among QMNC domains. Thus, items were retained in specific domains according to the WHO theoretical framework and no changes needed to the questionnaire.

The scoring system was shown to have the following properties: (1) easy to understand; (2) captured variability across domains; (3) captured variability across hospitals; (4) sensitive to change (change in single items translated to a change of the final score). There was agreement among stakeholders that the score should be intended as a complementary (not substitutive) way to quantitatively measure QMNC in a synthetic format, and should be interpreted looking at detailed results of other indicators.

DISCUSSION

Collecting women's views on the QMNC is critical for improving it. As highlighted by previous systematic reviews[20] and by our literature review (online supplemental table 2), existing assessment tools to measure QMNC in high-income to middle-income countries

were not explicitly designed based on the WHO framework and WHO Standards.[18] The availability of a unified comprehensive approach to measure QMNC as defined by the WHO Quality Measures, through a validated tool, allow comparisons of data across settings and over time and enhance efforts to improve quality of care.[20 21]

This paper presents the results of the long process of designing, developing and validating a questionnaire based on the WHO Standards[18] for measuring women's perception of QMNC around childbirth in high-income to middle-income countries. We believe that this process had several strengths. The process included multiple phases, incorporating the existing recommendations on the methodology on how to develop a health questionnaire[26 33] and assess patients' experience and person-centred measures,[34 35] in addition to an extensive literature review (total 32 754 papers screened) on other existing similar questionnaires previously developed. The initial questionnaire was optimised through a sequence of logical steps, which included several rounds of revisions after recommendations from international experts and end-users (mothers) two rounds of field-testing evaluating acceptability and utility in the real-world setting (total 5539 mothers, 113 decision-makers), as well as formal statistical assessment of the relevant psychometric properties of the tool. Other questionnaires previously developed and used in large recent surveys did not go through all these steps.[21 22] Interestingly, a systematic review emphasised the lack of clear scientifically sound recommendations on methods to validate patient-reported outcomes measures.[28] Similarly, we could not identify reports of data utility, showing whether indicators collected were actually used for reporting to decision makers (too many indicators may be difficult to communicate), whether all indicators where perceived as valuable, and whether eventually they initiated a process of quality improvement. We believe that the evaluation we made of data use in a real-world setting was fundamental for optimising the questionnaire design.

As a limitation of the tool, we would like to acknowledge that the scoring system should be intended as a complementary (not substitutive) way to quantitatively measure QMNC in a synthetic format, and should always be interpreted looking at detailed results of the whole list of indicators collected. The questionnaire is now undergoing of additional validations and field-testing in 16 countries in the WHO European Region, and is being translated in 23 languages. Results of these ongoing efforts will be reported separately. Responsiveness of the score shall be evaluated in future studies. In the future the questionnaire may also be adapted for use in large 'quick' online surveys.

This article does not describe the process of development and validation of a second complementary tool, the IMAgiNE questionnaire for health workers, aimed at collecting other key WHO Quality Measures, relevant to health providers as source of data (eg, availability of resources, organisation of care). The two tools have

been conceived and developed in parallel, aiming at collecting, when used together, a substantial percentage (around 50%) of the WHO Quality Measures.[18] Findings of the development and validation of second tool will be reported separately.

The ultimate objective of the tool described in this paper is to help department directors and policy makers understand what works well and what needs to be changed or improved in the health facilities where women deliver and babies are born to ensure QMNC. In the future, more studies should report on how tools developed to measure QMNC are used in real-world practice outside pure research settings, and whether they can actually contribute in helping mothers and babies to receive the best possible care.

ORCID iDs
Marzia Lazzerini http://orcid.org/0000-0001-8608-2198
Ilaria Mariani http://orcid.org/0000-0001-8260-4788
Emanuelle Pessa Valente http://orcid.org/0000-0002-4741-4628

REFERENCES

1 World Health Organization. Maternal mortality. Available: https://www.who.int/news-room/fact-sheets/detail/maternal-mortality [Accessed 23 Oct 2020].

2 United Nations Children's Fund A Neglected Tragedy. The global burden of stillbirths. Report of the un Inter-agency group for child mortality estimation, 2020. Available: https://childmortality.org/wp-content/uploads/2020/10/UN-IGME-2020-Stillbirth-Report.pdf [Accessed 30 Nov 2020].

3 United Nations Children's Fund. Levels and trends in child mortality. Report, 2020. Available: https://childmortality.org/wp-content/uploads/2020/09/UNICEF-2020-Child-Mortality-Report.pdf [Accessed 30 Nov 2020].

4 Harrison S, Alderdice F, Henderson J. *You and your baby: a national survey of health and care*. Oxford: National Perinatal Epidemiology Unit, University of Oxford, 2020. ISBN: 978-0-9956854-5-1.

5 Hundley V, Rennie AM, Fitzmaurice A, *et al*. A national survey of women's views of their maternity care in Scotland. *Midwifery* 2000;16:303–13.

6 Declercq ER, Sakala C, Corry MP. *Listening to Mothers III: report of the third national U.S. survey of women's childbearing experiences*. New York: Childbirth Connection, 2013.

7 Euro-Peristat Project. European perinatal health report. core indicators of the health and care of pregnant women and babies in Europe in 2015, 2015. Available: https://www.europeristat.com/images/EPHR2015_web_hyperlinked_Euro-Peristat.pdf [Accessed 12 Dec 2020].

8 Direzione Generale della digitalizzazione, del Sistema Informativo Sanitario e della Statistica. Ufficio di Statistica. Certificato di assistenza al parto (CeDAP). Analisi dell'evento nascita -Anno 2017. Roma, 2020. Available: http://www.salute.gov.it/imgs/C_17_pubblicazioni_2931_allegato.pdf [Accessed 12 Dec 2020].

9 Lauria L, Lamberti A, Buoncristiano M. Pre- and post-natal assistance: promotion and assessment of operational models quality. The 2008-2009 and 2010-2011 surveys (in Italian). In: *Rapporti ISTISAN 12/39*. Istituto Superiore di Sanit, 2012.

10 Bohren MA, Vogel JP, Hunter EC, *et al*. The mistreatment of women during childbirth in health facilities globally: a mixed-methods systematic review. *PLoS Med* 2015;12:e1001847.

11 Khosla R, Zampas C, Vogel JP, *et al*. International human rights and the mistreatment of women during childbirth. *Health Hum Rights* 2016;18:131–43.

12 World Health Organization. The prevention and elimination of disrespect and abuse during facility-based childbirth, 2014. Available: http://www.who.int/reproductivehealth/topics/maternal_perinatal/statement-childbirth/en/ [Accessed 22 Oct 2020].

13 Jomeen J, Redshaw M. Ethnic minority women's experience of maternity services in England. *Ethn Health* 2013;18:280–96.

14 Redshaw M, Hockley C. Institutional processes and individual responses: women's experiences of care in relation to cesarean birth. *Birth* 2010;37:150–9.

15 Ravaldi C, Skoko E, Battisti A, *et al*. Abuse and disrespect in childbirth assistance in Italy: a community-based survey. *Eur J Obstet Gynecol Reprod Biol* 2018;224:208–9.

16 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.

17 Tunçalp Ö, Were WM, MacLennan C, *et al*. Quality of care for pregnant women and newborns-the who vision. *BJOG* 2015;122:1045–9.

18 World Health Organization. Standards for improving quality of maternal and newborn care in health facilities, 2016. Available: http://www.who.int/reproductivehealth/publications/maternal_perinatal_health/improving-mnh-health- facilities/en/

19 Agency for healthcare research and quality. Plan-Do-Study-Act (PDSA) directions and examples. content last reviewed February 2015, 2015. Available: https://www.ahrq.gov/health-literacy/quality-resources/tools/literacy-toolkit/healthlittoolkit2-tool2b.htm

20 Brizuela V, Leslie HH, Sharma J, *et al*. Measuring quality of care for all women and newborns: how do we know if we are doing it right? A review of facility assessment tools. *Lancet Glob Health* 2019;7:e624–32.

21 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.

22 Bohren MA, Mehrtash H, Fawole B, *et al*. How women are treated during facility-based childbirth in four countries: a cross-sectional study with labour observations and community-based surveys. *Lancet* 2019;394:1750–63.

23 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.

24 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.

25 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.

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

27 Salkind NJ. *Encyclopedia of research design*. Thousand Oaks, CA: SAGE Publications, Inc, 2010.

28 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.

29 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.

30 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.

31 Taherdoost H. Validity and reliability of the research instrument; how to test the validation of a Questionnaire/Survey in a research. *International Journal of Academic Research in Management* 2016;5:hal-02546799.

32 Oluwatayo JA. Validity and reliability issues in educational research. *Journal of educational and social research* 2012;2 https://www.richtmann.org/journal/index.php/jesr/article/view/11851

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 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.

35 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.

36 The Universal Rights of Childbearing Women. *Respectful maternity care chart*, 2011.

37 World health organisation regional office for Europe. Hospital care for children: quality assessment and improvement tool. Copenhagen:, 2015. Available: http://www.euro.who.int/en/health-topics/Life-stages/child-and-adolescent-health/publications/2015/hospital-care-for-children-quality-assessment-and-improvement-tool [Accessed 12 Dec 2020].

38 Bacci A, Hodorogea S, Khachatryan H, *et al*. What is the quality of the maternal near-miss case reviews in who European region? cross-sectional study in Armenia, Georgia, Latvia, Republic of Moldova and Uzbekistan. *BMJ Open* 2018;8:e017696.

39 Uxa F, Bacci A, Mangiaterra V, *et al*. Essential newborn care training activities: 8 years of experience in eastern European, Caucasian and central Asian countries. *Semin Fetal Neonatal Med* 2006;11:58–64.

40 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. Jun 1.

41 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.

42 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.

43 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* 2020.

44 DeVellis R. *Scale development: theory and applications*. 4 edn. Chapel Hill, USA: University of North Carolina, 2016.

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