Quality of care during childbirth in low-resource settings: Applying an epidemiology lens to an implementation problem

Stephanie A. Kujawski

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

Quality of care during childbirth in low-resource settings: Applying an epidemiology lens to an implementation problem

Stephanie A. Kujawski

While significant progress has been made towards improving health outcomes in low-resource settings, unacceptably high maternal mortality remains a problem. Efforts to improve maternal mortality in low-resource settings did not yield intended results. One hypothesized reason for insufficient maternal mortality progress is poor interpersonal quality of care during childbirth at health facilities. Qualitative studies support the assumptions of quality of care frameworks that connect structural inputs (e.g. drugs and supplies, equipment, human resources) to interpersonal quality. However, there is no quantitative evidence for this relationship. Further, although maternal health researchers developed quantitative tools to measure interpersonal quality of care, the construct is mainly operationalized as a single, bipolar dimension, measured as respectful maternity care (good care) or disrespect and abuse (poor care). To address these limitations, this dissertation used an epidemiologic perspective to test the underlying assumptions of quality of care frameworks and to create a robust measure of interpersonal quality of care. This dissertation consists of three parts: an empirical study to test the hypothesis that structural inputs have a positive effect on interpersonal quality of care; a systematic review of the literature of instruments measuring the construct of interpersonal quality of care and their reliability, validity, and dimensionality; and an empirical study to assess the dimensionality and construct validity of the Maternal Health Interpersonal Quality Scale, a measure of interpersonal quality of care.

The first empirical study did not find meaningful associations between HIV structural inputs and maternal health structural inputs and interpersonal quality of care during childbirth. These results do not support the assumptions of quality of care frameworks nor qualitative evidence linking structural inputs and interpersonal quality of care. The systematic review suggested that the construct of interpersonal quality of care is not well-defined, that few instruments met psychometric standards for adequate reliability and validity, and that studies that assessed the instruments were generally of poor quality. The second empirical study found that interpersonal quality of care formed a two-dimensional, correlated structure, with one dimension measuring respectful maternity care and one dimension measuring disrespect and abuse. Overall, this dissertation used an epidemiologic lens to address an implementation problem in maternal health. While there is a need to improve interpersonal quality of care during childbirth, in order to impact change and to avoid implementation failure, it is imperative to ensure interventions have a strong evidence base and to use validated measures of the construct.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Acknowledgments

I am deeply indebted to Dr. Sharon Schwartz, my sponsor and mentor, for her extraordinary mentorship, support, patience, and instruction. She devoted countless hours to me and my dissertation and challenged me to become more of a critical thinker, researcher, and writer. She shaped my doctoral experience, and I am forever grateful.

I am also thankful to the rest of my dissertation committee. Thank you to Dr. Susie Hoffman, my chair, for her thoughtful feedback and close reads of my dissertation, and to Dr. Batya Elul for her guidance and insight as my T32 mentor and second reader. A special thank you to Dr. Andrea Norcini Pala for his incredible patience and for teaching me the art of measurement. I am extremely grateful to Dr. Margaret Kruk for her mentorship and support and for inspiring me to research quality of care and to enter the field of epidemiology.

I would like to acknowledge the research teams at the Averting Maternal Death and Disability Program at Columbia University and Ifakara Health Institute who are wonderful and supportive collaborators and provided the data for Chapter 3. Warmest thanks to Lynn Freedman for introducing me to the topic of interpersonal quality of care and for her invaluable and critical feedback.

Thank you to the Epidemiology Department at Columbia University, specifically my fellow doctoral students for their companionship and inspiration, and Liliane Zaretsky for her support and always having an open door. I would like to acknowledge the Global HIV Implementation

Science Research Training Fellowship for providing me with funding and protected research time during my doctoral studies.

Finally, my deep gratitude and love to my family and friends who have encouraged me always, especially my husband, Kyle Winnick; my parents, Elizabeth Szancer, Nathan Kujawski, and Tom Zoufaly; and Melissa and Andy Rosen. I would not have made it to this point without you. Special thanks to Mom and Melissa for picking up the phone at all hours, and to Kyle for his patience and positivity and for being by my side on this journey.

Dedication

To Mom, for her unwavering support and for all of the opportunities she has provided me; and to Kyle, for his infinite encouragement and love

Chapter 1: Introduction

1.1. Background

In the early 2000s, low-resource settings were plagued by persistently high maternal mortality with the maternal mortality ratio as high as 846 per 100,000 live births in sub-Saharan Africa in 2000.[1] In response, low-resource countries in the Millennium Development Goal era focused on promoting facility-based deliveries. The strategy was two-pronged. First, facilities were equipped with appropriate supplies and staff skilled to handle complications and to treat life-threatening conditions, such as eclampsia and postpartum hemorrhage.[2,3] Second, known barriers to facility-based care, such as cost, distance, and lack of transportation, were removed to increase access to health facilities.[2,4]

Utilization of facility-based delivery services in low-resource settings improved over time, with an increase in skilled birth attendants for delivery from 57% to 70% from 1990-2014.[5] However, these improvements did not translate into the expected decrease in maternal mortality, with a reduction in the maternal mortality ratio by 46% in low-resource settings over the same time period—far below the goal of a 75% reduction.[5]

The reasons for the intractability of high maternal mortality ratios are many, and several fall within the domain of quality of care, including facility infrastructure deficiencies, lack of skilled personnel, non-compliance with technical quality standards, and poor interpersonal quality of care.[6-9] These deficiencies underscore the implementation failures of maternal health programs

and interventions aimed at reducing maternal mortality through increased facility-based deliveries.

Given these deficits, the maternal health field looked to quality of care frameworks to explain the failure of increased facility use to meet expected improvement goals and to guide intervention. One commonly applied framework, developed by Avedis Donabedian, proposes three linked domains: structure, process, and outcomes.[10] Structure is defined as the elements of the care setting, such as drugs and supplies, equipment, human resources, and the organizational structure needed to provide the care. Process denotes the services provided to the patient and is broken down into technical quality and interpersonal quality. Technical quality of care refers to the knowledge and skills needed to conform to the best practice or gold standard. Interpersonal quality of care concerns the relationship between the patient and the provider, which must meet individual and social expectations and standards. Outcome is the effect of the care and services on the patient's health. The framework hypothesizes that that structure enables processes, and processes allow for favorable outcomes.[10] Specific to maternal health, in 2015, the World Health Organization proposed a quality of care framework for maternal and newborn health based on Donabedian's framework.[11] It includes the same three core domains of structure, process, and outcomes. However, the World Health Organization's framework is not explicit as to what is included in the structure domain.

In the past few years, the maternal health field has focused on interpersonal quality of care during childbirth as a point of intervention. Anecdotal reports and qualitative studies identified various manifestations of poor interpersonal quality of care by providers in health facilities, such

as physical abuse, verbal abuse, physical privacy violations, inappropriate demands for payment, and neglect in time of need. A recent systematic review of qualitative literature from over 30 countries reported consistent evidence of poor interpersonal quality of care during labor and delivery.[12] Studies indicate that this poor care may act as a deterrent to future utilization of maternal health services at health facilities.[6,12,13]

These studies suggest that poor interpersonal quality of care during childbirth at health facilities may have inhibited maternal mortality progress. In response to the evidence, the maternal health field created a respectful maternity care movement, aimed to promote respectful care and combat poor interpersonal quality of care during childbirth.[14,15] The research and movement to date demonstrate that an interpersonal quality problem exists based on qualitative studies and prevalence studies of poor care. There are two salient issues with the existing evidence base. First, the assumptions underlying the quality of care frameworks linking structural inputs and interpersonal quality is limited to qualitative studies.[12,16] Few interventions have addressed interpersonal quality of care in the maternal health context. These interventions were multicomponent intervention packages that included structural inputs, but were not able to isolate specific aspects of the intervention that were responsible for the outcomes.[17,18] However, before intervention and scale up, a useful next step is to examine the quantitative evidence for causal effects. Without causal identification, interventions may be poorly targeted and fail to replicate at scale.[19] Second, while maternal health researchers developed quantitative tools to measure interpersonal quality of care,[20-22] the utility of the tools are stymied by imprecise construct operationalization and improper validation. The construct is mainly measured as either

good or poor interpersonal quality of care, implying that the construct is opposite ends of a single continuum.[20,22,23]

This dissertation aims to build on the qualitative work with robust epidemiologic evidence and to test the connections in quality of care frameworks and to ensure that measures are appropriately reflecting the underlying construct. Chapter 2 tests the links in quality of care frameworks by examining whether HIV structural inputs, as a distal factor, and maternal health inputs, as a proximal factor, have positive effects on interpersonal quality of care. Chapter 3 presents a systematic review of the literature on instruments measuring the construct of interpersonal quality of care in health care settings and their validity, dimensionality, and reliability. Chapter 4 assesses the dimensionality and the construct validity of a measure of interpersonal quality of care during childbirth, the Maternal Health Interpersonal Quality Scale, with the particular goal of determining whether questions about positive and negative aspects of interpersonal quality form a unidimensional or two-dimensional scale. Overall this dissertation aimed to bring an epidemiologic perspective to an implementation problem in maternal health in order to guide the development and implementation of interventions to improve and address quality of care.

1.2. References

1. World Health Organization. *Trends in Maternal Mortality: 1990 to 2015: Estimates by WHO, UNICEF, UNFPA, World Bank Group, and the United Nations Population Division.* Geneva, Switzerland: World Health Organization; 2015

2. Campbell OM, Graham WJ. Strategies for reducing maternal mortality: getting on with what works. *Lancet.* 2006;368:1284-1299.

3. Paxton A, Maine D, Freedman L, Fry D, Lobis S. The evidence for emergency obstetric care. *International Journal of Gynecology and Obstetrics.* 2005;88:181-193.

4. Thaddeus S, Maine D. Too far to walk: Maternal mortality in context. *Soc Sci Med.* 1994;38:1091-1110.

5. United Nations. *The Millennium Development Goals Report 2015.* New York: United Nations; 2015. http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG 2015 rev (July 1).pdf

6. Bohren MA, Hunter EC, Munthe-Kaas HM, Souza JP, Vogel JP, Gülmezoglu AM. Facilitators and barriers to facility-based delivery in low- and middle-income countries: a qualitative evidence synthesis. *Reprod Health.* 2014;11:71.

7. Campbell OMR, Calvert C, Testa A, et al. The scale, scope, coverage, and capability of childbirth care. *Lancet.* 2016;388:2193-2208.

8. Kruk ME, Leslie HH, Verguet S, Mbaruku GM, Adanu RMK, Langer A. Quality of basic maternal care functions in health facilities of five African countries: an analysis of national health system surveys. *Lancet Glob Health.* 2016;4:e845-e855.

9. Souza JP, Gulmezoglu AM, Vogel J, et al. Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): A cross-sectional study. *Lancet.* 2013;381:1747-1755.

10. Donabedian A. The quality of care. How can it be assessed? *JAMA.* 1988;260:1743-1748.

11. Tunçalp Ö, Were W, MacLennan C, et al. Quality of care for pregnant women and newborns-the WHO vision. *BJOG.* 2015.

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

13. D'Ambruoso L, Abbey M, Hussein J, et al. Please understand when I cry out in pain: women's accounts of maternity services during labour and delivery in Ghana. *BMC Public Health.* 2005;5:140.

14. White Ribbon Alliance. *Respectful Maternity Care Charter: The Universal Rights of Childbearing Women.* Washington, DC: White Ribbon Alliance; 2011. http://whiteribbonalliance.org/wp-content/uploads/2013/10/Final_RMC_Charter.pdf

15. World Health Organization. The prevention and Elimination of Disrespect and Abuse during Facility-Based Childbirth. 2014; http://apps.who.int/iris/bitstream/10665/134588/1/WHO_RHR_14.23_eng.pdf?ua=1&ua=1.

16. Mselle LT, Moland KM, Mvungi A, Evjen-Olsen B, Kohi TW. Why give birth in health facility? Users' and providers' accounts of poor quality of birth care in Tanzania. *BMC Health Serv Res.* 2013;13:174.

17. Kujawski S, Freedman L, Ramsey K, et al. Community and health system intervention to reduce disrespect and abuse during childbirth in Tanga Region, Tanzania: A comparative before-and-after study. *PLoS Med.* 2017;14.

18. Abuya T, Ndwiga C, Ritter J, et al. The effect of a multi-component intervention on disrespect and abuse during childbirth in Kenya. *BMC Pregnancy Childbirth.* 2015;15:224.

19. Schwartz S, Gatto NM, Campbell UB. Causal identification: a charge of epidemiology in danger of marginalization. *Ann Epidemiol.* 2016;26:669-673.

20. Kruk ME, Kujawski S, Mbaruku G, Ramsey K, Moyo W, Freedman L. Disrespectful and abusive treatment during facility delivery in Tanzania: a facility and community survey. *Health Policy Plan.* 2014:czu079.

21. Sheferaw ED, Mengesha TZ, Wase SB. Development of a tool to measure women's perception of respectful maternity care in public health facilities. *BMC Pregnancy Childbirth.* 2016;16:67.

22. Rosen HE, Lynam PF, Carr C, et al. Direct observation of respectful maternity care in five countries: a cross-sectional study of health facilities in East and Southern Africa. *BMC Pregnancy Childbirth.* 2015;15:306.

23. Abuya T, Warren CE, Miller N, et al. Exploring the prevalence of disrespect and abuse during childbirth in Kenya. *PLoS One.* 2015;10(4):e0123606.

Chapter 2: Do HIV and maternal health structural inputs have a positive effect on interpersonal quality of care during childbirth? An examination of proximal and distal structural inputs

2.1. Abstract

Background: The maternal health field has recently focused on the importance of interpersonal quality of care and continues to cite structural deficits as a contributor to poor interpersonal treatment. This hypothesis is supported by qualitative evidence; however, there is no quantitative evidence of this relationship. This study tested the effect of HIV structural inputs, as a distal factor, and maternal health structural inputs, as a proximal factor, on interpersonal quality of care during childbirth. A secondary analysis tested whether maternal health structural inputs were a mediator of the relationship between HIV structural inputs and interpersonal quality of care during childbirth.

Methods: Analyses were conducted using data from the 2013-2014 Malawi Service Provision Assessment, which documented the availability and quality of health facility services and included the observation of laboring and delivering women. The exposure variables, HIV and maternal health structural inputs, were measured using a sum of facility infrastructure variables that reflected the prevailing global standards and guidelines for quality maternity care and service readiness. The top 25% of the sum scores per variable were categorized as having high structural inputs and the bottom 75% as low structural inputs. The outcome, interpersonal quality of care, was measured as a sum score of 12 items collected during the observations. The effects of distal structural inputs and proximal structural inputs on interpersonal quality of care were assessed using linear regression with cluster robust standard errors to account for clustering of individual observations within health facilities. As part of the mediation analysis, the effect of HIV structural inputs on maternal health structural inputs was tested using a generalized linear

model with a Poisson distribution with a log link to estimate a risk ratio, and cluster robust standard errors to account for clustering within facilities.

Results: 474 observations of delivering women were completed in 222 health facilities, with 429 observations in 204 facilities included in the complete case analysis. 24.2% of participants delivered in a facility with high HIV structural inputs, 19.4% delivered in a facility with high maternal health structural inputs, and the mean for the interpersonal quality of care score was 9.0 (SD: 1.8). HIV structural inputs (β: -0.15, 95% CI: -0.75, 0.46) and maternal health structural inputs (β: -0.06, 95% CI: -0.70, 0.58) had small, non-meaningful effects on interpersonal quality of care during childbirth. There was a strong association between the distal exposure (HIV structural inputs) and mediator (maternal health structural inputs, RR: 3.44, 95% CI: 1.56, 7.57), but no evidence that maternal health structural inputs were a mediator of the relationship between HIV structural inputs and interpersonal quality of care during childbirth.

Discussion: These findings do not verify the quality of care frameworks or qualitative evidence that support the relationship between structure and interpersonal quality of care. While structural inputs are important for health system performance, the results of this study suggest that they might not be necessary for a respectful childbirth experience at a health facility.

2.2. Introduction

While significant progress has been made toward improving health outcomes in low-resource settings, maternal mortality remains a problem. As of 2015, the maternal mortality ratio in sub-Saharan Africa was 546 per 100,000 live births, 46 times that of high-resource countries.[1] The Millennium Development Goal 5 aimed to reduce maternal mortality by 75% globally from 1990-2015. To achieve that goal, low-resource countries, where maternal mortality was highest, focused on moving childbirth from homes to adequately equipped and staffed facilities to manage life-threatening delivery complications, such as eclampsia and hemorrhage.[2,3] Emphasis was placed on removing barriers to facility-based care, such as transportation and cost, and on scaling up capacity at such facilities to provide high quality obstetric care.[2,4] From 1990 to 2014, these methods contributed to the increase in births attended by skilled personnel—a measure of facility utilization—from 57% to 70% in low-resource countries.[5]

Despite this improvement in facility utilization, the Millennium Development Goal was not met. In low-resource settings, maternal mortality decreased by 46% from 1990-2013, which, while impressive, was a far from the 75% goal.[5] Low-resource countries, where 99% of all maternal deaths occur, remain plagued by high maternal mortality rates.[6]

Among the factors hypothesized for the intractability of high maternal mortality ratios is poor interpersonal quality of care during childbirth in health facilities. Evidence of this poor care emerged from anecdotal reports, qualitative studies, and prevalence studies in low-resource settings.[7-9] Examples include physical abuse, verbal abuse, neglectful care, lack of consent for surgical procedures, and non-confidential care.[7-10]

To guide intervention to improve interpersonal quality of care, the maternal health field looked to quality of care frameworks, one of which proposes that structural inputs (e.g. drugs and supplies, equipment, human resources) impact the interpersonal quality of care provided.[11,12] In the context of maternal health, mainly qualitative studies support this claim.[8,13] These studies suggest that structural constraints in maternity wards directly limit the provision of adequate interpersonal quality of care.[13-15] For example, lack of supplies such as gloves, may lead providers to scold or neglect patients who do not bring supplies or to ask patients to buy their own.[13,15] This practice may be perceived as an unnecessary payment or bribe, especially when these services and supplies, by policy, are supposed to be free. Shortage of human resources and supplies can also lead to psychological stress for the providers, what is referred to as moral distress (inability to carry out moral decisions due to contextual constraints) and burnout (emotional exhaustion, depersonalization, increased dissatisfaction with one's work), which in turn may result in poor provider behavior and lack of empathy and compassion for patients.[13,16-20] However, there is no quantitative evidence of the relationship between structural inputs and interpersonal quality of care during childbirth. Some interventions to address interpersonal quality of care included structural inputs, but, as multicomponent interventions, they were not able to isolate which aspects of the intervention were responsible for change.[21,22] Prior to further intervention development and scale-up, there is a need to test the theory that maternal health structural inputs affect interpersonal quality of care during childbirth.

In addition to proximal health system factors, such as maternal health structural inputs, distal health system factors, such as HIV structural inputs, may also have an effect on the quality of care provided for maternal health. In sub-Saharan Africa, since 2005, the HIV epidemic resulted in an influx of resources and investments in the health system to scale up and address the HIV

health needs of the population.[23] Given the overhaul of the health system with the introduction of HIV programs, there is increased attention to examining the spillover effects of HIV programs on other health services, specifically to examine if HIV structural inputs, such as infrastructure improvements and increases in HIV-related human resources, influenced, either positively or negatively, the quality of non-HIV services in the same health facility.[24,25] However, studies to date used crude measures of the presence and/or funding of HIV services as a proxy measure for structural inputs and service utilization as the outcome,[24,26] which do not account for the heterogeneity in the implementation of facility improvements and provision or comprehensiveness of services. In this study, operationalizing the exposure as HIV structural inputs and the outcome as a quality measure may provide a more robust approximation of how the effect of HIV programs on non-HIV services.

Using data from health facilities in Malawi, the aim of this study was to test whether HIV structural inputs, as a distal factor, and maternal health structural inputs, as a proximal factor, have positive effects on interpersonal quality of care. A secondary goal was to test whether maternal health structural inputs were a mediator of the relationship between HIV structural inputs and interpersonal quality of care during childbirth.

2.3. Methods

Data source: sample and design

Data for this aim came from the 2013-2014 Malawi Service Provision Assessment.[27] Funded by the United States Agency for International Development (USAID) and implemented by ICF International and ministries of health, the Service Provision Assessments are nationally-

representative cross-sectional surveys that document the availability and quality of health facility services. Surveys include questionnaire items related to the infrastructure, resources, and capacity of key health facility services (family planning, maternal and newborn health, HIV/AIDS, child health, etc.), and the observation of certain client services. The 2013-2014 Malawi Service Provision Assessment also included the direct observation of women during delivery and childbirth in a sample of hospitals and health centers.

The 2013-2014 Malawi Service Provision Assessment was designed as a census of all 1,060 health facilities in the country. Of these, 540 facilities provided delivery services and 222 (41%) were chosen for observation based on whether there were delivering women available when the data collectors were at the facilities.[27] There were 474 observations of delivering women completed. This analysis was restricted to the health facilities where the 474 observations of delivering women occurred. Data were collected from June-August 2013 and November 2013-February 2014. More information about the data collection is available elsewhere.[27] Data are publicly available from the DHS Program (dhsprogram.com/data/available-datasets.cfm).

ICF International, the Service Provision Assessment implementing agency, obtained ethical approval to conduct this study. Providers and patients who were observed provided informed consent. This analysis was exempt from human subjects review by the Institutional Review Board of Columbia University.

Measures

Exposures

Distal structural inputs: HIV structural inputs

The first exposure of interest is HIV structural inputs, representing distal structural inputs. This was defined as the sum of 25 facility-level structural input variables for HIV testing and counseling, antiretroviral treatment (ART), and prevention of mother to child transmission (PMTCT) (Table 2.1). These variables were guided by global standards and indicate the facility's service readiness and capacity to provide HIV services.[28,29] Variables were coded as 1 if the HIV infrastructure input was present and 0 if it was not present. The variables were summed and then the score was split into quartiles. The top 25% were categorized as having high HIV structural inputs and the bottom 75% as low HIV structural inputs. This cutoff was chosen to represent realistic expectations of available resources in low-resource settings.[30] Alternative thresholds were also examined to ensure that the data support this cut off.

Proximal structural inputs: maternal health structural inputs

The second exposure of interest is proximal structural inputs, represented by maternal health structural inputs. This was similarly measured using a sum score of 26 facility-level maternal health infrastructure indicators that were available from the Service Provision Assessment surveys (Table 2.2). Variables were coded as 1 if the maternal health infrastructure input was present and 0 if it was not present. The variables included reflect the global standards and guidelines for quality maternity care and service readiness at the time of data collection.[29,31] The sum score was split into quartiles, with the top 25% categorized as having high maternal health structural inputs and the bottom 75% as low maternal health structural inputs. As with HIV structural inputs, this cutoff was chosen to represent realistic expectations of available resources in low-resource settings.[30] Again, alternative thresholds were also examined to ensure that the

data support this cut off. See Appendix 2.1 for more information about how the exposure variables were created.

Outcome: interpersonal quality of care during childbirth

The outcome is interpersonal quality of care during childbirth. Items measuring interpersonal quality of care were collected as part of the observations of delivering women in health facilities. Ten items measured respectful maternity care and two items measured negative interpersonal treatment (Table 2.3). The negative items were reverse coded. Variables were coded as 1 if the interpersonal quality of care item was reported and 0 if it was not reported. The items were then summed to create a scale. To use as much of the data as possible and to maintain consistency with the underlying construct, the variable was treated as continuous.

Potential confounding variables

Potential confounding variables for the relationship between distal structural inputs and interpersonal quality of care and proximal structural inputs and interpersonal quality of care were hypothesized based on the relevant HIV and maternal health literatures.[7,8,25,26] Directed acyclic graphs (DAG) were used to assist in determining a minimally sufficient set of confounding variables for which to control in the analyses (Figure 2.1).

Facility-level variables, such as health facility ownership, level of facility, if the facility receives any donor funding, whether the facility charged user fees for services, and the number of clinical staff would likely influence the level of structural inputs at the facility, both for HIV and maternal health, and may also affect how women are treated during childbirth. For example, a

higher-level facility, such as a hospital, would likely have more structural resources, as would facilities that receive outside funding, either through donors or collection of user fees. This facility would likely have a higher patient load and women may therefore be less likely to experience high interpersonal quality care. Health facility level was categorized into hospital vs. health center or clinic. Facility ownership was categorized into public ownership vs. other (private non-profit, private for-profit, or company owned). Donor funding was defined as the facility receiving funding either from an outside donor (e.g. foreign government), non-governmental organization, or faith-based program. A facility was categorized as charging user fees if it collected fees for any service provided. The number of clinical staff was treated as a continuous variable of the sum of skilled providers at a health facility from any of the following categories: doctor, clinical officer, clinical technician, medical assistant, registered nurse, or enrolled nurse. Facility variables were collected as part of the Service Provision Assessment and were measured using a close-ended facility checklist.

Individual-level factors, such as socioeconomic status (unmeasured), parity and HIV status, may also confound the relationships between the exposures and the outcome. For example, women who are primipara are more likely to choose higher-level facilities,[32] which have more structural resources, and they are also more likely to be treated poorly.[7] It is possible that HIV-positive women may be more likely to seek HIV care at a facility that is of higher HIV quality, but may be treated more poorly during childbirth than HIV-negative women due to HIV stigma. The evidence of the effect of HIV status on interpersonal quality of care during childbirth is mixed. One study found no significant differences in how women were treated during childbirth between HIV-positive and HIV-negative women.[33] Qualitative research indicated that HIV is

normalized due to trainings and prevention services.[33] However, the lack of infection prevention supplies, such as gloves, may influence how providers treat women if there is fear of infection transmission.[33,34]

Delivery experience factors, such as complications experienced during delivery or having a Caesarean section may also confound the relationship between proximal structural inputs and the outcome. Women who develop complications during labor and delivery or require a Caesarean section may be referred from a lower-level to a higher-level facility, which has more structural resources. Women who have complications during childbirth are also more likely to be treated poorly during childbirth, while women who have a Caesarean section are less likely to be treated poorly.[7] Any of the following was considered as experiencing a complication: eclampsia, major blood loss, fever, antibiotic use, or failure to progress/prolonged labor. Observers collected individual-level variables as part of the labor and delivery observation tool.[27]

Sample size and power

In order to calculate power, I needed to adjust for the design effect, which accounts for the clustering by facility.[35] Applying the survey's design effect of 1.36 to my full sample size of 474, my effective sample size is 349. Accounting for missing data, my complete case sample size is 429, with an effective sample size of 318 (design effect = 1.35).

For both main analyses, I calculated the minimum detectable effect size given an alpha of 0.05, a ratio of unexposed to exposed of 3, and a sample size of 318 or 349. For the distal exposure, with a standard deviation of 2.0 in the exposed and 1.8 in the unexposed, I would have at least 0.80

power to detect a mean difference of 0.70. For the proximal exposure, with a standard deviation of 1.7 in the exposed and 1.9 in the unexposed, I would have at least 0.80 power to detect a mean difference of 0.63. Power calculations were performed using Open Epi 3.01 (www. openepi.com).

For the mediation analysis, I used the MedPower app (https://davidakenny.shinyapps.io/MedPower/).[36] With a sample size of 349 and an alpha of 0.05, I would have at least 0.80 power to detect a 0.15 change in the standard deviation for the direct effect and 0.031 change in the standard deviation for the indirect effect. With a sample size of 318 and an alpha of 0.05, I would have at least 0.80 power to detect a 0.16 change in the standard deviation for the direct effect and a 0.032 change in the standard deviation for the indirect effect. See Appendix 2.2 for more details about the mediation power analysis and calculation of the design effect.

Statistical analysis

I used univariate statistics to explore the distribution of the exposures, outcome, and confounding variables. I then conducted bivariate analyses between distal structural inputs and interpersonal quality of care and proximal structural inputs and interpersonal quality of care.

To assess the effect of distal structural inputs and proximal structural inputs on interpersonal quality of care, I used linear regression with cluster robust standard errors to account for the clustering of individual observations in health facilities. The Service Provision Assessment data typically uses complex survey weights to account for the sampling approach and to make inferences at the national level. This analysis did not apply complex survey weights as the

women observed during delivery were chosen as a convenience sample. I tested for confounding by looking for indicators in my data, specifically that potential confounding variables, as outlined above, were associated with the exposures and the outcome in each analysis at $p<0.20$.[37] I then controlled for confounding guided by the DAGs.

As a secondary analysis, I tested whether maternal health structural inputs mediated the relationship between HIV structural inputs and interpersonal quality of care. As there was no interaction between HIV structural inputs and maternal health structural inputs, the mediation analysis followed methods as specified by Baron and Kenny.[38] First, I tested the relationship between HIV structural inputs and maternal health structural inputs (path *a*) (Figure 2.2.) using a generalized linear model with a Poisson distribution, a log link, and cluster robust standard errors. This approach estimates a risk ratio. As path *b* and the total effect (path *c*) were already tested in the main analyses, I next determined whether maternal health structural inputs were a mediator, by comparing the parameter estimate of the total effect to the parameter estimate of the direct effect (path *c'*) using linear regression with cluster robust standard errors. All analyses were performed using Stata version 14.2 (Stata Corp, College Station, TX, USA).

Missing data

In determining my approach to missing data, I considered the percent of missing data, the mechanism through which the missing data arose, and the consequences of the missing data, particularly in relation to the exposure and outcome variables in each analysis. Exploring the missing data revealed that there was no pattern to the missingness (i.e. pattern was not monotone). The majority of the missing data were for the outcome variable (9%). HIV structural

inputs were missing for 0.42% of the sample and there was no missing data for maternal health structural inputs or any of the confounding variables. For the outcome, the majority of the missing data was from variables that reflect the initial patient-provider interaction. It is possible that the observers may have missed these initial encounters due to when the observers started their observations on a given day. This likely occurred at random and was not related to the exposure variables or any of the confounding variables. Therefore, the missing data for this variable are likely missing completely at random (MCAR). The missing data were minimal; thus, I proceeded with a complete case analysis.

Sensitivity analyses

Additional analyses were planned to better understand the observed results. First, I assessed the sensitivity to the cut point for high structural inputs by defining high structural inputs for both HIV and maternal health as the top 10% of the distribution instead of the top 25%. I then compared these results with the main results to determine whether this more conservative measure showed the same relationships between each exposure and the outcome and in the mediation analysis.

Second, I assessed the effect of the missing data on the results. I performed a worst-case scenario sensitivity analysis by treating the missing outcome variables included in the composite outcome measure as 0, and a best-case scenario by treating the missing values as 1. This created an upper and lower bound for the interpersonal quality of care score. I compared these imputed results using the full dataset with my main analysis to see, under these best or worse cases, whether my results still held.

2.4. Results

Sample description

There were 540 facilities in Malawi that offered delivery services, of which 222 were chosen for observation based on whether there were delivering women available when the data collectors were at the facilities. As seen in columns 2, 3, and 4 in Table 2.4, there were some significant differences between the facilities chosen for observation and those that were not. Facilities with observations were more likely to be hospitals, in urban settings, and have more clinical staff. These facilities also had higher levels of HIV structural inputs and maternal health structural inputs. Of the 222 health facilities chosen for observation, 204 were included in the complete case analysis (Table 2.4, column 5). Of the 204 health facilities, 31.9% were hospitals (N=65) and the majority were publicly owned (N=140, 68.6%).

Characteristics of the participants in the total study sample and complete case sample are shown in Table 2.5. Of the 474 observations of laboring women in 222 health facilities, 429 observations in 204 facilities were available in the complete case sample (9% were missing for interpersonal quality of care, and 0.4% were missing for HIV structural inputs). On average, two women were observed at each facility. Participants were a mean age of 25.1 (SD = 6.4). For about a quarter of the sample (N=118, 27.5%), this was their first birth, and 5.8% (N=25) were HIV-positive. The sample was almost evenly distributed between hospitals (N=209, 48.7%) and health centers or clinics (N=220, 51.3%). The majority of the sample was observed in public (N=316, 73.7%) and in rural facilities (N=273, 63.6%).

Table 2.6 and Figure 2.3 show the distribution of the three main variables: HIV structural inputs (distal structural inputs), maternal health structural inputs (proximal structural inputs), and interpersonal quality of care. About one-quarter of the 429 participants (N=104, 24.2%) delivered in a health facility with high HIV structural inputs, while about one in five women (N=83, 19.4%) delivered in a health facility that met the threshold for high maternal health structural inputs. The facilities where women gave birth scored better on maternal health inputs (mean: 20.7, 80% of 26 items endorsed) than HIV inputs (mean: 16.6, 66% of 25 items endorsed). Nearly all women gave birth in a health facility with injectable oxytocin (97.2%), injectable magnesium sulfate (95.1%), the availability of neonatal bag or mask for resuscitation (96.0%) and suction apparatus (96.3%). There were key facility deficiencies: 40.1% of facilities had an examination light and 62.7% of facilities had staff that received any in-service training in intrapartum care in the last 24 months. For the delivery of HIV care, the majority of women gave birth in facilities that had gloves (96.3%) and first-line antiretroviral treatment (80.4%) available, but that lacked laboratory diagnostic capacity (CD4 tests or viral load:12.6%; renal or liver function test: 28.2%; full blood count: 30.5%).

The mean on the interpersonal quality of care scale was 9.0 (SD: 1.8), corresponding to the endorsement of 75% of the 12 items. Nearly all women were greeted respectfully by health providers (95.8%) and were not shouted at (99.8%) or slapped/hit (99.3%). However, a little more than a third of women were covered during labor with a drape (34.5%) and were asked if they had any questions during labor (36.1%).

Confounding variables

Potential confounders were tested to confirm their associations with each of the three main variables (Table 2.7). The only confounder of the distal exposure-outcome relationship was the number of clinical staff in a facility. Women who delivered in facilities with more clinical staff were more likely to be in a facility with high HIV structural inputs and more likely to have a higher interpersonal quality of care score. Women who were primipara, experienced any complication during childbirth or who had a Caesarean section were more likely to deliver at a health facility with high maternal health structural inputs and had a higher interpersonal quality of care score. At the facility level, women who delivered in facilities that collected user fees and that had more clinical staff were more likely to have a higher interpersonal quality of care score and to be in a facility with high maternal health inputs.

Effect of distal and proximal structural inputs on interpersonal quality of care during childbirth

Table 2.8 shows the unadjusted and adjusted models for the effects of distal and proximal structural inputs on interpersonal quality of care. Distal structural inputs (HIV structural inputs) had a small and non-meaningful association with interpersonal quality of care in both unadjusted and adjusted models. Controlling for the number of clinical staff in a health facility, women who delivered in a facility with high HIV structural inputs scored, on average, 0.15 (95% CI: -0.75, 0.46) points lower on the interpersonal quality of care scale than those in facilities with low HIV structural inputs.

Similar to the distal effects, proximal structural inputs (maternal health structural inputs) had a small and non-meaningful association with interpersonal quality of care in both adjusted and unadjusted models (Table 2.8). Women who delivered in a facility with high maternal health

structural inputs scored, on average, 0.06 (95% CI: -0.70, 0.58) points lower on the interpersonal quality of care scale than those in facilities with low maternal health structural inputs, adjusting for confounders.

Mediation analysis

While there was no effect of the distal exposure on the outcome, it is possible that maternal health structural inputs are an inconsistent mediator. If the effect of the exposure on the mediator is the opposite sign of the effect of the mediator on the outcome, the total effect may appear to be null, representing a balanced effect.[36] However, there was no effect of the proximal exposure on the outcome, so inconsistent mediation is not likely the cause of the null effect of the distal exposure on the outcome. To fulfill the original aims of this paper, I proceeded with a mediation analysis. Table 2.9 shows the unadjusted and adjusted models for the mediation analysis. There was a strong association between the distal exposure (HIV structural inputs) and the mediator (maternal health structural inputs). Women who delivered in facilities with high HIV structural inputs were 3.44 (95% CI: 1.56, 7.57) times more likely to also be in a facility with high maternal health structural inputs than a facility with low maternal health structural inputs, controlling for facility level and number of clinical staff. The effect of HIV structural inputs on interpersonal quality of care did not change substantially with the addition of maternal health structural inputs, the hypothesized mediator (β: -0.20, 95% CI: -0.89, 0.50).

Sensitivity analyses

Three sensitivity analyses were performed to determine the robustness of the results to changes in the two exposure variables and the outcome variable (Table 2.10). First, changing the cut point

for both structural input measures to the top 10% of the distribution instead of the top 25% did not significantly affect the results. The effect estimates remained small and non-significant. Second, I treated any missing variable included in the interpersonal quality of care composite measure first as 0 and then as 1, thus providing an upper and lower bound for the values interpersonal quality of care score. By using single imputation, the sample size increased to 472. In both the best-case and worst-case scenarios, point estimates were all small and non-significant, and fell within the confidence intervals of the main analyses.

2.5. Discussion

The aim of this study was to test the assumptions of quality of care frameworks that link structural inputs and interpersonal quality of care. To my knowledge, this is the first study to quantitatively explore the effect of distal and proximal structural inputs on interpersonal quality of care during childbirth. I found small, non-significant effects of structural inputs—for both HIV and maternal health—on interpersonal quality of care during childbirth in health facilities in Malawi. While I observed evidence of spillover between HIV and maternal health structural inputs, maternal health structural inputs did not mediate the relationship between HIV structural inputs and interpersonal quality of care during childbirth. These results do not support the quality of care frameworks or the qualitative evidence of the link between structure and interpersonal process in maternal health.[8,11-13]

Based on the data, structural inputs do not impact interpersonal quality of care during childbirth. However, there are several reasons possible reasons for the failure to find these effects. First, qualitative studies focus on the facilitators of poor, as opposed to good, interpersonal quality of

care. However, the outcome measure used in this analysis contained mostly positive aspects of interpersonal quality of care. The presence of good interpersonal care may not preclude experiences of poor interpersonal care. Second, the outcome was restricted to data from the early stages of labor. Measures of interpersonal quality of care throughout the entire labor and delivery process and across the quality of care continuum, including both positive and negative measures of the care experience, may have resulted in a different effect of structure on the interpersonal quality of care received. Third, the majority of women in the sample delivered in health facilities that had high structural inputs for both HIV and maternal health and scored high on the interpersonal quality of care scale, which is encouraging. However, the lack of variation in the measures may have resulted in a ceiling effect that contributed to the lack of association. It is possible that facilities with much lower levels of structure, or a sample with a wider distribution of structural inputs would yield a relevant association. It is not surprising, then, that restricting the exposures to the top 10% of their distributions did not yield meaningful results. Fourth, it is possible that unmeasured confounding, such as the socioeconomic status of the woman or the socioeconomic status of where the facility was located, may have biased the results. However, these variables would likely have biased the results in a positive direction, as high socioeconomic status would likely lead to a woman to select a facility with higher structural quality, either for HIV or maternal health, and also to be treated better. Therefore, controlling for these variables, either alone or together, would not have resulted in effect sizes of greater than 1—the minimum size for a meaningful effect.

The strong, positive relationship between the HIV structural inputs and maternal health structural inputs is consistent with previous research that demonstrates positive spillover effects of the

scale up of HIV programs on non-HIV services in sub-Saharan Africa.[24-26,39] By operationalizing the spillover as structural inputs, it provides a closer approximation of the mechanism for spillover than previous studies that explored the effect of the presence or funding of HIV programs. Given the scale up and omnipresence of HIV services in sub-Saharan Africa, this illustrates that variation in HIV quality, as opposed to merely the presence of HIV services, may be an important factor affecting non-HIV services.

This study had several limitations. First, the Service Provision Assessment was designed as a census of all health facilities in Malawi, but the maternity ward observations were performed in facilities where delivering women were present on the day of data collection. The sample thus constitutes health centers and hospitals that have high maternity ward patient volume. A comparison of facilities where the observations were performed to those where no observations were performed revealed that a higher proportion of observation facilities were hospitals and located in urban settings. They also had higher structural inputs for both HIV and maternal health. Therefore, the results cannot be extrapolated to all health facilities in Malawi. Second, the presence of the observers may have caused the health providers to alter their behaviors for the better, known as the Hawthorne effect.[40] This may have resulted in an overly positive estimate of the interpersonal quality of care provided during childbirth at Malawi health facilities. Third, the indicators included in the structural input variables were all weighted the same. It is possible that specific indicators are more important for interpersonal quality than others. Fourth, the cross-sectional nature of the study design limits the ability to confirm the temporality of the distal and proximal structural inputs. Based on the literature and historical trends of donor aid, it is more likely that HIV structural inputs cause maternal health structural inputs than vice versa.[41] In

addition, due to the cross-sectional design, the data only reflect stock outs and structural elements on the day of the data collection. Relatedly, no information was available on the number of providers in the maternity ward, only in the facility as a whole, nor the patient to provider ratio, which are hypothesized in the qualitative literature as structural contributors to the interpersonal quality of care issue.[16] A more nuanced measure of the change in structure over time may provide a wider picture of the relationship between structure and interpersonal quality.

Conclusion

This study did not find an effect of structural inputs on interpersonal quality of care during childbirth. The maternal health field has recently focused on the importance of interpersonal quality of care and continues to cite structural deficits as a contributor to poor interpersonal treatment. Structural inputs are essential for the performance of technical quality of care in maternal health, but the results of this study suggest that they might not be necessary for a respectful childbirth experience at a health facility. While further studies are necessary to validate these findings, other potential causes of interpersonal quality of care during childbirth, such as power dynamics between patients and providers, lack of accountability, and organizational culture, warrant quantitative exploration.

2.6. References

1. Alkema L, Chou D, Hogan D, et al. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group. *Lancet.* 2016;387(10017):462-474.

2. Campbell OM, Graham WJ. Strategies for reducing maternal mortality: getting on with what works. *Lancet.* 2006;368:1284-1299.

3. Graham W, Bell J, Bullough C. Can skilled attendance at delivery reduce maternal mortality in developing countries? In: De Brouwere V, Van Lerberghe W, eds. *Safe Motherhood Strategies: A Review of the Evidence*. Antwerp: ITGPress; 2001:97-130.

4. Thaddeus S, Maine D. Too far to walk: Maternal mortality in context. *Soc Sci Med.* 1994;38:1091-1110.

5. United Nations. *The Millennium Development Goals Report 2015.* New York: United Nations; 2015.http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG 2015 rev (July 1).pdf

6. World Health Organization. *Trends in Maternal Mortality: 1990 to 2015: Estimates by WHO, UNICEF, UNFPA, World Bank Group, and the United Nations Population Division.* Geneva, Switzerland: World Health Organization; 2015

7. Kruk ME, Kujawski S, Mbaruku G, Ramsey K, Moyo W, Freedman L. Disrespectful and abusive treatment during facility delivery in Tanzania: a facility and community survey. *Health Policy Plan.* 2014:czu079.

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

9. Abuya T, Warren CE, Miller N, et al. Exploring the prevalence of disrespect and abuse during childbirth in Kenya. *PLoS One.* 2015;10(4):e0123606.

10. Bowser D, Hill K. *Exploring Evidence for Disrespect and Abuse in Facility-Based Childbirth: Report of a Landscape Analysis.* Washington, DC: Harvard School of Public Health, University Research Co., LLC; 2010

11. Donabedian A. The quality of care. How can it be assessed? *JAMA.* 1988;260:1743-1748.

12. Tunçalp Ö, Were W, MacLennan C, et al. Quality of care for pregnant women and newborns-the WHO vision. *BJOG.* 2015.

13. Mselle LT, Moland KM, Mvungi A, Evjen-Olsen B, Kohi TW. Why give birth in health facility? Users' and providers' accounts of poor quality of birth care in Tanzania. *BMC Health Serv Res.* 2013;13:174.

14. Afulani PA, Kirumbi L, Lyndon A. What makes or mars the facility-based childbirth experience: thematic analysis of women's childbirth experiences in western Kenya. *Reprod Health.* 2017;14(1):180.

15. McMahon SA, George AS, Chebet JJ, Mosha IH, Mpembeni RN, Winch PJ. Experiences of and responses to disrespectful maternity care and abuse during childbirth; a qualitative study with women and men in Morogoro Region, Tanzania. *BMC Pregnancy Childbirth.* 2014;14:268.

16. Bradley S, Kamwendo F, Chipeta E, Chimwaza W, de Pinho H, McAuliffe E. Too few staff, too many patients: a qualitative study of the impact on obstetric care providers and on quality of care in Malawi. *BMC Pregnancy Childbirth.* 2015;15:65.

17. Ganle JK, Parker M, Fitzpatrick R, et al. A qualitative study of health system barriers to accessibility and utilization of maternal and newborn healthcare services in Ghana after user-fee abolition. *BMC Pregnancy Childbirth.* 2014;14:425.

18. Maslach C, Jackson SE, Leiter MP. Maslach Burnout Inventory. In: Zalaquett C, Wood R, eds. *Evaluating Stress: A Book of Resources*. Lanhan, MD: The Scarecrow Press; 1997:191-218.

19. Wilkinson JM. Moral Distress in Nursing Practice: Experience and Effect. *Nurs Forum.* 1987;23:16-29.

20. Munabi-Babigumira S, Glenton C, Lewin S, Fretheim A, Nabudere H. Factors that influence the provision of intrapartum and postnatal care by skilled birth attendants in low- and middle-income countries: a qualitative evidence synthesis. *Cochrane Database Syst Rev.* 2017;11:CD011558.

21. Abuya T, Ndwiga C, Ritter J, et al. The effect of a multi-component intervention on disrespect and abuse during childbirth in Kenya. *BMC Pregnancy Childbirth.* 2015;15:224.

22. Kujawski S, Freedman L, Ramsey K, et al. Community and health system intervention to reduce disrespect and abuse during childbirth in Tanga Region, Tanzania: A comparative before-and-after study. *PLoS Med.* 2017;14.

23. UNAIDS., The Henry J. Kaiser Family Foundation. *Donor Government Funding for HIV in Low- and Middle-Income Countries in 2016.* Menlo Park, CA: UNAIDS and The Henry J. Kaiser Family Foundation; 2017.http://files.kff.org/attachment/Report-Donor-Government-Funding-for-HIV-in-Low-and-Middle-Income-Countries-in-2016

24. Brugha R, Simbaya J, Walsh A, Dicker P, Ndubani P. How HIV/AIDS scale-up has impacted on non- HIV priority services in Zambia. *BMC Public Health.* 2010;10:540.

25. Kruk ME, Jakubowski A, Rabkin M, Elul B, Friedman M, El-Sadr W. PEPFAR programs linked to more deliveries in health facilities by African women who are not infected with HIV. *Health Aff (Millwood).* 2012;31:1478-1488.

26. Kruk ME, Jakubowski A, Rabkin M, et al. Association Between HIV Programs and Quality of Maternal Health Inputs and Progresses in Kenya. *Am J Public Health.* 2015;105:S207-210.

27. Ministry of Health (MoH) [Malawi]., ICF International. *Malawi Service Provision Assessment (MSPA) 2013-14.* Lilongwe, Malawi and Rockville, Maryland: MoH and ICF International; 2014

28. World Health Organization. *Service Availability and Readiness Asessment (SARA): An Annual Monitoring System for Service Delivery.* Geneva, Switzerland: World Health Organization; 2015

29. World Health Organization., USAID. *Measuring Service Availability and Readiness: A Health Facility Assessment Methodology for Monitoring Health System Strengthening.* Geneva, Switzerland: World Health Organization; 2012

30. Leslie HH, Fink G, Nsona H, Kruk ME. Obstetric Facility Quality and Newborn Mortality in Malawi: A Cross-Sectional Study. *PLoS Med.* 2016;13:e1002151.

31. World Health Organization. *WHO Safe Childbirth Checklist.* Geneva: World Health Organization; 2015

32. Kruk ME, Hermosilla S, Larson E, Mbaruku GM. Bypassing primary care clinics for childbirth: a cross-sectional study in the Pwani region, United Republic of Tanzania. *Bull World Health Organ.* 2014;92:246-253.

33. Sando D, Kendall T, Lyatuu G, et al. Disrespect and abuse during childbirth in Tanzania: are women living with HIV more vulnerable? *J Acquir Immune Defic Syndr Hum Retrovirol.* 2014;67 Suppl 4:S228-234.

34. Turan JM, Miller S, Bukusi Ea, Sande J, Cohen CR. HIV/AIDS and maternity care in Kenya: how fears of stigma and discrimination affect uptake and provision of labor and delivery services. *AIDS Care.* 2008;20:938-945.

35. United Nations Department of Economic and Social Affairs. *Guidelines and principles for the development of disability statistics.* New York: United Nations; 2001

36. Kenny DA. MedPower App. https://davidakenny.shinyapps.io/MedPower/.

37. Hosmer D, Lemeshow S, Sturdivant R. *Applied Logistic Regression.* 3rd ed. Hoboken, NJ: Wiley; 2013.

38. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986;51:1173-1182.

39. Rasschaert F, Pirard M, Philips MP, et al. Positive spill-over effects of ART scale up on wider health systems development: Evidence from Ethiopia and Malawi. *J Int AIDS Soc.* 2011;14:1-11.

40. Leonard K, Masatu MC. Outpatient process quality evaluation and the Hawthorne Effect. *Soc Sci Med.* 2006;63:2330-2340.

41. Shiffman J. Has donor prioritization of HIV/AIDS displaced aid for other health issues? *Health Policy Plan.* 2007;23:95-100.

42. VanderWeele TJ. *Explanation in Causal Inference: Methods of Mediation and Interaction.* New York: Oxford University Press; 2015.

43. Fritz MS, MacKinnon DP. Required sample size to detect the mediated effect. *Psychol Sci.* 2007;18:233-239.

2.7. Tables and figures

Table 2.1. Items included in the measure of HIV structural inputs (distal structural inputs)

HIV testing and counseling
Trained staff in HIV testing and counseling (in-service training in last 24 months)
Condom availability
Availability of HIV rapid tests in HIV testing and counseling clinic
Availability of HIV testing and counseling guidelines
Visual and auditory privacy for HIV testing
Infection control where testing is done:
Running water and soap or alcohol-based hand disinfectant
Latex gloves
Sharps container
Waste receptacle
Antiretroviral Treatment (ART)
Trained staff on ART (in-service training in last 24 months)
ART guidelines
Laboratory diagnostic capacity for full blood count (observed and in working order)
Laboratory diagnostic capacity for CD4 test or viral load (observed and in working order)
Laboratory diagnostic capacity for renal or liver function test (observed and in working order)
First-line ARV treatment available
Prevention of mother to child transmission (PMTCT)
Trained staff in PMTCT (in-service training in last 24 months)
Trained staff in infant and young child feeding (in-service training in last 24 months)
Guidelines for PMTCT
Guidelines on infant and young child feeding
Visual and auditory privacy for HIV testing
Rapid tests available
Dried blood spot (DBS) testing available for infant diagnosis
Nevirapine (NVP) syrup for ARV prophylaxis
Zidovudine (AZT) syrup for ARV prophylaxis
PMTCT Option B+ regimen (TDF/3TC/EFV) in stock

Table 2.2. Items included in the measure of maternal health structural inputs (proximal structural inputs)

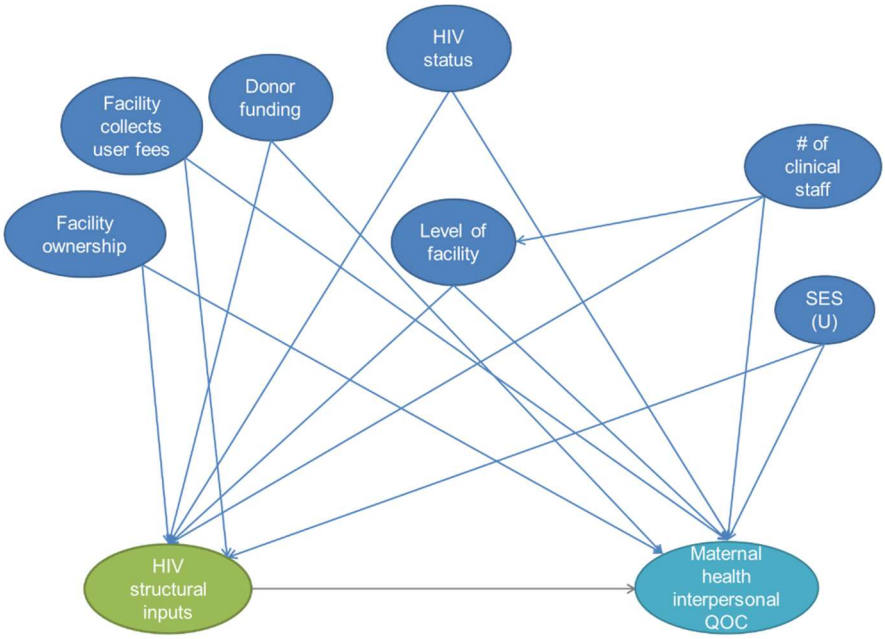
Drugs, supplies, and equipment	
Injectable magnesium sulfate	Examination light
Injectable antibiotics	Vacuum aspirator or dilation and curettage kit
Injectable uterotonic (Oxytocin)	Manual vacuum extractor
Injectable diazepam	Neonatal bag and mask
Intravenous fluids with infusion set	Suction apparatus (bulb or machine)
Delivery pack (scissors or blade, cord clamps or ties, episiotomy scissors, suture material with needle, needle holder)	Running water and soap or alcohol-based disinfectant
Skin disinfectant	Waste bin
Antibiotic eye treatment for newborns	Sharps container
Gloves	Syringes (single-use or auto-destruct)
Availability of partograph	Access to functioning ambulance
Electricity	Communication for referrals and consultations
Human resources	
In-service training for management of obstetric emergencies (in-service training in last 24 months)	
Organizational structure	
Duty schedule for 24-hour on-call or on-site staff	Up-to-date protocols for assessing intrapartum care
24-hour staff in facility or on-call	

Table 2.3. Items included in the measure of interpersonal quality of care during childbirth (outcome)

Initial client assessment
Respectfully greets pregnant woman
Encourages the woman to have a support person present during labor and birth
Asks woman (and support person) if she has any questions
Explains procedures to woman before proceeding
Informs woman of findings from initial examination
Care during labor
At least once, explains what will happen in labor
At least once, encourages woman to consume fluids/food during labor
At least once, encourages/assists woman to ambulate and assume different positions during labor
Drapes woman (one drape under buttocks, one over abdomen)
Explains procedures to woman before proceeding
Shouts, insults, or threatens woman during labor*
Slaps, hits, or pinches woman during labor*
*Reverse coded

Figure 2.1. Directed Acyclic Graphs (DAG)

A. Relationship between HIV structural inputs and interpersonal quality of care during childbirth



B. Relationship between maternal health structural inputs and interpersonal quality of care during childbirth

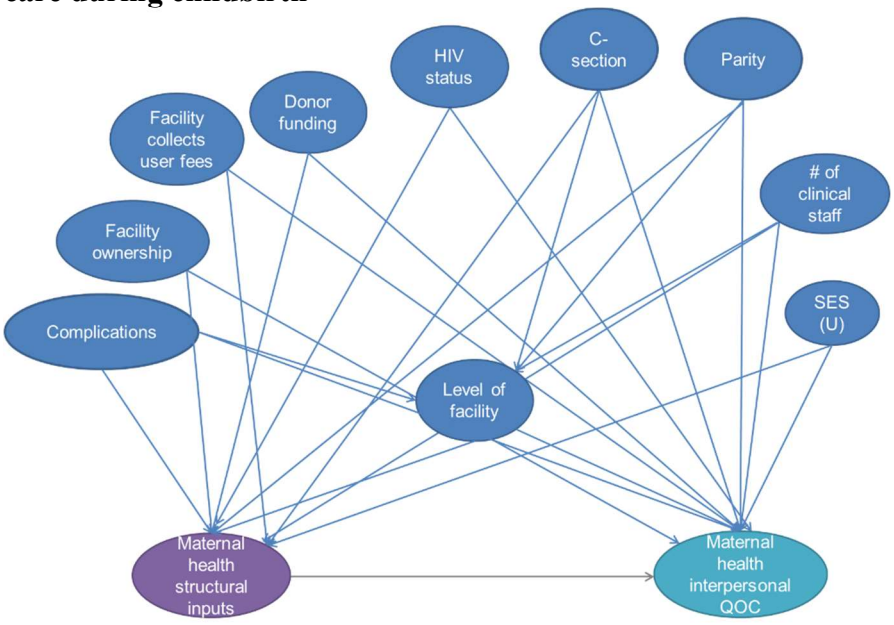


Figure 2.2. Traditional mediation model

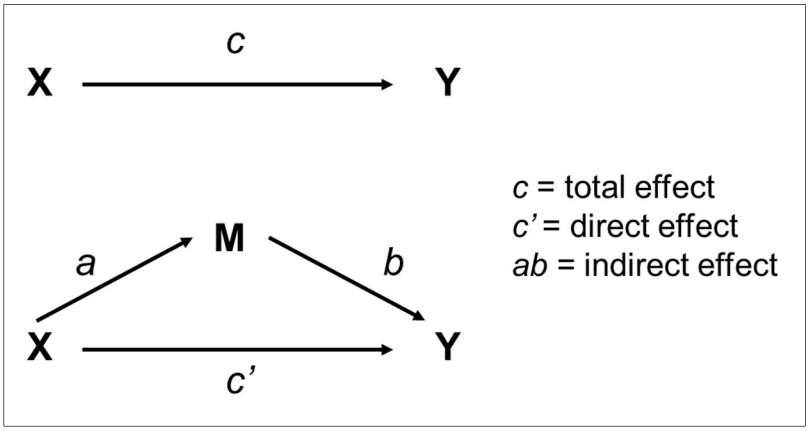


Table 2.4. Characteristics of all delivery facilities, delivery facilities without observations, delivery facilities with observations, and delivery facilities with observations that are included in the complete case sample, Malawi Service Provision Assessment 2013-2014

	All delivery facilities (N=540)		Delivery facilities without observations (N=318)		Delivery facilities with observations (N=222)			Delivery facilities with vs. without observations	Facilities with observations: Complete case (N=204)	
	N	%	N	%	N	%	% missing	p-value	N	%
Facility level										
Hospital	98	18.15	30	9.4	68	30.6	0.0	<0.0001	65	31.9
Health center or clinic	442	81.85	288	90.6	154	69.4			139	68.1
Facility ownership										
Public	356	66.5	202	64.5	154	69.4	0.0	0.24	140	68.6
Other (private non-profit, private for profit, company)	179	33.5	111	35.5	68	30.6			64	31.4
Facility receives donor funding	214	39.6	116	36.5	98	44.1	0.0	0.07	91	44.6
Facility charges user fees	199	36.9	114	35.9	85	38.3	0.0	0.56	81	39.7
Facility location										
Urban	80	14.8	36	11.3	44	19.8	0.0	0.01	43	21.1
Rural	460	85.2	282	88.7	178	80.2			161	78.9
Highest level clinician on site										
Doctor	65	12.0	21	6.6	44	19.8	0.0	<0.0001	43	21.1
Clinical officer or clinical technician	79	14.6	43	13.5	36	16.2		0.10	32	15.7
Medical assistant	348	64.4	219	68.9	129	58.1		0.18	116	56.9
Registered nurse	6	1.1	4	1.3	2	0.9		0.71	2	1.0
Enrolled nurse	41	7.6	30	9.4	11	5.0		0.55	11	5.4
Total clinical staff (mean, SD)	13.0	35.8	5.3	7.9	24.1	53.1		<0.0001	25.5	55.2
HIV structural inputs (mean, SD)	15.2	3.0	14.8	2.7	15.6	3.3	0.9	0.002	15.7	3.3
Maternal health structural inputs (mean, SD)	18.8	3.2	18.3	3.3	19.7	3.0	0.0	<0.0001	19.9	2.9

Table 2.5. Characteristics of women observed during labor and delivery at Malawi health facilities, Malawi Service Provision Assessment 2013-2014

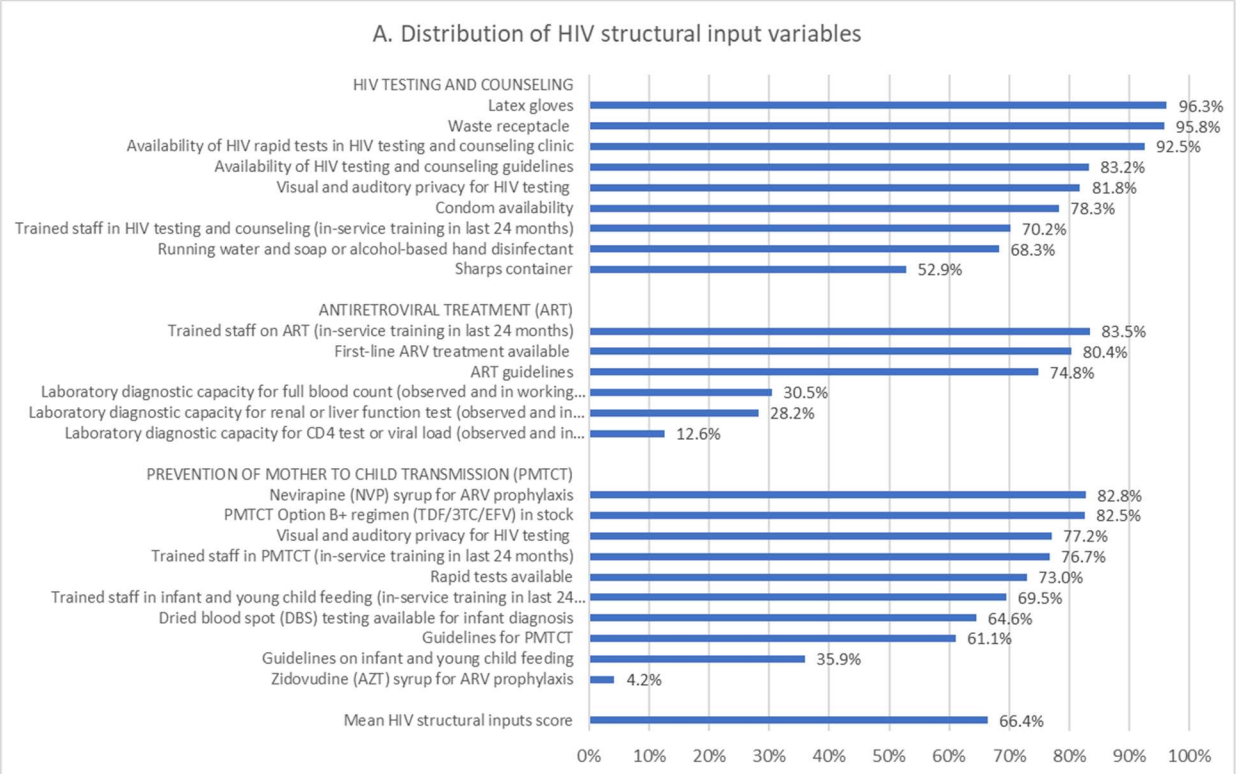
	Total sample (N=474)			Complete case (N=429)	
	N	%	% missing	N	%
Demographics					
Age (mean, SD)	25.0	6.6	0.0	25.1	6.4
Age categories					
15-19	107	22.6	0.0	95	22.1
20-34	319	67.3		289	67.4
35+	48	10.1		45	10.5
First birth	132	27.9	0.0	118	27.5
HIV-positive	30	6.3	0.0	25	5.8
Delivery characteristics					
Experienced any complication during childbirth[a]	42	8.9	0.0	40	9.3
Had a Caesarean section	18	3.8	0.0	17	4.0
Facility characteristics					
Facility level					
Hospital	229	48.3	0.0	209	48.7
Health center or clinic	245	51.7		220	51.3
Facility ownership					
Public	350	73.8	0.0	316	73.7
Other (private non-profit, private for profit, company)	124	26.2		113	26.3
Facility receives donor funding	232	49.0	0.0	210	49.0
Facility charges user fees	182	38.4	0.0	167	38.9
Facility location					
Urban	171	36.1	0.0	156	36.4
Rural	303	63.9		273	63.6
Highest level clinician on site					
Doctor	180	38.0	0.0	164	38.2
Clinical officer or clinical technician	79	16.7		69	16.1
Medical assistant	195	41.1		177	41.3
Registered nurse	3	0.6		3	0.7
Enrolled nurse	17	3.6		16	3.7
Total clinical staff (mean, SD)	42.2	65.9		43.2	67.2

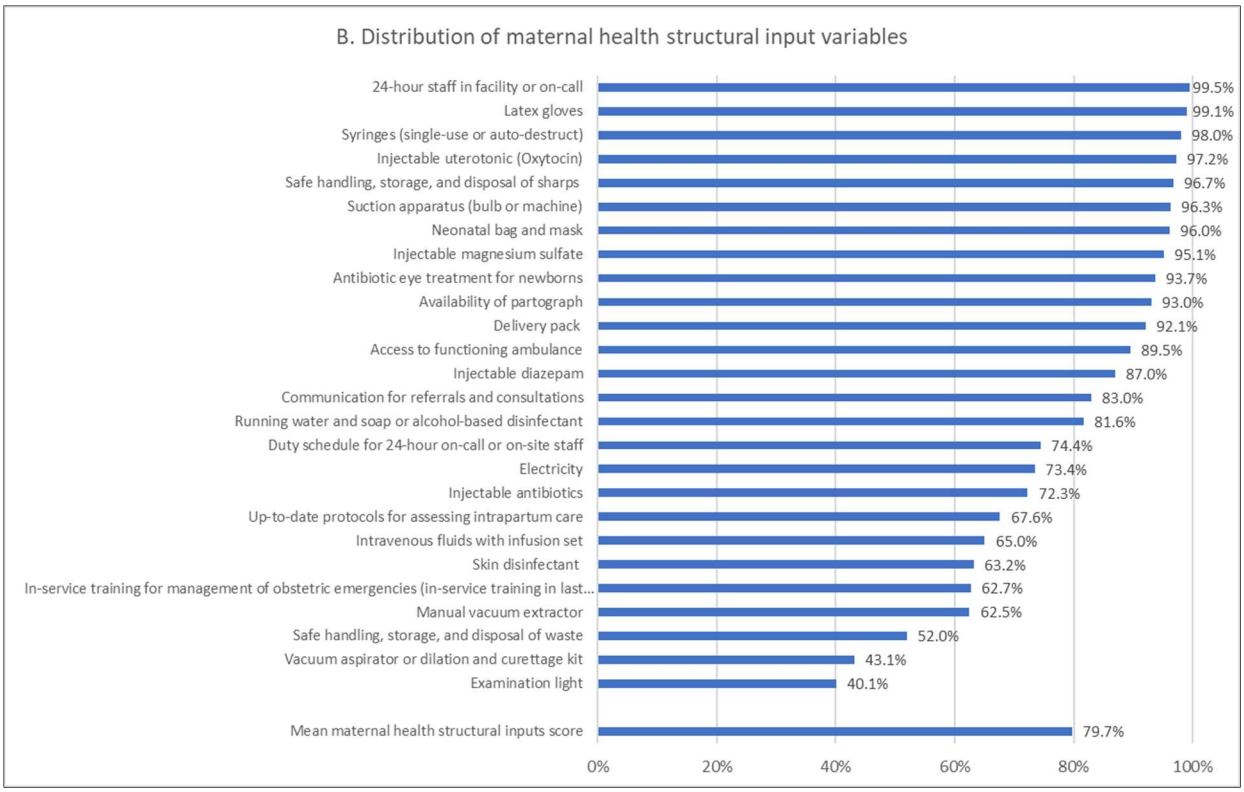
[a]Any complications during childbirth included eclampsia, major blood loss, fever, antibiotic use, or failure to progress/prolonged labor

Table 2.6. Distribution of HIV structural inputs, maternal health structural inputs, and interpersonal quality of care during childbirth among women observed during labor and delivery at Malawi health facilities, Malawi Service Provision Assessment 2013-2014

	Total sample (N=474)			Complete case (N=429)	
Distal exposure: HIV structural inputs (26 variables)	**N**	**%**	**% missing**	**N**	**%**
High HIV structural inputs (top 25%)	111	23.5		104	24.2
Low HIV structural inputs (bottom 75%)	361	76.5	0.4	325	75.8
Mean (SD)	16.5	3.6		16.6	3.6
Proximal exposure: Maternal health structural inputs (25 variables)					
High maternal structural inputs (top 25%)	90	19.0		83	19.4
Low maternal structural inputs (bottom 75%)	384	81.0	0.0	346	80.7
Mean (SD)	20.6	3.0		20.7	2.9
Outcome: Interpersonal quality of care (12 variables)					
Mean (SD)	9.0	1.8	9.0	9.0	1.8

Figure 2.3. Distribution of variables included in A. HIV structural inputs, B. maternal health structural inputs, and C. interpersonal quality of care in Malawi health facilities, Malawi Service Provision Assessment 2013-2014




B. Distribution of maternal health structural input variables
24-hour staff in facility or on-call 99.5%
Latex gloves 99.1%
Syringes (single-use or auto-destruct) 98.0%
Injectable uterotonic (Oxytocin) 97.2%
Safe handling, storage, and disposal of sharps 96.7%
Suction apparatus (bulb or machine) 96.3%
Neonatal bag and mask 96.0%
Injectable magnesium sulfate 95.1%
Antibiotic eye treatment for newborns 93.7%
Availability of partograph 93.0%
Delivery pack 92.1%
Access to functioning ambulance 89.5%
Injectable diazepam 87.0%
Communication for referrals and consultations 83.0%
Running water and soap or alcohol-based disinfectant 81.6%
Duty schedule for 24-hour on-call or on-site staff 74.4%
Electricity 73.4%
Injectable antibiotics 72.3%
Up-to-date protocols for assessing intrapartum care 67.6%
Intravenous fluids with infusion set 65.0%
Skin disinfectant 63.2%
In-service training for management of obstetric emergencies (in-service training in last... 62.7%
Manual vacuum extractor 62.5%
Safe handling, storage, and disposal of waste 52.0%
Vacuum aspirator or dilation and curettage kit 43.1%
Examination light 40.1%
Mean maternal health structural inputs score 79.7%
0% 20% 40% 60% 80% 100%

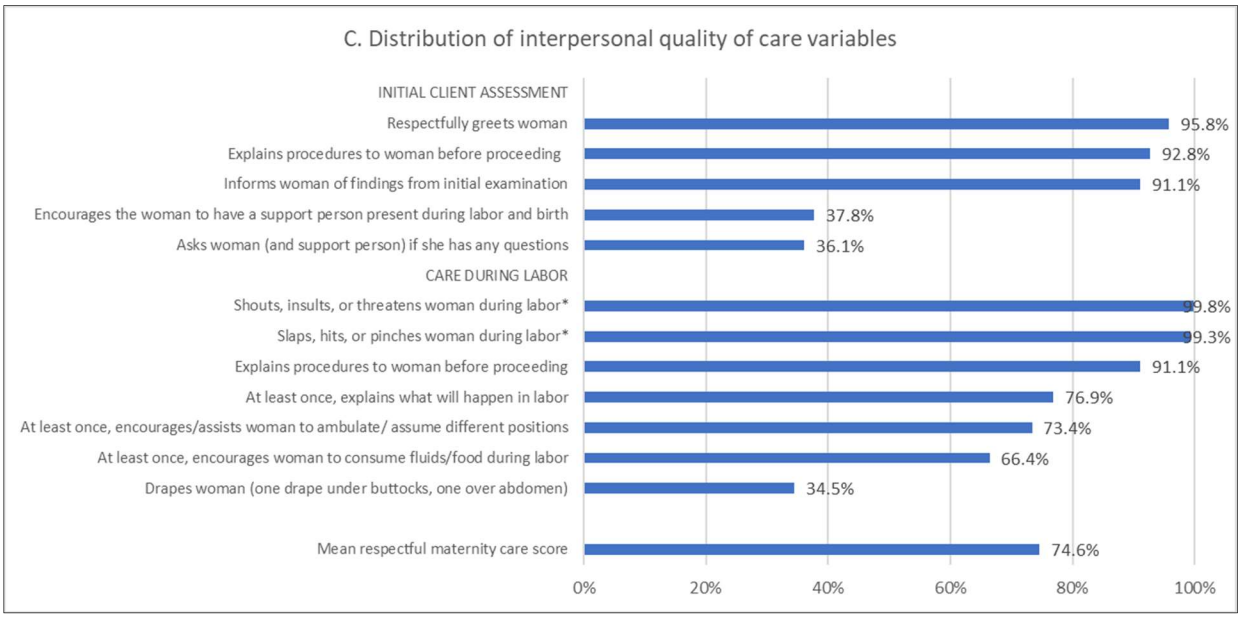

C. Distribution of interpersonal quality of care variables
INITIAL CLIENT ASSESSMENT
Respectfully greets woman 95.8%
Explains procedures to woman before proceeding 92.8%
Informs woman of findings from initial examination 91.1%
Encourages the woman to have a support person present during labor and birth 37.8%
Asks woman (and support person) if she has any questions 36.1%
CARE DURING LABOR
Shouts, insults, or threatens woman during labor* 99.8%
Slaps, hits, or pinches woman during labor* 99.3%
Explains procedures to woman before proceeding 91.1%
At least once, explains what will happen in labor 76.9%
At least once, encourages/assists woman to ambulate/ assume different positions 73.4%
At least once, encourages woman to consume fluids/food during labor 66.4%
Drapes woman (one drape under buttocks, one over abdomen) 34.5%
Mean respectful maternity care score 74.6%
0% 20% 40% 60% 80% 100%

Table 2.7. Associations of potential confounders with HIV structural inputs, maternal health structural inputs, and interpersonal quality of care

	Distal exposure: HIV structural inputs			Proximal exposure: Maternal health structural inputs			Outcome: Interpersonal quality of care		
	%	RR	95% CI	%	RR	95% CI	Mean (SE)	β	95% CI
Demographics									
First birth	-	-	-	37.4	**1.57**	**(1.12, 2.19)**	9.3 (0.1)	**0.52**	**(0.13, 0.92)**
HIV-positive	3.9	0.65	(0.275, 1.57)	6.0	1.04	(0.46, 2.35)	9.4 (0.3)	0.52	(-0.07, 1.11)
Delivery characteristics									
Experienced any complication during childbirth	-	-	-	19.3	**2.32**	**(1.26, 4.281)**	9.6 (0.2)	**0.72**	**(0.21, 1.22)**
Had a Caesarean section	-	-	-	12.1	**3.31**	**(1.82, 6.04)**	9.5 (0.4)	**0.60**	**(-0.22, 1.43)**
Facility characteristics									
Facility level									
Health center or clinic		Reference			Reference			Reference	
Hospital	86.5	**6.77**	**(2.50, 18.30)**	91.6	**11.43**	**(4.48, 29.17)**	8.9 (0.1)	-0.12	(-0.56, 0.32)
Facility ownership									
Other (private non-profit, private for profit, company)		Reference			Reference			Reference	
Public	87.5	**2.50**	**(0.90, 6.95)**	77.1	1.20	(0.47, 3.07)	8.9 (0.1)	-0.13	(-0.57, 0.31)
Facility receives donor funding	60.6	1.60	(0.72, 3.58)	67.5	2.16	(0.82, 5.68)	8.9 (0.1)	-0.007	(-0.45, 0.43)
Facility collects user fees	32.7	0.76	(0.31, 1.85)	57.8	**2.15**	**(0.91, 5.06)**	9.1 (0.1)	**0.32**	**(-0.12, 0.75)**
# of clinical staff		**1.00**	**(1.00, 1.01)**		**1.00**	**(1.00, 1.01)**		**0.003**	**(-0.00, 0.01)**

Bolded variables met p<0.20 criteria

Table 2.8. Regression models for the effects of distal and proximal structural inputs on interpersonal quality of care during childbirth

	Unadjusted model		Adjusted Model	
	β	**95% CI**	**β**	**95% CI**
Distal structural effects				
High HIV structural inputs	-0.04	(-0.64, 0.57)	-0.15	(-0.75, 0.46)
# of clinical staff			0.003	(0.00, 0.01)
Proximal structural effects				
High maternal health structural inputs	0.24	(-0.40, 0.88)	-0.06	(-0.70, 0.58)
Primipara			0.49	(0.09, 0.90)
Experienced any complication during childbirth			0.62	(0.04, 1.19)
Had a Caesarean section			0.04	(-0.76, 084)
Facility charges user fees			0.29	(-0.14, 0.71)
# of clinical staff			0.002	(-0.00, 0.00)

Table 2.9. Mediation analysis

	Unadjusted model		Adjusted Model	
	RR	**95% CI**	**RR**	**95% CI**
Path *a*: HIV structural inputs on maternal health structural inputs				
High HIV structural inputs	6.48	(3.06, 13.73)	3.44	(1.56, 7.57)
Hospital (vs. health center)			5.51	(1.80, 16.80)
# of clinical staff			1.00	(1.00, 1.01)
	β	**95% CI**	**β**	**95% CI**
Mediation Model				
High HIV structural inputs	-0.19	(-0.94, 0.55)	-0.20	(-0.89, 0.50)
High maternal health structural inputs	0.34	(-0.42, 1.11)	0.04	(-0.69, 0.77)
Primipara			0.49	(0.09, 0.90)
Experienced any complication during childbirth			0.63	(0.05, 1.20)
Had a Caesarean section			0.02	(-0.77, 0.80)
Facility charges user fees			0.26	(-0.16, 0.67)
# of clinical staff			0.002	(-0.00, 0.00)

Table 2.10. Sensitivity analyses for interpersonal quality of care analyses

Sensitivity 1: exposure and mediator as top 10% of distribution (N=429)

	Unadjusted model		Adjusted Model[a]	
Distal structural effects	**β**	**95% CI**	**β**	**95% CI**
High HIV structural inputs	0.001	(-0.70, 0.70)	-0.12	(-0.82, 0.59)
Proximal structural effects	**β**	**95% CI**	**β**	**95% CI**
High maternal health structural inputs	-0.09	(-1.16, 0.97)	-0.41	(-1.45, 0.64)
Path *a*: HIV structural inputs on maternal health structural inputs	**RR**	**95% CI**	**RR**	**95% CI**
High HIV structural inputs	6.01	(1.47, 24.61)	2.94	(0.71, 12.25)
Mediation Model	**β**	**95% CI**	**β**	**95% CI**
High HIV structural inputs	0.03	(-0.69, 0.76)	-0.10	(-0.73, 0.53)
High maternal health structural inputs	-0.10	(-1.22, 1.02)	-0.38	(-1.46, 0.70)

Sensitivity 2 & 3: Single imputation for missing outcome data (N=472): worst-case scenario and best-case scenario

	Sensitivity 2: worst-case scenario				Sensitivity 3: best-case scenario			
	Unadjusted model		Adjusted Model[a]		Unadjusted model		Adjusted Model[a]	
Distal structural effects	**β**	**95% CI**	**β**	**95% CI**	**β**	**95% CI**	**β**	**95% CI**
High HIV structural inputs	0.16	(-0.42, 0.74)	0.07	(-0.51, 0.66)	-0.10	(-0.69, 0.49)	-0.23	(-0.82, 0.37)
Proximal structural effects	**β**	**95% CI**	**β**	**95% CI**	**β**	**95% CI**	**β**	**95% CI**
High maternal health structural inputs	0.34	(-0.26, 0.94)	0.09	(-0.03, 0.75)	0.22	(-0.43, 0.88)	-0.08	(-0.73, 0.57)
Path *a*: HIV structural inputs on maternal health structural inputs	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**
High HIV structural inputs	6.84	(3.22, 14.51)	3.80	(1.71, 8.41)	6.84	(3.22, 14.51)	3.80	(1.71, 8.41)
Mediation Model	**β**	**95% CI**	**β**	**95% CI**	**β**	**95% CI**	**β**	**95% CI**
High HIV structural inputs	0.004	(-0.70, 0.71)	0.01	(-0.66, 0.69)	-0.28	(-1.00, 0.45)	-0.28	(-0.95, 0.39)
High maternal health structural inputs	0.34	(-0.375, 1.05)	0.09	(-0.639, 0.80)	0.38	(-0.39, 1.15)	0.07	(-0.65, 0.80)

[a] Distal structural effects model adjusted for number of clinical staff; Proximal structural effects model adjusted for parity, any complication during childbirth, Caesarean section, facility charges user fees, number of clinical staff; Path *a* adjusted for facility level, number of clinical staff; mediation model adjusted for parity, any complication during childbirth, Caesarean section, facility charges user fees, number of clinical staff

2.8. Appendices

Appendix 2.1. Creation of exposures variables and missing data

The exposure variables were dichotomized—top 25% vs. bottom 75%—based on the distribution of the data. I chose to create the exposure variables based on the full distribution of data available and then drop the missing data for the complete case analysis, rather than vice versa. I think this better represents the true distribution of the variables, as it was based on all available data, compared to a possibly skewed distribution based on the complete case data.

Appendix 2.2. Mediation power analysis accounting for design effect

The current methodologies for determining power and sample size for mediation analyses are not well developed. VanderWeele notes the limitations of current methods in his recent textbook on mediation[42] and suggests using Fritz and MacKinnon's 2007 method, which calculates minimum sample sizes needed with a power of 0.80 and alpha of 0.05 based on various combinations of pre-specified small, medium, and large effect sizes of the *a*, *b*, and *c'* pathways (Figure 2.2).[43] This paper, however, does not allow the researcher to specify the sample size, power, or the size of the effects, but can act as a guide for approximate sample size given the combinations provided.[42] Recently, Kenny developed an app to calculate power or sample size for mediation given any input values for the effect sizes of the *a*, *b*, and *c'* pathways. I therefore used Kenny's app (MedPower: https://davidakenny.shinyapps.io/MedPower/)[36] to calculate power given a specified sample size, accounting for design effect, for mediation analyses to determine the minimal detectable effect size for the direct and the indirect effect of HIV structural inputs on maternal health interpersonal quality of care, mediated by maternal health structural inputs.

Per Cohen's standards, power for both scenarios correspond to a medium path *a* when X is dichotomous (0.16-0.18), and a small-medium path *b* effect size when M is dichotomous and Y is continuous (0.18). These calculations roughly correspond to a minimum sample size between 224 and 427 in Fritz and MacKinnon when the direct effect=0.14, path *a*=0.14-0.26 and path *b*=0.26.[43]

Calculation of design effect Design effect formula: 1 + icc(n-1) Where icc = intraclass correlation coefficient; n = average number of observations per cluster		
	Full sample size (N=474)	**Complete case (N=429)**
# of clusters	222	204
Average # of observations per cluster	2.1	2.1
ICC	0.32	0.32
Design effect	1.36	1.35
Sample size account for design effect	349	318

Power analysis based on design effect sample size				
	N=349		**N=318**	
	Minimum detectable effect size	Power	Minimum detectable effect size	Power
path *a*	0.16	0.85	0.18	0.90
path *b*	0.18	0.93	0.18	0.90
Direct effect (path *c′*)	0.15	0.81	0.16	0.82
Indirect effect	0.029	0.79	0.032	0.81
Total effect (path *c*)	0.18	0.92	0.19	0.94

Chapter 3: Interpersonal quality of care: A systematic review of the reliability and validity of instruments to measure the construct in health care settings

3.1. Abstract

Background: In the past few years, the maternal health field has increasingly focused on interpersonal quality of care as an important aspect of the childbirth experience at health facilities. There is a growing need to develop interventions and to measure and evaluate the effects of interpersonal quality of care. However, the quality and scope of current instruments to measure this construct is unknown. The purpose of this paper was to systematically review the literature on the instruments measuring the construct of interpersonal quality of care in health care settings and their reliability, validity, and dimensionality.

Methods: Ovid MEDLINE, Health and Psychosocial Instruments database, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL) were searched to identify peer-reviewed articles published between January 1, 1988-December 31, 2017 that included the development or validation of an instrument to measure interpersonal quality of care. I evaluated the quality of the results for content validity, internal consistency, test-retest reliability, and structural validity for each included instrument. Methodological quality of the studies was also assessed for each psychometric domain.

Results: Twenty-six articles representing 27 instruments were included in the review. The 27 instruments measured 16 constructs within the domain of interpersonal quality of care. The factor structure of the instruments ranged from one to eight dimensions. While the majority of instruments received high ratings for content validity and internal consistency, about a third had poor content validity or did not meet the minimum Cronbach's alpha value for adequate reliability (≥0.70). Seven instruments evaluated test-retest reliability, but only two had an

adequate correlation coefficient value, and methodological quality was low. Twenty-three instruments were factor analyzed, with 13 instruments receiving positive ratings for structural validity, but most of the validation studies were not methodologically robust. Instruments that were self-completed had higher internal consistency and factor loadings than those that were interview-administered. Context did not contribute to the variability in construct definitions or factor structure, with the exception of instruments measuring interpersonal quality of care during labor and delivery.

Conclusions: This review demonstrates that the construct of interpersonal quality of care is not well-defined. Few instruments met psychometric standards of adequate reliability and validity, and the methodological quality of the studies was generally poor, limiting the ability to draw confident recommendations about measurement tools for interpersonal quality of care. Future research should aim to create a unifying definition of interpersonal quality of care, followed by the development of instruments using rigorous psychometric methods.

3.2. Introduction

Due to the failure to meet Millennium Development Goal 5, to reduce maternal mortality by 75%, the maternal health field has recently increased its focus on quality improvement efforts, with particular attention to interpersonal quality of care during childbirth.[1,2] This mirrors the growing awareness outside of maternal health of the importance of interpersonal quality in the provision of health care. For example, the Institute of Medicine and the World Health Organization included patient-centered care—care that is respectful and responsive to individual needs—as a key component for improving quality of care in the United States and globally.[3,4]

There is currently no consensus on the definition of interpersonal quality of care, and it is commonly defined as the relationship between the patient and the provider, specifically in relation to aspects of communication, support, and respect for patients' preferences.[1,5-7] Interpersonal quality of care is related to person-centered care and individualized care, which focus on care that is guided by patient's individual needs and preferences.[8-10] Specific to maternal health, the World Health Organization's quality of care framework for maternal and newborn health defines three key interpersonal quality of care components as effective communication, respect and dignity, and emotional support.[1] However, these constructs are distinguished from patient satisfaction and generalized measures of quality, which often reflect the totality of the patient experience and patient expectations of care.[6]

To meet the demand to improve interpersonal quality of care in maternal health, there is an increasing need to develop interventions and to measure and evaluate the effects of interpersonal quality of care. This requires clarity of the construct and reliable and valid instruments to

monitor quality improvements and intervention success. To understand the quality and breadth of current instruments, a review of the literature on measures of interpersonal quality of care relevant to the maternal health context is necessary. While there are systematic reviews of interpersonal quality of care instruments, they are too narrow in scope or do not encompass the full definition of interpersonal quality of care. For example, previous reviews focused on one aspect of interpersonal quality, like communication skills[11] or person-centered care,[12] or only included nursing care[13] or specific health care settings.[14]

Therefore, the purpose of this paper is to systematically review the literature on the instruments measuring the construct of interpersonal quality of care in health care settings to assess their reliability, validity, and dimensionality.

3.3. Methods

Eligibility criteria

All articles that included the development or validation of an instrument to measure interpersonal quality of care were eligible for inclusion. Based on the literature, I defined interpersonal quality of care as the relationship between the patient and the provider, specifically in relation to aspects of communication, support, and respect for patients' preferences.[1,5-7] Articles were included if they were peer-reviewed, conducted with adult subjects, published in English, and published between January 1, 1988 and December 31, 2017. The year 1988 was chosen as a lower limit to correspond to the publishing of a prominent quality of care framework by Avedis Donabedian, which integrated interpersonal quality of care as a main component of health care quality.[5] Commentaries, letters to the editor, unpublished manuscripts, and conference abstracts were

excluded, as were articles that tested the performance in a new language of an already validated instrument. The instrument also had to measure interpersonal quality of care from the patient's perspective; instruments using only provider, family, or third-party perspectives were excluded.

To ensure the articles reviewed only measured interpersonal quality of care and were distinct from similar constructs, I imposed additional exclusions. First, in order to determine the dimensionality of the interpersonal quality of care construct, I excluded studies that captured interpersonal quality of care in a subscale of a broader construct, like the general patient experience or perceptions of quality or satisfaction, or studies that included measures of clinical quality. Second, as I was interested in the full construct of interpersonal quality, rather than specific domains, I also excluded studies that only measured communication. Third, I excluded studies that measured interpersonal quality in nursing homes, care for those with impaired cognition, or palliative care. These patient populations and settings often involve different interpersonal needs, including long-term, live-in care and dependency on the caregiver.[14]

Sources and search strategy

I searched Ovid MEDLINE, Health and Psychosocial Instruments database, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). Reference lists of articles selected for review were searched for additional articles. I used keywords and Medical Subject Heading terms (MeSH) that included the following: 1) interpersonal quality of care (e.g. patient-centered care, person-centered care, professional-patient relations, patient-provider relationship, respectful care, quality of care, disrespect, abuse, dignified, dignity, patient experience, interpersonal quality, quality of care) and 2) measurement terms (e.g. instrument, index, scale, measurement,

measure, weights and measures, surveys and questionnaires, psychometrics, validation studies). In Ovid MEDLINE, the search was limited to those articles categorized as validation studies. In CINAHL, the search was limited to "research", "questionnaire/scale", and "research instrument" publication types. As in-process citations in Ovid MEDLINE are not yet tagged to MeSH terms or to certain limiters, I repeated the Ovid MEDLINE search for the year 2017 without the limiters for publication type. See Table 3.1 for the full search strategy used in each database.

Study selection

Articles were screened by title based on eligibility criteria and imported into Mendeley reference software. Two independent reviewers then screened the article abstracts based on inclusion criteria (SK and SS[a]). The two reviewers discussed any discrepancies in the categorization of abstracts until consensus was reached. When eligibility was unclear from the abstract, the full text was reviewed. The full text of articles retained by the abstract review were then read by SK to determine final inclusion.

Data collection process and data synthesis

Data were extracted from each article using an Excel spreadsheet. In the event that an article validated multiple instruments, each instrument was considered separately and treated as a separate validation study. In addition to the title of the instrument, author, and year of publication, I extracted the following information: construct definitions, context in which the instruments were validated (country, patient population, health care setting), methods (sample size, mode of instrument administration, response scale of the instrument, reliability and validity

[a] Stephanie Kujawski (SK) and Suzue Saito (SS)

methods), and results (content validity, internal consistency, test-retest reliability, number of items included in the final scale, structural validity, dimensions and number of items per dimension, fit statistics).

Synthesis of results

For each included instrument, where available, I evaluated the results of the content validity, internal consistency, test-retest reliability, and structural validity. The quality of the instrument's content validity, internal consistency, and test-retest reliability was assessed using the quality criteria for measurement properties developed by Terwee et al.[15] The Terwee et al. criteria use the following rating scale: positive, indeterminate, negative, or no information available. As existing guidelines for assessing structural validity are dated and include some subjective principles,[16,17] I developed criteria to evaluate instrument quality for this measurement property, guided by current structural validity standards and in consultation with a psychometrician.[18-21] Each included item for structural validity was rated as positive, indeterminate, negative, or no information available. Table 3.2 provides a list of all instrument quality criteria and the rating scales for each domain.

To determine if context contributed to any variability in the construct definitions, results were compared by country settings and within comparable health care settings. It is plausible that instruments developed and measured in similar contexts would have more consistent construct definitions or instrument dimensions. Countries were divided into high income vs. middle- or low-income countries for this comparison based on World Bank designations.[22] To assess

whether data collection methods could explain variation in instrument quality, results were compared by mode of instrument administration (self-completed vs. interview-administered).

Quality of the reliability and validity studies

After evaluating the reliability and validity of the instruments, I examined the methodological quality of each study to determine the level of confidence in the results. The quality of each study's internal consistency and test-retest reliability methods were guided by the COSMIN checklist, a standardized tool developed for measuring the quality of health measurement studies in systematic reviews.[23] As some items in the COSMIN checklist are not well-defined or are subjective in nature, I adapted the checklist for this review. The COSMIN checklist uses a 4-point scale: excellent, very good, fair, or poor. A score for each measurement property is determined by the lowest rated criterion in the checklist category. For structural validity, I developed criteria to assess the quality of each study based on current best practices for factor analysis.[17-20,23,24] Table 3.3 provides a list of all study quality criteria by domain, the rating scales for each domain, and citations for the criteria.

3.4. Results

Study selection

Figure 3.1 details the article selection procedure. A total of 9,155 articles were identified through the database search, with 3,268 from Ovid MEDLINE, 918 from Health and Psychosocial Instruments, and 4,969 from CINAHL. An additional 20 articles were found through searching the reference lists of included studies. After removing duplicates (N=26), the remaining 9,149 articles were reviewed by title, and 8,943 were excluded mainly due to wrong subject matter.

Abstracts were reviewed for 206 articles and 90 were excluded. The full text of 116 articles were reviewed. Ninety articles were excluded for the following reasons: provider perspective (N=25), wrong construct (N=15), validation of existing instrument in another language (N=12), nursing home or elderly care (N=8), not a validation study (N=9), contains aspects of clinical quality (N=6), third-party assessment (N=5), non-adult population (N=4), family perspective (N=3), and systematic review (N=3). Twenty-six articles were included for review, which accounted for 27 instruments and 29 validation studies.[6,25-49] Two instruments had two separate articles validating their structure.[34,35,48,49] One article, van der Kooy et al., presented the validation of an instrument in the same population for three different aspects of perinatal care: antenatal, labor and delivery, and postpartum.[47] As no information was provided in the text as to the validation of the total instrument across the three health care settings, each instrument was reviewed separately.

Characteristics of the included instruments

Constructs and construct definitions

Instruments were included if they met the definition of interpersonal quality of care, as outlined in the methods, as the relationship between the patient and the provider, specifically in relation to aspects of communication, support, and respect for patients' preferences. The 27 instruments measured 16 constructs under this umbrella of interpersonal quality of care. While many of the instruments measured constructs that were not explicitly named "interpersonal quality of care," there were commonalities across the construct names and definitions that indicated the construct of interpersonal quality of care (Table 3.4). Seven of the instruments measured constructs that were not defined in the articles.[26,33-36,38,39,46] Only three instruments labeled the construct with the word "interpersonal."[6,33,36] The instruments' construct names and definitions mainly fell into five

categories: patient-provider relationship (N=7),[26-28,32,34,35,38,46] therapeutic and caring interactions (N=5),[30,31,37,44,48,49] individual needs (N=5),[25,40,42,43] interpersonal processes or skills (N=5),[6,33,36,41] and responsiveness (N=4).[45,47] One instrument's construct, perceptions of maternity care, measured by the Perceptions of Care Adjective Checklist,[39] did not fit within a category based on its construct name or definition, but the items included in the tool reflected the construct of interpersonal quality of care. See Table 3.4 for a complete list of the constructs and definitions by category.

Country and health care settings

Table 3.5 provides the characteristics of the context in which each instrument was measured. The majority of the instruments (N=21, 77.8%) were validated in high-income settings,[6,27-30,32-39,41-43,46-49] five (18.5%) in middle- or low-income settings,[25,26,31,40,44] and one instrument was validated in a mix of 41 high-income and middle- or low-income countries.[45] The health care setting in which interpersonal quality of care was measured varied, with the majority in outpatient/primary care (N=8),[6,27,36,37,41,45,46] followed by labor and delivery (N=5),[25,39,40,44,47] hospital inpatient care (N=4),[30,42,43,48,49] antenatal care (N=2),[26,47] oncology departments (N=2),[34,35,38] dentistry (N=1),[33] HIV primary care (N=1),[28] intensive care unit (N=1),[31] postpartum care (N=1),[47] and rehabilitation services (N=1).[29] The health care setting was not provided for one instrument.[32]

Data collection and scales

Twelve instruments were self-completed,[27,29,30,32,33,39,41-43,46,48,49] nine were interviewed-administered,[6,25,26,28,40,45,47] and one was completed via the telephone or internet[36] (Table 3.6).

Five instruments did not specify the data collection method.[31,34,35,37,38,44] The majority of the instruments used either a 4-point or 5-point Likert scale for the responses to the scale items.[25-35,37,38,40,41,43-45,47-49] One instrument used a binary option (circle items that apply)[39] and three instruments did not specify the response scales.[6,42,46] Fifteen instruments contained positively-worded items only,[27,30-33,36,38,43,45-49] while 12 contained at least one negatively-worded item.[6,25,26,28,29,34,35,37,40-42,44]

Reliability and validity methods

Content validity was evaluated for 24 instruments. The majority of studies used a mix of literature reviews, expert review, and qualitative methods that involved the target population (e.g. focus group discussions, in-depth interviews, cognitive interviews) (Table 3.6).[6,25–32,34–38,40–42,44–49] Reliability was assessed for 26 instruments with a minimum of Cronbach's alpha, a measure of internal consistency.[6,25–49] Several instruments had their reliability assessed before and after factor analysis. Where available, the reliability estimates included in the table were those calculated after factor analysis. Test-retest reliability was also performed for seven instruments.[27–29,35,42,45,46] The retest period ranged from three days to two months. Test-retest reliability was estimated with different methods: intraclass correlation coefficient (N=1),[29] Kappa (N=1),[45] and Pearson correlation coefficient (N=5).[27,28,35,42,46]

Twenty-three of the 27 instruments were validated using a factor analysis technique.[25–28,30–32,34,36,38–49] Seven instruments were validated both via a data-driven method (exploratory factor analysis or principal component analysis) and by using confirmatory factor analysis to verify the factor structure.[25,32,34–36,43,44,48,49] The remaining 16 instruments were validated with just one

technique: seven used principal component analysis,[26,28,30,38,40,42,46] six used exploratory factor analysis,[27,31,45,47] and three used confirmatory factor analysis.[39,41]

Reliability and validity results

Table 3.7 presents the results of the instruments and Table 3.8 provides the quality ratings of the results by content validity, internal reliability, test-retest reliability, and structural validity.

Content validity

The content validity of each instrument was assessed using Terwee et al.'s criteria.[15] Fifteen out of the 24 instruments that assessed content validity received positive ratings.[25,27–29,31,32,37,38,40,45,47] These instruments all had a clear description of the measurement aim, target population, the concepts being measured, how the items were selected, and included in the instrument development both the target population and experts. Six instruments received negative ratings, as they did not include the target population in the item selection process.[6,30,42,44,46,48,49] Three instruments received indeterminate ratings, mainly due to the lack clarity in the concepts being measured.[26,34–36] Two instruments did not have any information about their content validity,[33,39] and one instrument referenced the content validity of an earlier iteration of the instrument.[43]

Reliability: internal consistency

A scale is considered adequately reliable if Cronbach's alpha is at least 0.70 for each of its factors.[21] Sixteen of the 26 instruments that had their internal consistency assessed met this criterion; however, 10 instruments had a Cronbach's alpha < 0.70 for at least one factor.[6,25,31,40,45,47,49] Of these 10 instruments, the factors with inadequate internal consistency had

less than four items, which is not surprising since Cronbach's alpha is influenced by the number of items.

Reliability: test-rest reliability

As with Cronbach's alpha, adequate test-retest reliability is a minimum of 0.70.[15] Only two of the seven instruments, the Client-Centered Rehabilitation Questionnaire and the Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD), met this minimum value.[29,35]

Structural validity

Dimensions

Across the 23 instruments that were factor analyzed, the number of dimensions (factors) ranged from one to eight. Six instruments were unidimensional, five of which measured the patient-provider relationship/interaction.[26,27,32,38,46] For the instruments with more than one dimension (N=17), similar factors emerged.[25,28,30,31,34,39–45,47–49] Thirteen instruments had a factor representing respectful or compassionate care,[25,30,31,34,35,39–41,44,47–49] 10 had a communication factor,[25,28,34,41,43,45,47–49] and nine had a factor related to decision-making or patient involvement.[25,30,31,42,43,45,47]

Two validations of the same instrument, the Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD), yielded different factor structures.[34,35] The original validation using principal component analysis extracted four factors, while the subsequent validation, using exploratory factor analysis followed by a confirmatory factor analysis, showed a two-factor

structure. This inconsistency was likely due to the different methods used; principal component analysis does not account for unique variance, while exploratory factor analysis does, which affects the percent of variance explained.[18]

Validation with Exploratory factor analysis or principal component analysis and confirmatory factor analysis

Seven instruments were validated both using both exploratory and confirmatory methods. Two instruments, the Scale of Supportive Care Given During Labor and the Patient Evaluation of Emotional Care during Hospitalization (PEECH), had sufficient model fit and factor loadings of at least 0.30 (the minimum correlation that is indicative that the item is a good reflection of the underlying construct), loadings that were statistically significant, or loadings that were < 0.30 but were justified by the authors to retain them.[44,48,49] The five other instruments either did not have sufficient fit, did not meet the factor loading cut off for both the exploratory and confirmatory methods, or I was unable to assess the results as no information was provided.[25,32,34–36,43] None of the instruments had Heywood cases (factor loadings > 1.0), an indication that the model is overfit.

Validation with Exploratory factor analysis or principal component analysis only

Of the 13 instruments that were validated with either exploratory factor analysis or principal component analysis only, nine instruments had factor loadings that met one of the following criteria: at least loadings of 0.30, loadings were statistically significant, or loadings were < 0.30 but the authors discussed the low factor loadings and justified the decision to retain them.[26–28,30,31,40,42,46,47] Three instruments did not provide any information on the factor loadings,[38,47] and, two of the 13 instruments had Heywood cases (factor loadings > 1.0).[45,47] As none of the

instruments that used exploratory factor analysis with maximum likelihood provided fit statistics, I was unable to assess the quality of the model fit.

Validation with Confirmatory factor analysis only

Of the three instruments that were validated via confirmatory factor analysis only, two had factor loadings ≥ 0.30 (the third did not provide factor loadings),[39,41] and all three met the cut offs for sufficient model fit.[39,41] There were no Heywood cases in the confirmatory factor analyses.

Sources of heterogeneity

High-income vs. middle- or low-income countries

As the search discovered that there was variability in the construct definitions, I assessed whether contextual factors, specifically country settings and health care settings, could explain this variation. When comparing instruments within high-income countries and within middle- or low-income countries, no specific trends were identified regarding the constructs assessed or the numbers or types of dimensions.

Health care settings

Instruments that were validated in the same health care settings were also compared. For outpatient/primary care, antenatal care, and inpatient hospital care, no specific patterns were detected. For the two instruments that measured interpersonal care in oncology departments, both measured similar constructs—the patient-provider relationship—but had different factor structures.[34,35,38]

The most similar tools were the five instruments measuring interpersonal quality of care during labor and delivery.[25,39,40,44,47] All of the instruments included a dimension of respect or compassionate care. While the number of factors varied from two to eight, three of the instruments, the Respectful Maternity Scale, Perceptions of Care Adjective List, and the Scale of Supportive Care Given During Labor, all included one factor with positive care attributes and one factor with negative care attributes.[39,40,44] For example, the Scale of Support Care Given During Labor had one factor that measured comfortable behaviors and one that measured disturbing behaviors.[44] Similarly, the Respectful Maternity Care scale included a friendly care factor and an abuse-free care factor (items were reverse coded).[40]

Data collection

I compared instruments that were self-completed vs. interview-administered to determine if the mode of instrument administration impacted the quality of the validity and reliability results. A higher proportion of instruments that were interview-administered had poor internal consistency and lower factor loadings than instruments that were self-completed.

Methodological quality of the included studies

Reliability: internal consistency

Table 3.9 summarizes the methodological quality of the studies. Using the modified COSMIN checklist for internal consistency quality, 22 of the 29 studies received an excellent rating,[25–27,30–32,34–36,38–44,46–49] one received a good rating,[28] and five received a poor rating.[6,29,33,37,45] The five that received a poor rating failed to perform factor analysis as a prerequisite for testing internal consistency.[6,29,33,37,45] One study did not assess the internal consistency of the instrument.[41]

Reliability: test-retest reliability

The quality ratings of test-retest reliability ranged from good to poor. Reasons for these low-quality ratings included small sample sizes,[28,42,46] the use of inferior statistical methods to calculate the test-retest reliability (e.g. Pearson correlation coefficient instead of Kappa or intraclass correlation coefficient),[27,28,35,42,46] the omission of details regarding the stability of the participant populations over time,[35,42,45,46] and different test conditions between the two measurements.[27,35,43]

Structural validity

Validation with exploratory factor analysis or principal component analysis

Twenty-one studies performed either an exploratory factor analysis (N=11)[25,27,31,32,35,44,45,47,48] or a principal component analysis (N=10).[26,28,30,34,36,38,40,42,43,46] I assessed whether the studies provided adequate justification for their choice of exploratory method. If the aim of the analysis was item reduction, principal component analysis was appropriate. If the aim was to explain correlations between indicators or identify latent factors, exploratory factor analysis was the appropriate approach.[18] In choosing between the two exploratory methods, eight provided adequate justification for their choice of analysis,[25,27,30–32,35,44,45] nine provided no justification,[36,42–44,46–48] and four studies provided inappropriate reasoning for their method choice.[26,28,38,40] These four used a principal component analysis and cited the goal of identifying or assessing factor structure, aims which would have been more appropriate for an exploratory factor analysis. None of the 21 studies used a power calculation to determine an appropriate sample size to achieve statistical power for their study. Two studies used rule of thumb methods

for determining sample size; however, rule of thumb techniques fail to account for study design, correlations, and model specifications (e.g. scaling, estimator type) that affect power.[19,28,40] About half of the studies described how missing data were handled.[25,27,28,30,32,34,38,45,47] While 18 of the 21 studies cited a rotation method (orthogonal or oblique) for their analysis, only three provided a rationale for their choice.[25,31,40] Sixteen studies presented the factor loadings of their model,[25–28,30–32,35,40,42–48] but goodness of fit tests, appropriate for assessing the model fit for exploratory factor analyses using maximum likelihood or weighted least squares estimation, were not presented in any study.

Validation with Confirmatory factor analysis

Ten studies performed confirmatory factor analyses.[25,32,35,36,39,41,43,44,49] Similar to the exploratory methods, no study used a power calculation to determine adequate sample size to achieve a desired level of power, one study used a rule of thumb method,[49] and 40% discussed how they handled missing data.[25,32,41] Three studies stated the estimator used,[39,41] four studies provided factor loadings,[39,41,44,49] and seven provided goodness of model fit statistics for the instruments.[32,35,39,41,44,49]

Six studies performed an exploratory method and confirmatory method on the same instrument in the same study. Three studies used different samples for each method,[32,35,36] as appropriate, while three used the same samples.[25,43,44] In two of the studies that used the same samples, the text suggested that the exploratory method was the primary focus of the study, and the confirmatory factor analysis was used to endorse their final exploratory model.[25,43]

3.5. Discussion

In a systematic review of published, peer-reviewed literature, I identified 27 validated instruments from 26 articles that measured interpersonal quality of care in health care settings from the patient's perspective. Three conclusions emerged from this review. First, the construct of interpersonal quality of care is not well defined. Second, there were few instruments that met both psychometric standards of adequate reliability and validity and methodological quality. The lack of study quality, particularly for structural validity, limits the ability to draw confident conclusions about this body of research. Third, heterogeneity in contextual factors and mode of administration explained some of the findings. Variability in construct definitions and instrument dimensions were not explained by country setting or health care setting, with the exception of the labor and delivery setting, while mode of administration affected the quality of the reliability and validity results.

As evidenced by this review, interpersonal quality of care is not a well-defined construct. Sixteen different constructs were measured in 27 instruments. While there was considerable variation in the constructs, five common themes emerged from the construct names and definitions, all within the boundaries of interpersonal quality of care: patient-provider relationships, therapeutic and caring interactions, individual needs, interpersonal processes or skills, and responsiveness. However, this conceptual ambiguity hinders the quality of care field. It encourages redundancy and limits the ability to share and learn about the same construct across different health care settings. Relatedly, no clear factor structure emerged for interpersonal quality of care, with the number of dimensions extracted ranging from one to eight. These issues restricted the

comparison of instruments. There is a need for a unifying definition and taxonomy of interpersonal quality of care to guide measurement and instrument development.

I compared instruments measured in similar country or health care settings, expecting that context would influence the variability in construct definitions and factor structure. Surprisingly, there were no specific patterns in constructs or dimensions when comparing instruments validated within high-income countries, middle- to low-income countries, or similar health care settings. This lack of cohesion may be a product of the construct ambiguity discussed above. One exception was the five instruments measuring interpersonal quality of care in labor and delivery. While there was no trend in the number of factors, three of the five instruments included separate factors that represented positive care and negative care.[39,40,44] Despite 44% of instruments having at least one negative item included in the measure, this separation of factors based on positive or negative interactions was only present in two other instruments, the Interpersonal Processes of Care – Revised and the Interpersonal Processes of Care – Short Form, which were created for a primary care setting.[41] While experiencing both positive and negative aspects of interpersonal quality of care is not specific to maternity care, the intense exposure to and interaction with health care providers and the vulnerability of the patients during labor and delivery may thus promote the interest in measuring both facets of the care experience. Subsequent measures of interpersonal quality of care developed for maternal health use should consider the inclusion of both types of items.

The instruments reviewed varied in both instrument quality and methodological quality. The majority of the instruments received high ratings for content validity and internal consistency,

which is paramount to provide confidence in the utility of the instrument. However, about a third of the tools either had questionable content validity or did not meet the threshold value of Cronbach's alpha for adequate reliability. This may indicate the early stages of development and that the instruments require further refinement. One-fifth of the instruments evaluated test-retest reliability, and only two had a correlation coefficient value that met the 0.70 cut off to qualify as reliable.[29,35] Overall, the methodological quality of this domain was low, with no study scoring an excellent rating.

More striking, however, were the deficits in the methodological quality for the structural validity domain, compromising the evaluation of the results and utility of the instruments. For example, while 48% of all instruments had adequate factor loadings, none of the studies that performed an exploratory factor analysis provided fit statistics, and 62% of the studies that performed an exploratory method either provided inappropriate or no justification for their method of choice. Across all studies that performed any structural validation, none used a power calculation. Few used rule of thumb methods to determine sample size, an inferior method that nevertheless illustrates consideration of sample size in the study.

Taken together, eight instruments generally had good reliability and validity results. However, the confidence in these findings is affected by shortfalls in the quality of the studies.[26–28,30,42–44,46] For example, five instruments, the Caring Assessment Tool (CAT),[30] the Individualized Care Scale (ICS),[42] the Individualized Care Scale – Revised (ICS-R),[43] the Scale of Supportive Care Given During Labor,[44] and the Patient-Doctor Relationship Questionnaire (PDRQ-9)[46] did not have the target population involved in item selection, hindering content validity. Four

instruments, the Questionnaire on the Quality of Physician-Patient Interaction (QQPPI),[27] the Health Care Relationship Trust Scale,[28] the Patient-Doctor Relationship Questionnaire (PDRQ-9),[46] and the Individualized Care Scale (ICS)[42] had poor test-retest reliability, likely due to the differences in the test conditions between the two measurements. Two studies omitted key information necessary to evaluate the instruments' structural validity models: the Questionnaire on the Quality of Physician-Patient Interaction (QQPPI)[27] lacked fit statistics, and the Individualized Care Scale – Revised (ICS-R)[43] had good exploratory results, but did not provide the results for its confirmatory factor analysis. Lastly, the studies validating the Health Care Relationship Trust Scale[28] and the Patient-Provider Relationship Scale (PPRS)[26] provided an incorrect justification for the choice of structural validity method.

Despite these methodological limitations, the Quality of Physician-Patient Interaction (QQPPI)[27] instrument shows the most promise. This tool had good internal consistency, factor loadings, and content validity, but could benefit from additional validations. Specifically, validations should employ more rigorous methods and explicitly discuss the rationale for choice of methods to ultimately determine if the instruments are useful measures of interpersonal quality of care.

This review had several limitations. First, in Ovid MEDLINE, the search was limited to publications categorized as validation studies. It is thus possible that the search missed eligible articles for inclusion articles that were not appropriately tagged. However, review of the reference lists of included articles provided an additional source of articles. Second, I excluded articles that were validations of existing instruments in other languages, as the goal of this review was to understand the quality of reliability and structural validity, rather than cross-cultural

validity. A next step could be to determine if cross-cultural validity was assessed for any of the instruments that received high quality grades in this review. Third, while psychometric standards and published checklists guided the assessment of methodological quality, I adapted the COSMIN checklist and created my own criteria for structural validity. This may have influenced the quality ratings. For example, for the COSMIN checklist, I eliminated the questions about missing data in the internal consistency and test-retest reliability sections. Inclusion of this criterion would have downgraded the quality ratings for studies that did not include information on missing data.

Conclusion

This review found that the construct of interpersonal quality of care suffers from conceptual ambiguity. Overall, the reliability and validity of instruments measuring interpersonal quality of care were hindered by poor quality of the studies, which impairs confidence in the utility of the measures. One instrument showed promise and should be further validated using rigorous methods in a variety of settings to confirm its results. Future research is also encouraged to create a unified definition of interpersonal quality of care, which can guide the creation and validation of quality measures.

3.6. Acknowledgements

I acknowledge Suzue Saito (SS) who assisted with the search as my second reviewer.

3.7. References

1. Tunçalp Ö, Were W, MacLennan C, et al. Quality of care for pregnant women and newborns-the WHO vision. *BJOG.* 2015.

2. Koblinsky M, Moyer CA, Calvert C, et al. Quality maternity care for every woman, everywhere: a call to action. *Lancet.* 2016;388:2307-2320.

3. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: National Academy Press; 2001.

4. World Health Organization. *Quality of Care: A Process for Making Strategic Choices in Health Systems.* Geneva2006. 92 4 156324 9

5. Donabedian A. The quality of care. How can it be assessed? *JAMA.* 1988;260:1743-1748.

6. Stewart AL, Napoles-Springer A, Perez-Stable EJ. Interpersonal processes of care in diverse populations. *Milbank Q.* 1999;77:305-339.

7. Cleary PD, McNeil BJ. Patient satisfaction as an indicator of quality care. *Inquiry.* 1988;25(1):25-36.

8. Mead N, Bower P. Patient-centredness: a conceptual framework and review of the empirical literature. *Soc Sci Med.* 2000;51:1087-1110.

9. Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. *Med Care Res Rev.* 2013;70:351-379.

10. Radwin LE, Alster K. Individualized nursing care: An empirically generated definition. *Int Nurs Rev.* 2002;49:54-63.

11. Zill JM, Christalle E, Müller E, Härter M, Dirmaier J, Scholl I. Measurement of physician-patient communication—A systematic review. *PLoS One.* 2014;9:e112637.

12. Hudon C, Fortin M, Haggerty J, Lambert M, Poitras M. Measuring patients' perceptions of patient-centered care: a systematic review of tools for family medicine. *Ann Fam Med.* 2011;9:155-164.

13. Köberich S, Farin E. A systematic review of instruments measuring patients′ perceptions of patient-centred nursing care. *Nurs Inq.* 2015;22:106-120.

14. Edvardsson D, Innes A. Measuring person-centered care: a critical comparative review of published tools. *Gerontologist.* 2010;50:834-846.

15. Terwee C, Bot S, de Boer M, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007.

16. de Vet HCW, Adèr HJ, Terwee CB, Pouwer F, Fisher WP. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. *Qual Life Res.* 2005.

17. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol Assess.* 1995;7:286-299.

18. Raykov T, Marcoulides GA. *Introduction to Psychometric Theory.* New York: Routledge; 2011.

19. Brown TA. *Confirmatory Factor Analysis for Applied Research.* 2nd ed. New York: The Guilford Press; 2015.

20. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling.* 1999;6:1-55.

21. Kline RB. *Principles and Practice of Structural Equation Modeling.* 3rd ed. New York: The Guilford Press; 2011.

22. The World Bank. World Bank Country and Lending Groups. 2018; https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

23. Consensus based standards for the selection of health measurement instruments (COSMIN). *COSMIN Checklist with 4-Point Scale.* 2011.http://www.cosmin.nl/the_cosmin_checklist.html.

24. Byrne BM. Factor analytic models: viewing the structure of an assessment instrument from three perspectives. *J Pers Assess.* 2005;85(1):17-32.

25. Afulani PA, Diamond-Smith N, Golub G, Sudhinaraset M. Development of a tool to measure person-centered maternity care in developing settings: validation in a rural and urban Kenyan population. *Reprod Health.* 2017;14:118.

26. Barry OM, Bergh A-M, Makin JD, Etsane E, Kershaw TS, Forsyth BWC. Development of a measure of the patient–provider relationship in antenatal care and its importance in PMTCT. *AIDS Care.* 2012;24:680-686.

27. Bieber C, Muller K, Nicolai J, Hartmann M, Eich W. How does your doctor talk with you? Preliminary validation of a brief patient self-report questionnaire on the quality of physician – patient interaction. *J Clin Psychol Med Settings.* 2010:125-136.

28. Bova C, Fennie KP, Watrous E, Dieckhaus K, Williams AB. The health care relationship (HCR) trust scale: development and psychometric evaluation. *Res Nurs Health.* 2006;29(5):477-488.

29. Cott C, Teare G, McGilton K, Lineker S. Reliability and construct validity of the client-centred rehabilitation questionnaire. *Disabil Rehabil.* 2006;28:1387-1397.

30. Duffy JR, Hoskins L, Seifert RF. Dimensions of caring: psychometric Evaluation of the caring assessment tool. *ANS Adv Nurs Sci.* 2007;30:235-245.

31. Hariharan M, Chivukula U, Rana S. The intensive care unit psychosocial care scale: Development and initial validation. *Intensive Crit Care Nurs.* 2015;31:343-351.

32. Hausman AV, Mader D. Measuring social aspects in the physician/patient relationship. *Health Mark Q.* 2004;21:3-26.

33. Hurst YK, Prescott-Clements LE, Rennie JS. The patient assessment questionnaire: a new instrument for evaluating the interpersonal skills of vocational dental practitioners. *Br Dent J.* 2004;197:497-500.

34. Loblaw BDA, Bezjak A, Bunston T. Development and testing of a visit-specific patient satisfaction questionnaire: The Princess Margaret Hospital Satisfaction with Doctor Questionnaire. *J Clin Oncol.* 1999;17:1931-1938.

35. Loblaw DA, Bezjak A, Singh PM, et al. Psychometric refinement of an outpatient, visit-specific satisfaction with doctor questionnaire. *Psychooncology.* 2004;13:223-234.

36. Makoul G, Krupat E, Chang C-H. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. *Patient Educ Couns.* 2007;67:333-342.

37. Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Fam Pract.* 2004;21:699-705.

38. Ostacoli L, Costantini M, Beccaro M, et al. Development and validation of a patient-physician relationship index in the advanced cancer setting. *Tumori.* 2007;93:485-490.

39. Redshaw M, Martin CR. Validation of a perceptions of care adjective checklist. *J Eval Clin Pract.* 2009;15:281-288.

40. Sheferaw ED, Mengesha TZ, Wase SB. Development of a tool to measure women's perception of respectful maternity care in public health facilities. *BMC Pregnancy Childbirth.* 2016;16:67.

41. Stewart AL, Nápoles-Springer AM, Gregorich SE, Santoyo-Olsson J. Interpersonal Processes of Care Survey: Patient-Reported Measures for Diverse Groups. *Health Research and Educational Trust.* 2007.

42. Suhonen R, Valimaki M, Katajisto J. Developing and testing an instrument for the measurement of individual care. *J Adv Nurs.* 2000;32:1253-1263.

43. Suhonen R, Leino-Kilpi H, Välimäki M. Development and psychometric properties of the Individualized Care Scale. *J Eval Clin Pract.* 2005;11:7-20.

44. Uludağ E, Mete S. Development and testing of women's perception for the scale of supportive care given during labor. *Pain Manag Nurs.* 2015;16(5):751-758.

45. Valentine NB, Bonsel GJ, Murray CJL. Measuring quality of health care from the user's perspective in 41 countries: psychometric properties of WHO's questions on health systems responsiveness. *Qual Life Res.* 2007;16:1107-1125.

46. van der Feltz-Cornelis CM, van Oppen P, van Marwijk HWJ, de Beurs E, van Dyck R. A patient-doctor relationship questionnaire (PDRQ-9) in primary care: development and psychometric evaluation. *Gen Hosp Psychiatry.* 2004;26:115-120.

47. van der Kooy J, Valentine NB, Birnie E, et al. Validity of a questionnaire measuring the world health organization concept of health system responsiveness with respect to perinatal services in the Dutch obstetric care system. *BMC Health Serv Res.* 2014;14.

48. Williams AM, Kristjanson LJ. Emotional care experienced by hospitalised patients: Development and testing of a measurement instrument. *J Clin Nurs.* 2009;18:1069-1077.

49. Williams AM, Pienaar C, Toye C, Joske D, Lester L. Further psychometric testing of an instrument to measure emotional care in hospital. *J Clin Nurs.* 2011;20:3472-3482.

3.8. Tables and Figures

Table 3.1. Search strategy for Ovid MEDLINE, Health and Psychosocial Instruments, and CINAHL databases

Ovid MEDLINE search strategy:
(exp Patient-Centered Care/ OR person-centered care.mp OR exp Professional-Patient Relations/ OR "patient-provider relationship".mp OR respectful care.mp OR "Quality of Health Care"/ OR disrespect.mp OR abuse.mp OR digni*.mp OR patient experience.mp OR interpersonal quality.mp OR quality of care.mp) AND (measurement.mp OR measure.mp OR "Weights and Measures"/ OR instrument.mp OR exp "Surveys and Questionnaires"/ OR Validation Studies/ OR index.mp OR scale.mp OR exp Psychometrics/) and limit to (english language and humans and yr="1988-2017" and validation studies)

Health and Psychosocial Instruments search strategy:
[(Patient-centered care.mp OR patient-provider.mp OR professional patient.mp OR patient-provider relationship.mp OR respectful care.mp OR respect.mp OR disrespect.mp OR digni*.mp OR quality of care.mp OR interpersonal quality.mp OR patient experience.mp) OR (interpersonal.mp AND health care.mp) OR (abuse.mp AND health care.mp)] and limit to (english language and yr="1988-2017")

CINAHL search strategy:
(MH "Patient Centered Care" OR "patient centered care" OR MH "Physician-Patient Relations" OR MH "Professional-Patient Relations+" OR MH "Nurse-Patient Relations" OR "patient provider relationship" OR "respectful care" OR MH "Quality of Health Care+" OR "disrespect" OR "digni*" OR "patient experience" OR MH "Interpersonal Relations+" OR "interpersonal quality" OR MM "Quality of Health Care") AND (MH "Research Measurement+" OR MH "Instrument Validation" OR MH "Validation Studies" OR "index" OR MH "Psychometrics" OR MH "Reliability and Validity" OR MH "Psychometrics" OR MH "Research Instruments" OR MH "Instrument Construction" OR MH "Questionnaires+" OR MH "Scales" OR MH "Structured Questionnaires") AND limiters: Published Date: 19880101-20171231; English Language; Peer Reviewed; Human; Publication Type: Questionnaire/Scale, Research, Research Instrument

Figure 3.1. Flow diagram of article selection procedure

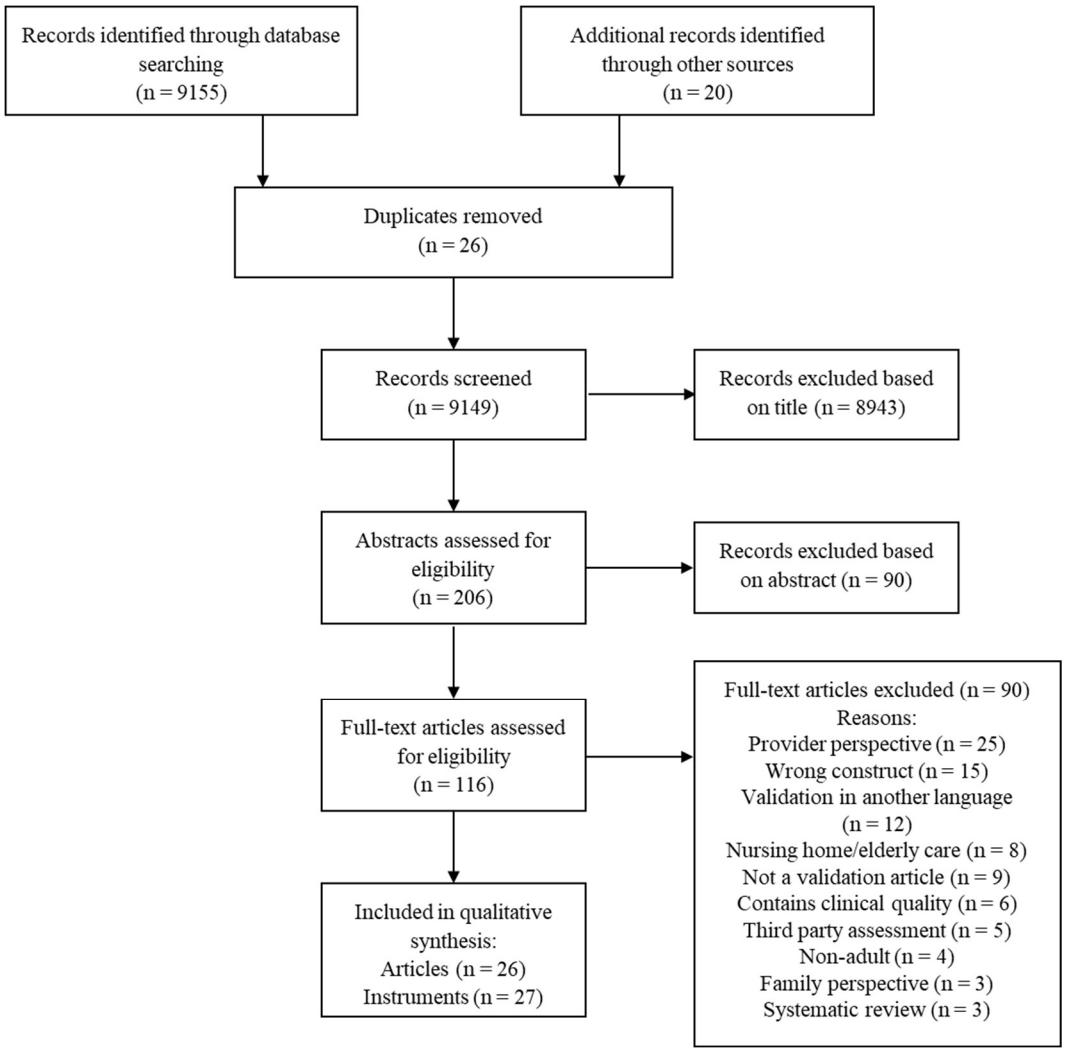


Table 3.2. Criteria for the quality of the instruments

Psychometric property	Rating system	Source
Content validity		
Clear description of the measurement aim, target population, concepts that are being measured, and the item selection, and target population and investigators/experts were involved in item selection	+ = positive rating	Terwee et al.[15]
A clear description of the measurement aim, target population, concepts that are being measured, and the item selection is lacking or only target population is involved in item selection	? = indeterminate	
No target population involvement	- = negative rating	
No information found on target population involvement	0 = no information available	
Internal consistency		
Cronbach's alpha or KR-20 ≥ 0.70	+ = positive rating, - = negative rating, 0 = no information available	Terwee et al.[15]
Test-retest reliability		
Reliability coefficient ≥ 0.70	+ = positive rating, - = negative rating, 0 = no information available	Terwee et al.[15]
Structural validity (assessed separately for EFA, PCA, CFA)		
Factor loadings ≥ 0.30 or statistically significant (if available) or discussion of retaining low factor loadings for clinical significance	+ = positive rating, - = negative rating, 0 = no information available	Brown[19]
No Heywood cases	+ = positive rating, - = negative rating, 0 = no information available	Brown[19]
Goodness of fit: CFI ≥ 0.90, TLI > 0.90, and RMSEA ≤ 0.08	+ = positive rating, - = negative rating, 0 = no information available	Hu & Bentler[20]

EFA = exploratory factor analysis, PCA = principal component analysis, CFA = confirmatory factor analysis

Table 3.3. Criteria for methodological quality of the studies

Psychometric property	Rating system	Source
Internal consistency		
Was the sample size included in the internal consistency analysis adequate?	Per item, 4-point scale: excellent, very good, fair, and poor. Overall score is determined by the lowest-rated criterion	COSMIN Checklist[23]
Was the unidimensionality of the scale checked? i.e. was factor analysis or item response theory model applied?		
Was an internal consistency statistic calculated for each scale separately?		
Was Cronbach's alpha or KR-20 calculated?		
Test-retest reliability		
Was the sample size included in the analysis adequate?	Per item, 4-point scale: excellent, very good, fair, and poor. Overall score is determined by the lowest-rated criterion	COSMIN Checklist[23]
Were at least two measurements available?		
Were the administrations independent?		
Was the time interval stated?		
Were patients stable in the interim period on the construct to be measured?		
Were the test conditions similar for both instruments?		
For continuous scores: Was an intraclass correlation coefficient calculated?		
For dichotomous/nominal/ordinal scores: Was kappa calculated?		
For ordinal scores: Was a weighted kappa calculated?		
For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic		

Psychometric property	Rating system	Source
Structural validity		
Choice of EFA vs. CFA is justified. If the factorial model had already been published, EFA is not appropriate	Per item, + = positive rating, - = negative rating, 0 = no information available, N/A = not applicable for the method	Byrne[24]
For EFA/PCA:		
Authors correctly justified their approach in using EFA vs. PCA. If the aim is item reduction, PCA is appropriate. If the aim is to explain correlations between indicators or identify latent factors, EFA is appropriate		Raykov & Marcoulides[18]
Power calculation using Monte Carlo simulation to determine sample size		Brown[19]
Rule of thumb method used to determine sample size (inferior method)		Brown[19]
Discussion of how missing data were handled		COSMIN Checklist[23]
Extraction method stated		Floyd & Widaman[17]
Extraction method used appropriate for the data (maximum likelihood for continuous variables, WLSMV for categorical variables, principal axis factoring, etc.)		Floyd & Widaman[17]
Rotation method stated (orthogonal or oblique)		Floyd & Widaman[17]
Rationale for rotation method provided		Floyd & Widaman[17]
Eigenvalues provided / # of factors extracted guided by eigenvalues		Raykov & Marcoulides[18]
Factor loadings provided		Floyd & Widaman[17]
For EFA with ML or WLSMV, goodness of model fit assessed (Preferably CFI/TLI/RMSEA that are not sensitive to sample size)		Hu & Bentler[20]
For CFA:		
If EFA and CFA performed in the same study, different samples used		Floyd & Widaman[17]
Power calculation using Monte Carlo simulation to determine sample size		Brown[19]
Rule of thumb method used to determine sample size (inferior method)		Brown[19]
Discussion of how missing data were handled		COSMIN Checklist[23]
Type of estimator stated		Brown[19]
Estimator used appropriate for the data (maximum likelihood for continuous variables, WLSMV for categorical variables)		Brown[19]
Factor loadings provided		Floyd & Widaman[17]
Goodness of model fit assessed (Preferably CFI/TLI/RMSEA that are not sensitive to sample size)		Hu & Bentler[20]

EFA = exploratory factor analysis, PCA = principal component analysis, CFA = confirmatory factor analysis, WLSMV = weighted least squares with mean and variance adjustment, ML = maximum likelihood, CFI = Comparative Fit Index, TLI = Tucker Lewis Index, RMSEA = Root Mean Square Error of Approximation

Table 3.4. Construct categories, constructs, and construct definitions of included instruments

Construct categories and constructs	Construct definition	Instruments
1. patient-provider relationship		
Patient-provider relationship	Not defined	Patient-Provider Relationship Scale (PPRS)[26]
Social aspects of the professional service relationships	The personal bond between client and professional[32]	Social Aspects of Professional Service Relationships (SAPSR)[32]
Patient-physician relationship	"…patient's positive bond with the therapist who is perceived as a helpful and supportive person"[38(p486)]	Patient-Physician Relationship Index (PPRI)[38]
Patient-doctor relationship	Not defined	Patient-Doctor Relationship Questionnaire (PDRQ-9)[46]
Patient satisfaction with doctor-patient interaction	Not defined	Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD)[34,35]
Quality of the patient-provider interaction	Designed to measure aspects of a good physician-patient relationship such as information exchange, patient involvement, and shared decision-making[27]	Questionnaire on the Quality of Physician-Patient Interaction (QQPPI)[27]
Trust	Includes the dimensions of competence, compassion, confidentiality, reliability, dependability, open communication, and reciprocity.[28]	Health Care Relationship Trust Scale[28]
2. Therapeutic and caring interactions		
Empathy	Understand the patient's situation and communicate and act to respond to the patient's situation in a therapeutic way.[37]	Consultation and Relational Empathy Measure (CARE)[37]
Emotional care	Therapeutic and interpersonal interactions[48,49]	Patient Evaluation of Emotional Care during Hospitalization (PEECH)[48,49]
Supportive care	Not defined	Scale of Supportive Care Given During Labor[44]
Psychosocial care	"…specific supportive interventions such as providing explanations, reassuring and raising faith and hope, cheering-up, strengthening patients' self-esteem, giving emotional warmth, offering empathetic listening, empathetic touch, and spending extra time with patients."[31(p344)]	Intensive Care Unit Psychosocial Care Scale (ICU-PU Scale)[31]
Caring	"…human interaction, mutuality, appreciating the uniqueness of individuals, and improving the welfare of patients and families"[30(p235)]	Caring Assessment Tool (CAT)[30]

3. Individual needs		
Individualized care	Personalization of care based on patient's feelings and preferences, and involving the patient in decision-making[42,43]	Individualized Care Scale (ICS), Individualized Care Scale (ICS) – Revised[42,43]
Client-centered rehabilitation	Client participation in decision-making and goal-setting, client-centered education, evaluation of outcomes from client's perspective, family involvement, emotional support, co-ordination/continuity, physical comfort[29]	Client-centered Rehabilitation Questionnaire (CCRQ)[29]
Person-centered maternity care	Adapted from Institute of Medicine's definition of person-centered care: "providing maternity care that is respectful and responsive to individual women and their families' preferences, needs, and values, and ensuring that their values guide all clinical decisions." [25(p3)]	Person-Centered Maternity Care Scale[25]
Respectful maternity care	"Respectful maternity care (RMC) encompasses the universal right of every childbearing woman to receive care that includes respect for the woman's autonomy, dignity, feelings, choices, and preferences including the choice of companionship and cultural rituals at birth…"[40(p2)]	Respectful Maternity Care Scale[40]
4. Interpersonal processes or skills		
Interpersonal processes	Three dimensions of interpersonal processes proposed: communication, decision making, and interpersonal style[6,41]	Interpersonal Processes of Care[6], Interpersonal Processes of Care - Revised,[41] Interpersonal Processes of Care - Short Form[41]
Interpersonal skills	Not defined	Patient Assessment Questionnaire (PAQ)[33]
Interpersonal and communication skills	Not defined	Communication Assessment Tool[36]
5. Responsiveness		
Health system responsiveness	"…non-clinical and non-financial dimensions of quality of care that reflected respect for human dignity and interpersonal aspects of the care process"[45(p1108)]	WHO Health System Responsiveness Questionnaire,[45] Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Antenatal care,[47] Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - labor and delivery,[47] Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) – postpartum[47]
6. Other		
Perceptions of maternity care	Not defined	Perceptions of Care Adjective Checklist (PCACL-R)[39]

Table 3.5. Characteristics of the context in which the instruments were measured

Title of instrument	Author	Year	Country	Patient population	Health care setting
Person-Centered Maternity Care Scale[a]	Afulani et al.[25]	2017	Kenya	Women	Labor and delivery at public hospitals and health centers
Patient-Provider Relationship Scale (PPRS)	Barry et al.[26]	2012	South Africa	Women	Antenatal care clinics
Questionnaire on the Quality of Physician-Patient Interaction (QQPPI)	Bieber et al.[27]	2010	Germany	Adults	Outpatient care
Health Care Relationship Trust Scale	Bova et al.[28]	2006	USA	Adults living with HIV	HIV primary care clinics
Client-centered Rehabilitation Questionnaire (CCRQ)	Cott et al.[29]	2006	Canada	Adults in rehabilitation hospitals	Rehabilitation hospitals
Caring Assessment Tool (CAT)	Duffy et al.[30]	2007	USA	Medical-surgical patients who were hospitalized at least 2 days	Hospital inpatient care
Intensive Care Unit Psychosocial Care Scale (ICU-PU Scale)	Hariharan et al.[31]	2015	India	Adults who underwent Coronary artery bypass graft	Intensive care unit in hospitals
Social Aspects of Professional Service Relationships (SAPSR)	Hausman et al.[32]	2004	USA	Adults	Not provided
Patient Assessment Questionnaire (PAQ)	Hurst et al.[33]	2002	Scotland	Adults	Dentist office
Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD)	Loblaw et al.[34]	1999	Canada	Cancer patients	Oncology departments
	Loblaw et al.[35]	2004	Canada	Cancer patients	Oncology departments
Communication Assessment Tool	Makoul et al.[36]	2007	USA	Adults	Outpatient care
Consultation and Relational Empathy Measure (CARE)	Mercer et al.[37]	2004	UK	Adults	Primary care
Patient-Physician Relationship Index (PPRI)	Ostacoli et al.[38]	2007	Italy	Cancer patients	Oncology departments
Perceptions of Care Adjective Checklist (PCACL-R)	Redshaw et al.[39]	2009	UK	Women	Labor and delivery at health facilities
Respectful Maternity Care Scale	Sheferaw et al.[40]	2016	Ethiopia	Women	Labor and delivery at public hospitals and health centers
Interpersonal Processes of Care	Stewart et al.[6]	1999	USA	Adults	Primary care
Interpersonal Processes of Care - Revised[a]	Stewart et al.[41]	2007	USA	Adults	Primary care
Interpersonal Processes of Care - Short Form	Stewart et al.[41]	2007	USA	Adults	Primary care
Individualized Care Scale (ICS)	Suhonen et al.[42]	2000	Finland	Surgical patients	Hospital inpatient care - surgery

Title of instrument	Author	Year	Country	Patient population	Health care setting
Individualized Care Scale – Revised (ICS-R)[a]	Suhonen et al.[43]	2005	Finland	Surgical patients	Hospital inpatient care - surgery
Scale of Supportive Care Given During Labor	Uludağ et al.[44]	2015	Turkey	Women	Labor and delivery at a hospital
WHO Health System Responsiveness Questionnaire[a]	Valentine et al.[45]	2007	41 countries	Adults	Outpatient care
Patient-Doctor Relationship Questionnaire (PDRQ-9)	van der Feltz-Cornelis et al.[46]	2004	Netherlands	Adults	Primary care and epilepsy clinic
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Antenatal care	van der Kooy et al.[47]	2014	Netherlands	Women	Antenatal care
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Labor and delivery	van der Kooy et al.[47]	2014	Netherlands	Women	Labor and delivery
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Postpartum	van der Kooy et al.[47]	2014	Netherlands	Women	Postpartum
Patient Evaluation of Emotional Care during Hospitalization (PEECH)	Williams et al.[48]	2009	Australia	Adults	Hospital inpatient care
	Williams et al.[49]	2011	Australia	Adults	Hospital inpatient care

[a] Instrument named for purposes of the systematic review

Table 3.6. Reliability and validity methods of the included instruments

Title of instrument	Author	Data collection	Sample size	Response scale	Content validity	Internal consistency	Test-retest reliability	Structural validity
Person-centered maternity care scale	Afulani et al.[25]	Interview-administered questionnaire	1387	4-point Likert scale: no, never; yes, a few times; yes, most of the time; yes, all the time	literature review, expert review, cognitive interviews	Cronbach's alpha	Not tested	EFA
						Not tested	Not tested	CFA
Patient-Provider Relationship Scale (PPRS)	Barry et al.[26]	Interview-administered questionnaire	192	4-point Likert scale: always, never	literature review, expert review, focus group discussion	Cronbach's alpha	Not tested	PCA
Questionnaire on the Quality of Physician-Patient Interaction (QQPPI)	Bieber et al.[27]	Self-completed questionnaire	147	5-point Likert scale: I do not agree, I partly agree, I agree, I strongly agree, I fully agree	literature review, in-depth interviews, expert review	Cronbach's alpha	- Retest period: 3 weeks - Pearson correlation coefficient	EFA
Health Care Relationship Trust Scale	Bova et al.[28]	Interview-administered questionnaire	99	5-point Likert scale: none of the time, some or a little of the time, occasionally or a moderate amount of the time, most of the time, all of the time	literature review, focus group discussions, expert review	Cronbach's alpha	- Retest period: 2-4 weeks - Pearson correlation coefficient	PCA
Client-centered Rehabilitation Questionnaire (CCRQ)	Cott et al.[29]	Self-completed questionnaire	1002	5-point Likert scale: strongly disagree, disagree, neither agree nor disagree, agree, strongly agree	literature review, expert review, focus group discussions, cognitive interviews	Cronbach's alpha	- Retest period: 2 weeks - ICC	N/A
Caring Assessment Tool (CAT)	Duffy et al.[30]	Self-completed questionnaire	557	5-point Likert scale: never to always	expert review	Cronbach's alpha	N/A	PCA
Intensive Care Unit Psychosocial Care Scale (ICU-PU Scale)	Hariharan et al.[31]	Details not provided	250	5-point Likert scale: never to always	literature review, in-depth interviews, expert review	Cronbach's alpha	Not tested	EFA

Title of instrument	Author	Data collection	Sample size	Response scale	Content validity	Internal consistency	Test-retest reliability	Structural validity
Social Aspects of Professional Service Relationships (SAPSR)	Hausman et al.[32]	Self-completed questionnaire	181	5-point Likert scale: strongly agree to strongly disagree	in-depth interviews, expert review	Cronbach's alpha	Not tested	EFA
			109			N/A	Not tested	EFA
			109			Cronbach's alpha	Not tested	CFA
			239			Cronbach's alpha	Not tested	CFA
Patient Assessment Questionnaire (PAQ)	Hurst et al.[33]	Self-completed questionnaire	5767	5-point Likert scale: poor, fair, good, very good, excellent	Not tested	Cronbach's alpha	Not tested	N/A
Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD)	Loblaw et al. (1999)[34]	Details not provided	174	4-point Likert scale: strongly agree, agree, disagree, strongly disagree	literature review, expert review including patients, pilot testing	Cronbach's alpha	Not tested	PCA
	Loblaw et al. (2004)[35]	Details not provided	80	4-point Likert scale: strongly agree, agree, disagree, strongly disagree	N/A	Cronbach's alpha	- Retest period: 3-5 days - Pearson correlation coefficient	EFA
			174			Not tested	Not tested	CFA
Communication Assessment Tool	Makoul et al.[36]	Details not provided	600	5-point Likert scale: poor, fair, good, very good, excellent	literature review, expert review, focus group discussions, pilot testing	Cronbach's alpha	Not tested	PCA
		Telephone or internet questionnaire	950	5-point Likert scale: poor, fair, good, very good, excellent		Cronbach's alpha	Not tested	CFA
Consultation and Relational Empathy Measure (CARE)	Mercer et al.[37]	Details not provided	10	5-point Likert scale: poor, fair, good, very good, excellent	literature review, pilot testing using semi-structured interviews, expert review	Cronbach's alpha	Not tested	N/A
Patient-Physician Relationship Index (PPRI)	Ostacoli et al.[38]	Details not provided	109	5-point Likert scale: not at all, a little, quite a bit, much, very much	qualitative interviews, expert review	Cronbach's alpha	Not tested	PCA
Perceptions of Care Adjective Checklist (PCACL-R)	Redshaw et al.[39]	Self-completed questionnaire	2960	Circling the adjectives if they apply	Not tested - face validity explored in previous studies	Cronbach's alpha	Not tested	CFA

Title of instrument	Author	Data collection	Sample size	Response scale	Content validity	Internal consistency	Test-retest reliability	Structural validity
Respectful Maternity Care Scale	Sheferaw et al.[40]	Interview-administered questionnaire	515	5-point Likert scale: strongly agree, agree, don't know, do not agree, strongly do not agree	literature review, in-depth interviews, expert review	Cronbach's alpha	Not tested	PCA
Interpersonal Processes of Care	Stewart et al.[6]	Interview-administered questionnaire	603	Not provided	literature review	Cronbach's alpha	Not tested	N/A
Interpersonal Processes of Care - Revised	Stewart et al.[41]	Self-completed questionnaire	1664	5-point Likert scale: never, rarely, sometimes, usually, always	Focus group discussions, literature review, cognitive interviews	Not tested	Not tested	CFA
Interpersonal Processes of Care - Short Form	Stewart et al.[41]	Self-completed questionnaire	1664	5-point Likert scale: never, rarely, sometimes, usually, always	Focus group discussions, literature review, cognitive interviews	Cronbach's alpha	Not tested	CFA
Individualized Care Scale (ICS)	Suhonen et al.[42]	Self-completed questionnaire	203	Not provided	literature review, expert review, pilot testing	Cronbach's alpha	- Retest period: 2 weeks - Pearson correlation coefficient	PCA
Individualized Care Scale – Revised (ICS-R)	Suhonen et al.[43]	Self-completed questionnaire	454	5-point Likert scale: fully disagree, disagree, neither disagree nor agree, agree, fully agree	N/A - content validity explored in previous studies	Cronbach's alpha	Not tested	PCA
						Not tested	Not tested	CFA
Scale of Supportive Care Given During Labor	Uludağ et al.[44]	Details not provided	360	4-point Likert scale: never to always	expert review	N/A	Not tested	EFA
						Cronbach's alpha	Not tested	CFA

Title of instrument	Author	Data collection	Sample size	Response scale	Content validity	Internal consistency	Test-retest reliability	Structural validity
WHO Health System Responsiveness Questionnaire	Valentine et al.[45]	Interview-administered questionnaire	50,876	4-point Likert scale: always, usually, sometimes, never OR 5-point Likert scale: very good, good, moderate, bad, very bad depending on the question	literature review, expert review, cognitive interviews, pilot testing	Cronbach's alpha	- Retest period: 8-30 days - Kappa	N/A
						Not tested	Not tested	EFA
Patient-Doctor Relationship Questionnaire (PDRQ-9)	van der Feltz-Cornelis et al.[46]	Self-completed questionnaire	255	Not provided	Pilot testing	Cronbach's alpha	- Retest period: mean = 61 days - Pearson correlation coefficient	PCA
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Antenatal care	van der Kooy et al.[47]	Interview-administered questionnaire	171	5-point Likert scale: very good, good, moderate, bad, very bad	expert review, qualitative interviews	Cronbach's alpha	Not tested	EFA
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Labor and delivery	van der Kooy et al.[47]	Interview-administered questionnaire	171	5-point Likert scale: very good, good, moderate, bad, very bad	expert review, qualitative interviews	Cronbach's alpha	Not tested	EFA
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Postpartum	van der Kooy et al.[47]	Interview-administered questionnaire	171	5-point Likert scale: very good, good, moderate, bad, very bad	expert review, qualitative interviews	Cronbach's alpha	Not tested	EFA
Patient Evaluation of Emotional Care during Hospitalization (PEECH)	Williams et al. (2009)[48]	Self-completed questionnaire	132	4-point Likert scale: all, most, some, none	literature review, expert review, pilot testing	Cronbach's alpha	Not tested	EFA
	Williams et al. (2011)[49]	Self-completed questionnaire	251	4-point Likert scale: all, most, some, none	Performed in previous study (Williams et al 2009)	Cronbach's alpha	Not tested	CFA

EFA = exploratory factor analysis, PCA = principal component analysis, CFA = confirmatory factor analysis

Table 3.7. Reliability and validity results of the included instruments

Title of instrument	Author	Internal consistency	Test-retest reliability	# of items included - final scale	Structural validity	Dimensions (# of items per dimension)	Fit statistics
Person-Centered Maternity Care Scale	Afulani et al.[25]	Dignified and respectful care: 0.63; communication and autonomy: 0.73; supportive care: 0.72; overall: 0.86		30	EFA	3 dimensions: dignified and respectful care (6), communication and autonomy (9), supportive care (15)	
				30	CFA	3 dimensions: dignified and respectful care (6), communication and autonomy (9), supportive care (15)	
Patient-Provider Relationship Scale (PPRS)	Barry et al.[26]	0.91		14	PCA	1 dimension	N/A
Questionnaire on the Quality of Physician-Patient Interaction (QQPPI)	Bieber et al.[27]	0.95	0.59	14	EFA	1 dimension	
Health Care Relationship Trust Scale	Bova et al.[28]	interpersonal connection: 0.85, respectful communication: 0.81, professional partnering: 0.89, overall: 0.92	0.59	15	PCA	3 dimensions: interpersonal connection (5); respectful communication (4); professional partnering skills (6)	N/A
Client-centered Rehabilitation Questionnaire (CCRQ)	Cott et al.[29]	decision-making: 0.87; education: 0.72; outcome evaluation: 0.82; family involvement: 0.88; emotional support: 0.85; co-ordination/continuity: 81; physical comfort: 0.78; overall: 0.97	decision-making: 0.78; education: 0.74; outcome evaluation: 0.85; family involvement: 0.83; emotional support: 0.77; co-ordination/continuity: 85; physical comfort: 0.84; overall: 0.85	30	N/A	7 dimensions: decision-making (5), education (4), outcome evaluation (4), family involvement (5), emotional support (4), co-ordination/continuity (4), physical comfort (4)	N/A

Title of instrument	Author	Internal consistency	Test-retest reliability	# of items included - final scale	Structural validity	Dimensions (# of items per dimension)	Fit statistics
Caring Assessment Tool (CAT)	Duffy et al.[30]	mutual problem solving: 0.89; attentive reassurance: 0.92; human respect: 0.90; encouraging manner: 0.92; appreciation of unique meanings: 0.90; healing environment: 0.86; affiliation needs: 0.82; basic human needs: 0.76		36	PCA	8 dimensions: mutual problem solving (5); attentive reassurance (5); human respect (5); encouraging manner (6); appreciation of unique meanings (4); healing environment (5); affiliation needs (3); basic human needs (3)	N/A
Intensive Care Unit Psychosocial Care Scale (ICU-PU Scale)	Hariharan et al.[31]	human dignity and rights: 0.82; transparency for decision making and care continuity: 0.78; sustained patient, family orientation: 0.47; overall: 0.86		14	EFA	3 dimensions: protection of human dignity and rights (6); transparency for decision making and care continuity (6); sustained patient, family orientation (2)	
Social Aspects of Professional Service Relationships (SAPSR)	Hausman et al.[32]	0.95		16	EFA	1 dimension	
		N/A		15	EFA	1 dimension	
		0.90		6	CFA	1 dimension	CFI: 0.96; RMSEA: 0.10
		0.92		6	CFA	1 dimension	CFI: 0.99; RMSEA: 0.09
Patient Assessment Questionnaire (PAQ)	Hurst et al.[33]	0.95		13	N/A	N/A	N/A
Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD)	Loblaw et al. (1999)[34]	information exchange: 0.92; interpersonal skills: 0.90; empathy: 0.88; quality of time: 0.88; overall: 0.97		29	PCA	4 dimensions: information exchange (10); interpersonal skills (8); empathy (6); quality of time (5)	N/A
	Loblaw et al. (2004)[35]	Time 1: physician disengagement: 0.92; perceived support: 0.85 Time 2: physician disengagement: 0.93; perceived support: 0.95	physician disengagement: 0.79; perceived support: 0.76; overall: 0.60	24	EFA	2 dimensions: physician disengagement (13); perceived support (11)	
				24	CFA	2 dimensions: physician disengagement (13); perceived support (11)	CFI: 0.88; RMSEA: 0.08

Title of instrument	Author	Internal consistency	Test-retest reliability	# of items included - final scale	Structural validity	Dimensions (# of items per dimension)	Fit statistics
Communication Assessment Tool	Makoul et al.[36]	0.98		15	PCA	1 dimension	N/A
		0.96		14	CFA	1 dimension	
Consultation and Relational Empathy Measure (CARE)	Mercer et al.[37]	0.93		10	N/A	N/A	N/A
Patient-Physician Relationship Index (PPRI)	Ostacoli et al.[38]	Time 1: 0.81 Time 2: 0.86		8	PCA	1 dimension	N/A
Perceptions of Care Adjective Checklist (PCACL-R)	Redshaw et al.[39]	positive: 0.78; negative: 0.73; overall: 0.81		16	CFA	2 correlated dimensions: positive (8); negative (8)	CFI: 0.97, TLI: 0.99, RMSEA: 0.03
Respectful Maternity Care Scale	Sheferaw et al.[40]	Friendly care: 0.89; abuse-free care: 0.75; timely care: 0.71; discrimination-free care: 0.67		15	PCA	4 dimensions: friendly care (7); abuse-free care (3); timely care (3); discrimination-free care (2)	N/A
Interpersonal Processes of Care	Stewart et al. (1999)[6]	general clarity: 0.70; elicitation and responsiveness of patients' concerns and expectations: 0.86; explanations of condition: 0.93; explanations of processes: 0.78; explanations of self-care: 0.83; explanations of medications: 0.74; empowerment: 0.84; responsiveness to patient preferences: 0.64; consideration of patients' ability to comply: 0.85; friendliness and courteousness: 0.76; respectfulness: 0.76; discrimination: 0.87; emotional support/reassurance: 0.75		41	N/A	13 dimensions: general clarity (2); elicitation and responsiveness of patients' concerns and expectations (4); explanations of condition (2); explanations of processes (4); explanations of self-care (2); explanations of medications (5); empowerment (2); responsiveness to patient preferences (4); consideration of patients' ability to comply (2); friendliness and courteousness (3); respectfulness (4); discrimination (4); emotional support/reassurance (4)	N/A

Title of instrument	Author	Internal consistency	Test-retest reliability	# of items included - final scale	Structural validity	Dimensions (# of items per dimension)	Fit statistics
Interpersonal Processes of Care - Revised	Stewart et al. (2007)[41]			29	CFA	7 second-order dimensions: Hurried communication (5); elicited concerns/responded (3); explained results/medications (4); patient-centered decision making (4); compassionate/respectful (5); discrimination (4); disrespectful office staff (4)	RMSEA: <0.04; CFI: >0.96
Interpersonal Processes of Care - Short Form	Stewart et al. (2007)[41]	Lack of clarity: 0.65; elicited concerns/responded: 0.80; explained results: 0.81; decided together: 0.75; compassionate/respectful: 0.71; discriminated due to race/ethnicity: 0.79; disrespectful office staff: 0.90		18	CFA	7 dimensions: Lack of clarity (2); elicited concerns/responded (3); explained results (2); decided together (2); compassionate/respectful (3); discriminated due to race/ethnicity (2); disrespectful office staff (4)	RMSEA: 0.04; CFI: 0.97
Individualized Care Scale (ICS)	Suhonen et al. (2000)[42]	patient's situation during hospitalization: 0.91; patient's personal life situation: 0.84; facilitating participation in decision making: 0.93	patient's situation during hospitalization: 0.65; patient's personal life situation: 0.79; facilitating participation in decision making: 0.82	37	PCA	3 dimensions: patient's situation during hospitalization (10); patient's personal life situation (10); facilitating participation in decision making (17)	N/A

Title of instrument	Author	Internal consistency	Test-retest reliability	# of items included - final scale	Structural validity	Dimensions (# of items per dimension)	Fit statistics
Individualized Care Scale - Revised (ICS-R)	Suhonen et al. (2005)[43]	Scale A: clinical situation: 0.88; personal life situation: 0.88; decisional control: 0.88; overall: 0.94; Scale B: clinical situation: 0.85; personal life situation: 0.85; decisional control: 0.83; overall: 0.93		38	PCA	2 scales with same 3 dimensions: Scales: ICS-A) patient's views of the support for individuality received from nurses through specific nursing interventions (19); ICS-B) patient's perceptions of individuality in his or her own care (19) Dimensions: clinical situation (7); personal life situation (5); decisional control (7)	N/A
		Not tested		38	CFA	2 scales with same 3 dimensions: Scales: ICS-A) patient's views of the support for individuality received from nurses through specific nursing interventions (19); ICS-B) patient's perceptions of individuality in his or her own care (19) Dimensions: clinical situation (7); personal life situation (5); decisional control (7)	
Scale of Supportive Care Given During Labor	Uludağ et al.[44]			33	EFA	3 dimensions: comfortable behaviors; education; disturbing behaviors	
		comfortable behaviors: 0.92; education: 0.85; disturbing behaviors: 0.97; overall: 0.94		33	CFA	3 dimensions: comfortable behaviors (15); education (8); disturbing behaviors (10)	RMSEA: 0.07, CFI: 0.97, TLI: 0.97, NFI: 0.95

Title of instrument	Author	Internal consistency	Test-retest reliability	# of items included - final scale	Structural validity	Dimensions (# of items per dimension)	Fit statistics
WHO Health System Responsiveness Questionnaire	Valentine et al.[45]	prompt attention: 0.65; dignity: 0.84; communication: 0.88; autonomy: 0.82; confidentiality: 0.83; choice of provider: 0.82; quality of basic amenities: 0.92; overall: 0.93	0.58-0.69	22	N/A	7 domains: prompt attention (2); dignity (4); communication (4); autonomy (3); confidentiality (3); choice of provider (3); quality of basic amenities (3)	N/A
		N/A		22	EFA	Developed countries: 5 domains: general factor, prompt attention-autonomy, basic amenities, communication, confidentiality Less-developed countries: 3 dimensions: general factor, basic amenities, choice Items for each factor not provided	
Patient-Doctor Relationship Questionnaire (PDRQ-9)	van der Feltz-Cornelis et al.[46]	0.94	0.61	9	PCA	1 dimension	N/A
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Antenatal care	van der Kooy et al.[47]	prompt attention: 0.67; dignity: 0.73; communication: 0.80; autonomy: 0.73; confidentiality: 0.82; choice and continuity: 0.77; quality of basic amenities: 0.57; social consideration: 0.76		25	EFA	8 domains: prompt attention (4); dignity (3); communication (5); autonomy (3); confidentiality (3); choice and continuity (3); quality of basic amenities (2); social consideration (2)	
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Labor and delivery	van der Kooy et al.[47]	prompt attention: 0.83; dignity: 0.86; communication: 0.92; autonomy: 0.787; confidentiality: 0.78; choice and continuity: 0.88; quality of basic amenities: 0.66; social consideration: 0.87		40	EFA	8 domains: prompt attention (7); dignity (5); communication (6); autonomy (3); confidentiality (6); choice and continuity (7); quality of basic amenities (3); social consideration (3)	

Title of instrument	Author	Internal consistency	Test-retest reliability	# of items included - final scale	Structural validity	Dimensions (# of items per dimension)	Fit statistics
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Postpartum	van der Kooy et al.[47]	prompt attention: 0.89; dignity: 0.87; communication: 0.95; autonomy: 0.94; confidentiality: 0.94; choice and continuity: 0.89; quality of basic amenities: 0.62; social consideration: 0.84		39	EFA	8 domains: prompt attention (4); dignity (5); communication (6); autonomy (5); confidentiality (6); choice and continuity (5); quality of basic amenities (3); social consideration (5)	
Patient Evaluation of Emotional Care during Hospitalization (PEECH)	Williams et al. (2009)[48]	Level of security: 0.68; level of connection: 0.69; level of knowing: 0.67; level of personal value: 0.87		22	EFA	4 dimensions: level of security (6); level of connection (3); level of knowing (3); level of personal value (10)	
	Williams et al. (2011)[49]	Level of security: 0.73; level of connection: 0.59; level of knowing: 0.73; level of personal value: 0.86		22	CFA	4 dimensions: level of security (6); level of connection (3); level of knowing (3); level of personal value (10)	CFI: 0.96; TLI: 0.95; RMSEA: 0.06

EFA = exploratory factor analysis, PCA = principal component analysis, CFA = confirmatory factor analysis, CFI = Comparative Fit Index, TLI = Tucker Lewis Index, RMSEA = Root Mean Square Error Approximation, NFI = Normed Fit Index, N/A = not applicable

Table 3.8. Instrument quality for content validity, internal consistency, test-retest reliability, and structural quality

Title of instrument	Author	Content validity	Internal consistency	Test-retest reliability	Structural quality					
					Exploratory factor analysis or principal component analysis			Confirmatory factor analysis		
					Factor loadings	No Heywood cases	Goodness of fit	Factor loadings	No Heywood cases	Goodness of fit
Person-Centered Maternity Care Scale	Afulani et al.[25]	+	-		-	+	N/A	0	0	0
Patient-Provider Relationship Scale (PPRS)	Barry et al.[26]	?	+		+	+	N/A			
Questionnaire on the Quality of Physician-Patient Interaction (QQPPI)	Bieber et al.[27]	+	+	-	+	+	0			
Health Care Relationship Trust Scale	Bova et al.[28]	+	+	-	+	+	N/A			
Client-centered Rehabilitation Questionnaire (CCRQ)	Cott et al.[29]	+	+	+						
Caring Assessment Tool (CAT)	Duffy et al.[30]	-	+		+	+	N/A			
Intensive Care Unit Psychosocial Care Scale (ICU-PU Scale)	Hariharan et al.[31]	+	-		+	+	0			
Social Aspects of Professional Service Relationships (SAPSR)	Hausman et al.[32]	+	+		+	+	0	0	0	-
Patient Assessment Questionnaire (PAQ)	Hurst et al.[33]	0	+							
Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD)	Loblaw et al. (1999)[34]	?	+		0	0	N/A			
	Loblaw et al. (2004)[35]	N/A	+	+	-	0	0	0	0	-
Communication Assessment Tool	Makoul et al.[36]	?	+		0	0	N/A	0	0	0
Consultation and Relational Empathy Measure (CARE)	Mercer et al.[37]	+	+							
Patient-Physician Relationship Index (PPRI)	Ostacoli et al.[38]	+	+		0	0	N/A			
Perceptions of Care Adjective Checklist (PCACL-R)	Redshaw et al.[39]	0	+					+	+	+
Respectful Maternity Care Scale	Sheferaw et al.[40]	+	-		+	+	N/A			

Title of instrument	Author	Content validity	Internal consistency	Test-retest reliability	Structural quality					
					Exploratory factor analysis or principal component analysis			Confirmatory factor analysis		
					Factor loadings	No Heywood cases	Goodness of fit	Factor loadings	No Heywood cases	Goodness of fit
Interpersonal Processes of Care	Stewart et al. (1999)[6]	-	-							
Interpersonal Processes of Care - Revised	Stewart et al. (2007)[41]	+						0	0	+
Interpersonal Processes of Care - Short Form	Stewart et al. (2007)[41]	+	-					+	+	+
Individualized Care Scale (ICS)	Suhonen et al. (2000)[42]	-	+	-	+	+	N/A			
Individualized Care Scale - Revised (ICS-R)	Suhonen et al. (2005)[43]	N/A	+		+	+	N/A	0	0	0
Scale of Supportive Care Given During Labor	Uludağ et al.[44]	-	+		+	+	0	+	+	+
WHO Health System Responsiveness Questionnaire	Valentine et al.[45]	+	-	-	-	-	0			
Patient-Doctor Relationship Questionnaire (PDRQ-9)	van der Feltz-Cornelis et al.[46]	-	+	-	+	+	N/A			
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Antenatal care	van der Kooy et al.[47]	+	-		0	0	0			
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Labor and delivery	van der Kooy et al.[47]	+	-		+	-	0			
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Postpartum	van der Kooy et al.[47]	+	-		0	0	0			
Patient Evaluation of Emotional Care during Hospitalization (PEECH)	Williams et al. (2009)[48]	-	-		+	+	0			
	Williams et al. (2011)[49]	N/A	-					+	+	+

Rating system: + = positive, - = negative, ? = indeterminate, 0 = no information available, N/A = not applicable. See Table 3.2 for more information.

Table 3.9. Methodological quality of included studies

A. Internal consistency and test-retest reliability

Title of instrument	Author	Internal consistency	Test-retest reliability
Person-Centered Maternity Care Scale	Afulani et al.[25]	Excellent	
Patient-Provider Relationship Scale (PPRS)	Barry et al.[26]	Excellent	
Questionnaire on the Quality of Physician-Patient Interaction (QQPPI)	Bieber et al.[27]	Excellent	Poor
Health Care Relationship Trust Scale	Bova et al.[28]	Good	Poor
Client-centered Rehabilitation Questionnaire (CCRQ)	Cott et al.[29]	Poor	Good
Caring Assessment Tool (CAT)	Duffy et al.[30]	Excellent	
Intensive Care Unit Psychosocial Care Scale (ICU-PU Scale)	Hariharan et al.[31]	Excellent	
Social Aspects of Professional Service Relationships (SAPSR)	Hausman et al.[32]	Excellent	
Patient Assessment Questionnaire (PAQ)	Hurst et al.[33]	Poor	
Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD)	Loblaw et al. (1999)[34]	Excellent	
	Loblaw et al. (2004)[35]	Excellent	Poor
Communication Assessment Tool	Makoul et al.[36]	Excellent	
Consultation and Relational Empathy Measure (CARE)	Mercer et al.[37]	Poor	
Patient-Physician Relationship Index (PPRI)	Ostacoli et al.[38]	Excellent	
Perceptions of Care Adjective Checklist (PCACL-R)	Redshaw et al.[39]	Excellent	
Respectful Maternity Care Scale	Sheferaw et al.[40]	Excellent	
Interpersonal Processes of Care	Stewart et al. (1999)[6]	Poor	
Interpersonal Processes of Care - Revised	Stewart et al. (2007)[41]	Not assessed	
Interpersonal Processes of Care - Short Form	Stewart et al. (2007)[41]	Excellent	
Individualized Care Scale (ICS)	Suhonen et al. (2000)[42]	Excellent	Poor
Individualized Care Scale - Revised (ICS-R)	Suhonen et al. (2005)[43]	Excellent	
Scale of Supportive Care Given During Labor	Uludağ et al.[44]	Excellent	
WHO Health System Responsiveness Questionnaire	Valentine et al.[45]	Poor	Good
Patient-Doctor Relationship Questionnaire (PDRQ-9)	van der Feltz-Cornelis et al.[46]	Excellent	Fair
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Antenatal care	van der Kooy et al.[47]	Excellent	
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Labor and delivery	van der Kooy et al.[47]	Excellent	
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Postpartum	van der Kooy et al.[47]	Excellent	
Patient Evaluation of Emotional Care during Hospitalization (PEECH)	Williams et al. (2009)[48]	Excellent	
	Williams et al. (2011)[49]	Excellent	

B. Exploratory factor analysis or principal component analysis

Title of instrument	Author	Choice of EFA vs. CFA justified	Choice of EFA vs. PCA justified	Power calculation using Monte Carlo simulation	Rule of thumb method for sample size	How missing data were handled	Extraction method stated	Appropriate extraction method used	Rotation method stated	Rationale for rotation method provided	Eigenvalues or variance explained provided	Factor loadings provided	For EFA with ML, goodness of model fit assessed
Person-Centered Maternity Care Scale	Afulani et al.[25]	+	+	-	-	+	+	+	+	+	+	+	-
Patient-Provider Relationship Scale (PPRS)	Barry et al.[26]	+	-	-	-	-	N/A	N/A	-	0	-	+	N/A
Questionnaire on the Quality of Physician-Patient Interaction (QQPPI)	Bieber et al.[27]	+	+	-	-	+	+	+	+	-	+	+	-
Health Care Relationship Trust Scale	Bova et al.[28]	+	-	-	+	+	N/A	N/A	+	-	+	+	N/A
Caring Assessment Tool (CAT)	Duffy et al.[29]	+	+	-	-	+	N/A	N/A	+	-	+	+	N/A
Intensive Care Unit Psychosocial Care Scale (ICU-PU Scale)	Hariharan et al.[31]	+	+	-	-	-	+	+	+	+	+	+	-
Social Aspects of Professional Service Relationships (SAPSR)	Hausman et al.[32]	+	+	-	-	+	+	+	+	-	+	+	-
Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD)	Loblaw et al. (1999)[34]	+	+	-	-	+	N/A	N/A	+	-	+	-	N/A
	Loblaw et al. (2004)[35]	-	N/A	-	-	-	-	0	+	-	+	+	-
Communication Assessment Tool	Makoul et al.[36]	+	0	-	-	-	+	N/A	+	-	+	-	-
Patient-Physician Relationship Index (PPRI)	Ostacoli et al.[38]	+	-	-	-	+	+	N/A	-	0	+	-	N/A
Respectful Maternity Care Scale	Sheferaw et al.[40]	+	-	-	+	-	N/A	N/A	+	+	+	+	N/A
Individualized Care Scale (ICS)	Suhonen et al. (2000)[42]	+	0	-	-	-	N/A	N/A	+	-	+	+	N/A
Individualized Care Scale - Revised (ICS-R)	Suhonen et al. (2005)[43]	+	0	-	-	-	N/A	N/A	+	-	+	+	N/A
Scale of Supportive Care Given During Labor	Uludağ et al.[44]	+	0	-	-	-	-	0	-	0	+	+	-

Title of instrument	Author	Choice of EFA vs. CFA justified	Choice of EFA vs. PCA justified	Power calculation using Monte Carlo simulation	Rule of thumb method for sample size	How missing data were handled	Extraction method stated	Appropriate extraction method used	Rotation method stated	Rationale for rotation method provided	Eigenvalues or variance explained provided	Factor loadings provided	For EFA with ML, goodness of model fit assessed
WHO Health System Responsiveness Questionnaire	Valentine et al.[45]	+	+	-	-	+	+	+	+	-	+	+	-
Patient-Doctor Relationship Questionnaire (PDRQ-9)	van der Feltz-Cornelis et al.[46]	+	0	-	-	-	N/A	N/A	+	-	+	+	N/A
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Antenatal care	van der Kooy et al.[47]	+	0	-	-	+	+	+	+	-	+	-	-
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Labor and delivery	van der Kooy et al.[47]	+	0	-	-	+	+	+	+	-	+	+	-
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) - Postpartum	van der Kooy et al.[48]	+	0	-	-	+	+	+	+	-	+	-	-
Patient Evaluation of Emotional Care during Hospitalization (PEECH)	Williams et al. (2009)[49]	+	0	-	-	+	+	+	+	-	-	+	-

Rating system: + = positive, - = negative, ? = indeterminate, 0 = no information available; N/A = not applicable. See Table 3.3 for more information.
EFA = exploratory factor analysis, PCA = principal component analysis, CFA = confirmatory factor analysis, ML = maximum likelihood

C. Confirmatory factor analysis

Title of instrument	Author	Choice of EFA vs. CFA justified	If EFA and CFA performed in the same study, different samples used	Power calculation using Monte Carlo simulation to determine sample size	Rule of thumb method for sample size	Description of how missing data were handled	Estimator stated	Appropriate estimator used	Factor loadings provided	Goodness of model fit assessed
Person-Centered Maternity Care Scale	Afulani et al.[25]	+	-	-	-	+	-	0	-	-
Social Aspects of Professional Service Relationships (SAPSR)	Hausman et al.[32]	+	+	-	-	+	-	0	-	+
Princess Margaret Hospital Satisfaction with Doctor Questionnaire (PMH/PSQ-MD)	Loblaw et al. (2004)[35]	-	+	-	-	-	-	0	-	+
Communication Assessment Tool	Makoul et al.[36]	+	+	-	-	-	-	0	-	-
Perceptions of Care Adjective Checklist (PCACL-R)	Redshaw et al.[39]	+	N/A	-	-	-	+	+	+	+
Interpersonal Processes of Care - Revised	Stewart et al. (2007)[41]	+	N/A	-	-	+	+	+	-	+
Interpersonal Processes of Care - Short Form	Stewart et al. (2007)[41]	+	N/A	-	-	+	+	+	+	+
Individualized Care Scale - Revised (ICS-R)	Suhonen et al. (2005)[43]	+	-	-	-	-	-	0	-	-
Scale of Supportive Care Given During Labor	Uludağ et al.[44]	+	-	-	-	-	-	0	+	+
Patient Evaluation of Emotional Care during Hospitalization (PEECH)	Williams et al. (2011)[49]	+	N/A	-	+	-	-	0	+	+

Rating system: + = positive, - = negative, ? = indeterminate, 0 = no information available; N/A = not applicable. See Table 3.3 for more information.
EFA = exploratory factor analysis, PCA = principal component analysis, CFA = confirmatory factor analysis

Chapter 4: Respectful maternity care and disrespect and abuse: opposite ends of a single continuum or two separate dimensions? A confirmatory factor analysis to determine the dimensionality of interpersonal quality of care during childbirth

4.1. Abstract

Background: As the maternal health field is increasingly focused on interpersonal quality of care as a possible intervention point to address maternal mortality, researchers have developed quantitative instruments to measure interpersonal quality of care during childbirth. The construct is often operationalized as opposite ends along a single continuum, measured either as respectful maternity care (good care) or disrespect and abuse (poor care). However, this conceptualization may not fully reflect the underlying construct, thereby limiting the utility of these measures. The aim of this study was to determine whether these two measures of interpersonal quality of care during childbirth form a unidimensional or two-dimensional scale of the Maternal Health Interpersonal Quality Scale.

Methods: Analyses were conducted using data from two data collection methods to measure interpersonal quality of care during childbirth in two health facilities in Tanzania: observations of laboring women in study facility maternity wards and women's self-report of their delivery experience. For each dataset (observation and self-report), two confirmatory factor analyses were fit with 11 items measuring interpersonal quality of care. A one-factor model was fit with all items, and a two-factor model specified a respectful maternity care factor (five items) and a disrespect and abuse factor (six items) with a correlation between the factors. Model fit was assessed using the Comparative Fit Index (CFI), Tucker Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA).[17] A CFI $\geq$ 0.95, TLI > 0.95, and RMSEA $\leq$ 0.05 indicated excellent fit, and a CFI $\geq$ 0.90, TLI > 0.90, and RMSEA $\leq$ 0.08 indicated an adequate

fit. The chi-square difference test (DIFFTEST in Mplus) was used to compare the nested models (e.g. one-factor vs. two-factor models). Reliability of the scales was calculated using ordinal alpha.

Results: 317 observations were completed, with 269 included in the complete case analysis. In the self-report sample, 1680 women completed an exit questionnaire, with 1638 included in the complete case analysis. Results revealed a better fitting two-factor structure than a one-factor structure in both samples (observation sample: CFI: 0.98, TLI: 0.98, RMSEA: 0.03; self-report sample: CFI: 0.97, TLI: 0.96, RMSEA: 0.06). The two-factor models performed differently when measured via observation vs. self-report, with overall higher factor loadings and reliability for the self-report model.

Conclusion: Using data from two different data collection methods, the confirmatory factor analyses demonstrated a two-factor structure for the Maternal Health Interpersonal Quality Scale. This supports the hypothesis that respectful maternity care and disrespect and abuse are not opposite ends of a single continuum but rather represent two separate, correlated dimensions of interpersonal quality of care.

4.2. Introduction

In the wake of persistently high maternal mortality in low-resource settings, there is increased attention to exploring the reasons for limited maternal health progress. Qualitative evidence points to poor interpersonal quality of care during childbirth as one possible reason, with several studies aiming to verify, quantify, and measure the extent of the interpersonal quality problem.[1-3] As an emerging field in maternal health, however, there is no consensus on the conceptualization of the construct of interpersonal quality of care during childbirth, nor is there a gold standard for measuring it.

Maternal health researchers typically operationalize interpersonal quality of care using two aspects of the construct: good interpersonal quality, called respectful maternity care and poor interpersonal quality, called disrespect and abuse. Respectful maternity care is defined as: "respect for women's basic human rights, including respect for women's autonomy, dignity, feelings, choices, and preferences".[4,p(1)] More specifically, it is conceptualized as a practice of good interpersonal care, often framed as the absence of disrespectful and abusive treatment.[4,5] On the other hand, researchers define disrespect and abuse during childbirth as interpersonal interactions deemed to be humiliating or undignified.[6] These include interactions that are both normalized by patients and providers and those that are agreed upon as disrespectful or abusive by local consensus.[6]

The majority of the studies operationalize and measure poor, as opposed to good, interpersonal care during childbirth. Five studies in health facilities in sub-Saharan Africa reported prevalence of disrespect and abuse during childbirth ranging widely from 15% to 98%.[2,3,7-9] This poor treatment included physical privacy violations, verbal abuse, physical abuse, and inappropriate

demands for payment. Two studies measured respectful maternity care, as opposed to disrespect and abuse, as adherence to respectful standards during labor and delivery such as proper communication, friendly support of patients, and encouraging positive birthing behaviors (e.g. assumption of different labor positions).[5,10] For example, using observations of laboring women in health facilities in six sub-Saharan African countries, one study noted that 66-93% of women were supported by providers in a friendly way and 60-95% were greeted in a respectful manner.[5] In a study in Ethiopia, women reported that 66% of nine respectful maternity care indicators were performed during their deliveries.[10]

The current operationalization of the construct suggests that respectful care and disrespect and abuse are opposite ends of a single dimension. However, it is not clear that the presence of respectful care (positive aspects) indicates the absence of disrespectful care (negative aspects), and vice versa. Labor and delivery is a complex process, introducing the possibility that a provider is supportive in one moment, and verbally abusive in the next. For example, in one study that measured respectful maternity practices, observers also noted that the same women not only experienced positive care, but also negative treatment, such as verbal abuse, during the course of childbirth.[5] This study points to the complexity of measuring women's interpersonal quality of care experiences and suggests that disrespectful treatment and respectful care may be two separate but correlated dimensions. Using either a negative or a positive lens to describe and measure interpersonal quality may be too restrictive to provide evidence for the full construct and to understand its effects. For example, it is possible that positive aspects and negative aspects of care may have different effects on maternal health outcomes, satisfaction with care, and perceptions of the health system. Thus, determining whether the construct of interpersonal

quality of care is one or two dimensions would impact how best to measure its effects and how to meaningfully target interventions.

The aim of this study was to determine whether questions about positive and negative aspects of interpersonal quality of care during childbirth form a unidimensional or two-dimensional scale. This was done using confirmatory factor analysis using data from two samples of women in Tanzania.

4.3. Methods

Data source: sample and design

Data for this aim come from the Staha Study, a cross-sectional study in the Tanga Region of Tanzania, which aimed to measure disrespect and abuse during facility-based childbirth and to develop and evaluate an intervention to address the problem. Two hospitals in the Tanga Region, Magunga Hospital in Korogwe District and Teule Hospital in Muheza District, were chosen for the study.

The study utilized data from two data collection methods to measure interpersonal quality of care during childbirth: observations of laboring women in study facility maternity wards and women's self-report of their delivery experience. For the observations, women who presented to the facility for childbirth, were at least 15 years of age, and were in active labor were eligible to participate. Trained nurse observers unaffiliated with the hospitals observed women from active labor to two hours postpartum. Observers worked in three 8-hour shifts to ensure 24-hour coverage in the maternity wards. One observer was assigned to each woman. Another observer

took over the observation if it was not complete within an 8-hour shift. For the self-report data, after discharge, women were approached to participate in an exit interview. Women who delivered in study facilities and were at least 15 years of age were eligible to participate. Women completed a closed-ended questionnaire about their delivery experience. Observation data were collected from September-October 2012 (N=317) and from November-December 2015 (N=357). Self-report data were collected from March-September 2015 (N=1680). See Table 4.1 for the sample sizes from the different waves of data collection. To maximize sample size and the number of items included in the confirmatory factor analyses, the observation data across the two time periods (Dataset 1 and Dataset 3) were combined and self-report data collected from Dataset 2 only were used for analysis. As further outlined in the statistical analysis section, I determined whether the observation data from the two time periods could be combined by examining the distribution of the items included in the confirmatory factor analysis and by running confirmatory factor analyses separately by time period.

All data were collected in Swahili. All participants provided informed consent. The study was approved by the Institutional Review Boards of Columbia University, Ifakara Health Institute, and the National Institute for Medical Research in Tanzania. More information about the study is available elsewhere.[2,11]

Measures

Items included in the confirmatory factor analysis

The observation instrument included 11 respectful maternity care (positive) and 14 disrespect and abuse (negative) items. Observers noted if any of these events happened during labor and delivery to the women they observed, and responses were recorded dichotomously as yes/no.

The self-report instrument included 9 respectful maternity care items (five of which overlapped with the observation instrument) and the same 14 disrespect and abuse items. In the exit interview, women were asked if they experienced these specific events during labor and delivery. Each item was asked as a separate question and categorized dichotomously (yes/no for the respectful maternity care questions, experienced/not experienced for the disrespect and abuse questions).

The majority of the respectful maternity care items were taken from the Johns Hopkins Maternal and Child Health Integrated Program's (MCHIP) Maternal and Newborn Quality of Care Survey and were supplemented with additional items included developed by the Staha study team.[12] Items covered the domains of communication, support during labor and delivery, and physical privacy. The disrespect and abuse items were based on Bowser and Hill's Disrespect and Abuse Landscape Analysis.[13] The Staha study team adapted the disrespect and abuse items for the cultural context and established face validity and content validity by expert opinion and focus group discussions with women in the study area. Items included in the study instruments covered the categories of physical privacy violations, verbal abuse, physical abuse, neglectful care, non-consented care, and inappropriate demands for payment. Table 4.2 provides a list of the items available in the observation and self-report instruments.

Descriptive variables

To provide the context in which I am examining the construct of interpersonal quality of care, demographic and delivery experience variables were chosen to describe the cohorts of women. In addition, in previous studies, these factors were associated with reports of disrespect and

abuse.[2,11] Thus, differences in these factors between the two samples may have contributed to how the confirmatory factor analyses performed. Demographic characteristics included age (continuous and categorical as 15-19, 20-34, ≥ 35), parity (first birth, 2-3 births, 4 or more births), education (attended secondary education or greater vs. less than secondary education), marital status (married vs. single or divorced/widowed), socioeconomic status, reported low mood or depression in the last 12 months (single item), and reported ever being physically abused or raped. Socioeconomic status was measured from 18 household asset questions from the exit questionnaire using a principal component analysis.[14] The principal component analysis index was split into quintiles, with the lowest 2 quintiles classified as poor. Delivery experience factors included having a Caesarean section, having any complications during childbirth, and length of stay for delivery (≤ 1 day vs. > 1 day). Complications during childbirth included experiencing any of the following: extreme pain, high blood pressure, seizures, blurred vision, severe headaches, swelling in the hands/feet, baby was in distress or too large, long labor (> 12 hours), excessive bleeding, or infection/fever. For the observation data, age, parity, and Caesarean section information were collected with the observation instrument. Women who were observed during labor and delivery were invited to participate in an exit questionnaire after discharge through which the remaining variables were collected for the observation sample. As not all participants who were observed participated in the exit questionnaire, only 77% of the observation sample has the remaining descriptive data available. For the self-report data, all items were collected with the exit questionnaire.

Statistical analysis

Data preparation

To account for the Hawthorne effect (the impact of the presence of observers on provider behavior), data from the first week of the observations were eliminated. It has been suggested that providers acclimate to the presence of observers after 10-15 patient interactions.[15]

To allow for informative findings, measurement items that had few endorsements (< 5) were eliminated from the analysis. In both the observation and self-report samples, these items included non-consent for procedures (tubal ligation, hysterectomy, Caesarean section), sexual harassment, rape, and detention. "Suggested or asked for a bribe" and "threatened to withhold treatment" were additionally excluded for the self-report sample. These items are not common in the Tanzanian setting or may be subject to social desirability bias.[2] I used a complete case analysis based on the remaining items available for each data collection method.

Descriptive analysis

I explored the frequency and distribution of the items to be included in the confirmatory factor analyses and of the descriptive variables separately for the observation and self-report samples. The observation sample was also stratified by time period. For the observation data, while the underlying factors should have remained invariant over time, there was an intervention implemented between the two time periods to reduce disrespect and abuse. Therefore, it was possible that this and other contextual factors could have contributed to changes in the measurement properties. I first ran two separate confirmatory factor analyses by time period for the observation data to determine if there were any time differences. Preliminary analyses revealed differences by time period both for the endorsement of the items and for the

confirmatory factor analyses. I therefore chose to use the observation data from Dataset 1 only. See Appendix 4.1 for the justification for this decision.

Confirmatory factor analyses

I performed two confirmatory factor analyses for data from each data collection method (observation and self-report): a one factor model underlying all of the items and a two-factor model specifying a respectful maternity care factor and a disrespect and abuse factor with a correlation between the factors. Because I tested a specific hypothesis, in this case, a one-factor vs. two-factor structure of the construct, I used a confirmatory factor analysis rather than an exploratory factor analysis.[16,17] For comparability, all models were fit with 11 overlapping items that were available from each data collection method: five respectful maternity care items and six disrespect and abuse items (Table 4.2).

All disrespect and abuse items were reverse coded in the confirmatory factor analyses for ease of interpretation. The confirmatory factor analyses were conducted in Mplus version 8.0 (Muthen & Muthen) using weighted least squares with mean and variance adjustment (WLSMV), which estimates the model using a tetrachoric correlation matrix. WLSMV is the default estimator for dichotomous dependent variables. I scaled the models using a standardized approach, fixing the first loading to 1.0. The one-factor model had 22 free parameters (11 items: 11 factor loadings, 11 error variances), with 44 degrees of freedom. The two-factor model estimated 23 free parameters (11 factor loadings, 11 error variances, 1 correlation between the factors) with 43 degrees of freedom.

The direction, magnitude, and statistical significance of the factor loadings of the models were examined. Large positive factor loadings (> 0.70) indicate that the item is a good reflection of the underlying construct.[18] However, items that are conceptually important with factor loadings greater than 0.30 may be retained if statistically significant. Because the chi-square fit statistic tends to over-reject models with large sample sizes, I examined the goodness of model fit using fit indices that are not influenced by sample size, including the Comparative Fit Index (CFI), Tucker Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA).[16] A CFI ≥ 0.95, TLI > 0.95, and RMSEA ≤ 0.05 indicate excellent fit, and a CFI ≥ 0.90, TLI > 0.90, and RMSEA ≤ 0.08 indicate an adequate fit.[19] The chi-square difference test (DIFFTEST in Mplus) was used to compare the nested models (e.g. one-factor vs. two-factor models). If the models did not fit well, I considered re-specifying the models by using modification indices or removing items with low factor loadings (e.g. <0.30). Modification indices, which estimate the improvement in model fit if parameters are freely estimated, were only considered if they indicated residual variance between similar items. Once the models were specified and the final models were chosen, the reliability of each factor was calculated by an ordinal alpha in RStudio version 1.0.136, *Psych* package, using tetrachoric correlation coefficients.[20] Factors were considered reliable if the ordinal alpha was ≥ 0.70.[21]

Validity analyses

To examine how the construct of interpersonal quality of care during childbirth may be similar or different when measured via observation vs. self-report, I qualitatively compared the results of the final confirmatory factor analysis models from the different data collection methods, specifically the number of factors, model fit, and the strength of the factor loadings.

To further examine the construct validity of the scales, I explored convergent validity (strong correlations between similar constructs) and divergent validity (weak correlations between dissimilar constructs) using polychoric correlations, appropriate for assessing correlations with categorical variables.[22] Variables to measure convergent and divergent validity were only available for the self-report instrument. As evidence of convergent validity, I hypothesized a high correlation (>0.70) between the self-report confirmatory factor analysis latent factor(s) and a single-item 5-category variable measuring women's ratings of the respect providers showed them for delivery (1-excellent, 2-very good, 3-good, 4-fair, 5-poor). To illustrate that interpersonal quality of care is distinct from the construct of satisfaction, I correlated the self-report latent factor(s) and a single-item 4-category variable (1-very satisfied, 2-somewhat satisfied, 3-somewhat dissatisfied, 4-very dissatisfied) of women's self-report of their satisfaction with their delivery experience. As interpersonal quality of care and satisfaction are conceptually related, relative to the convergent validity correlation, I would expect a lower correlation, as evidence of divergent validity.

Secondary analysis

As a secondary analysis, to examine the robustness of the factor structures of the final confirmatory factor analysis models, I evaluated the impact on the fit statistics of adding non-overlapping items to the models. As the number of items differed between the final models and these more saturated models, the models cannot be formally compared. For the observation models, six respectful maternity care items and two disrespect and abuse items were added. For the self-report models, four respectful maternity care items were added.

Sample size and power

I conducted power calculations for the one-factor and two-factor confirmatory factor analyses for each data collection method to determine the minimal effect size discernable given the specified sample sizes for the self-report data and the observation data. The sample sizes excluded missing data, and the observation data further excluded data for the Hawthorne effect. Power calculations were performed using Monte Carlo simulation in Mplus.[23] For the observation data, with a sample size of 269 and an alpha=0.05, the one-factor model would have at least 0.80 power to detect factor loadings of 0.32 and residual variance of 0.90, while the two-factor model would have at least 0.80 power to detect factor loadings of 0.40, residual variance of 0.84, and a correlation between the two factors of 0.40. For the self-report data, with a sample size of 1638 and an alpha=0.05, both the one-factor and two-factor model at least 0.80 power to detect factor loadings of 0.30, residual variance of 0.91, and 0.30 correlation between the factors, where applicable. For factor loadings, 0.30 is the minimum correlation that is indicative that the item is a good reflection of the underlying construct.[16] See Appendix 4.2 for the power and coverage per model and dataset.

4.4. Results

Sample description

For the observation data from Dataset 1, 317 women at the two study facilities were observed during labor and delivery. Of these, 26 were removed from analysis for the Hawthorne effect. Using a complete case analysis, the sample size was reduced to 269 (7.6% missing). For the self-report data, 1680 participated in the exit survey, and 1638 had complete data for this analysis (2.5% missing). Characteristics of the total study sample and complete case sample for each data collection method are shown in Table 4.3. In both samples, women were on average 26 years of

age (SD: 6) and about 40% (observation: N=105, self-report: N=707) delivered their first child. There were some significant differences in the background characteristics between the two samples. A higher proportion of women in the observation sample had a Caesarean section (observation: 12.6% vs. self-report: 1.7%, p<0.0001) and experienced any complications during childbirth (observation: 49.3% vs. self-report: 37.3%, p=0.001). It is possible that those who had a Caesarean section or experienced complications were less likely to participate in the exit questionnaire due to the difficulty to sit for an interview after surgery or complications or that there was a reduction in the Caesarean section rate over time. The self-report sample had a higher proportion of married participants (N=1421, 86.8%) compared to the observation sample (N=157, 75.9%, p<0.0001).

Table 4.4 shows the frequency and distribution of the items included in the confirmatory factor analyses for the total sample and complete case for each data collection method. For the overlapping items, RMC5: "supported during labor in a friendly way" (observation: N=218, 81.0%; self-report: N=1276, 77.9%) was the most endorsed respectfully maternity care item. For the disrespect and abuse items, "health providers shouted at or scolded woman" (DA2) was the most endorsed item in both samples (observation: N=121, 45.0%; self-report: N=88, 5.4%) followed by "health providers threatened woman for any reason or made negative or disparaging comments about the woman" (DA3) for the observation sample (N=119, 44.2%) and "health providers ignored or abandoned woman when in need or when she called for help" (DA4) for the self-report sample (N=84, 5.1%).

Confirmatory factor analyses

Table 4.5 shows the results of the confirmatory factor analyses for one-factor and two-factor models of the 11 overlapping items for each data collection method.

Observation sample

Neither the one-factor model nor the two-factor model fit well. Both models followed a similar pattern of factor loadings. (Appendix 4.3). I removed DA1 (women's body seen by other people during delivery), a weak indicator that loaded as 0.13 on the one-factor model and 0.16 on the two-factor model and then reran the models. Both models still fit poorly (Appendix 4.3). To further investigate whether the model fit could be improved, I explored the modification indices. In the two-factor model, the residual variance between DA4 (health providers ignored or abandoned woman when in need or when she called for help) and DA5 (delivered without any assistance) was associated with the highest modification index (MI=50.75). This may be due to the fact that both items measure aspects of neglect. In order to reduce redundancy, I combined DA4 and DA5 into one item ("neglect") and reran the models. The two-factor model then fit well (Table 4.5), with RMSEA = 0.03 (90% CI: 0.00, 0.06), CFI = 0.98, and TLI = 0.98. In comparison with the one-factor model, the two-factor model fit significantly better (x^2 = 24.42, df = 1, p<0.0001). While all items loaded significantly on the two factors, the neglect item had the lowest loading at 0.27, just below the cut point of 0.30 indicating that the item is a meaningful, though weak, reflection of the underlying construct. In the two-factor model, the reliability of the respectful maternity care factor was ordinal α = 0.77 and of the disrespect and abuse factor was ordinal α = 0.70 (Table 4.6).

Self-report sample

The one-factor model and two-factor model both fit well and all items loaded significantly (Table 4.5). The two-factor model fit better, with RMSEA = 0.06 (90% CI: 0.05, 0.06), CFI = 0.97 and TLI = 0.96. A chi-square difference test comparing the one-factor and two-factor models confirmed that the two-factor model fit better (x^2 = 65.33, df = 1, p<0.0001). There was a high correlation of 0.75 between the factors. Moving from the one-factor to two-factor model, the factor loadings increased for each item in the model. The highest loading item at 0.99 on the disrespect and abuse factor was DA4: "health providers ignored or abandoned woman when in need or when she called for help." The lowest loading item at 0.32 was RMC1: "respectfully greeted when you arrived at the maternity ward" on the respectful maternity care factor. There was also a high correlation between the factors of 0.75. In the two-factor model, the reliability of the respectful maternity care factor was ordinal α = 0.81 and of the disrespect and abuse factor was ordinal α = 0.92 (Table 4.6).

Validity analyses

Qualitative comparison: observation and self-report models

The observation and self-report samples both revealed two-factor, well-fitting models, but several differences between the two models emerged. First, DA1 (women's body seen by other people during delivery) was eliminated from the observation model, yet it loaded highly (0.71) in the self-report model. Similarly, the combined "neglect" item factor loading was 0.27 for the observation model, but each of these items had a high factor loading in the self-report model (health providers ignored or abandoned woman when in need or when she called for help = 0.91, delivered without any assistance = 0.99). Second, there was a higher correlation between the two

factors for the self-report sample (0.75) than the observation (0.50). Third, the self-report model generally had higher loadings and the scales had higher reliability than the observation model. When comparing the factor loadings of the same items between the models, the disrespect and abuse items all had higher loadings for the self-report (range: 0.71-0.99) than the observation (range: 0.27-0.89) data. For respectful maternity care items, "respectfully greeted when you arrived at the maternity ward" (RMC1) had a high loading for the observation (0.77) but a low factor loading for the self-report (0.32), while "asked if you had any questions during stay in maternity ward" (RMC2) and "encouraged to consume liquids/food throughout labor" (RMC4) had high factor loadings for self-report (RMC2: 0.71, RMC4: 0.83) but low factor loadings for observation (RMC2: 0.37, RMC4: 0.34).

Convergent and divergent validity

Table 4.7 shows measures of convergent and divergent validity for each of the latent factors from the two-factor self-report model. Convergent validity was tested by correlating a single-item question that asked women to rate the respect providers showed them for delivery. The correlation with the respectful maternity care factor was 0.48 and with the disrespect and abuse factor was 0.79. The correlation between women's rating of satisfaction with their delivery—a hypothesized measure of divergent validity—and the respectful maternity care factor was 0.50, and it was 0.79 with the disrespect and abuse factor. These results were surprising and might indicate that the disrespect and abuse factor is more predictive of global measures of the delivery experience than the respectful maternity care factor.

Secondary analysis

To examine the robustness of the two-factor structure**,** the two-factor models were fit with additional available items.

Observation sample

The two-factor observation model was fit with eight additional items, six additional respectful maternity care items and two disrespect and abuse items that were collected with the observation instrument. The factor loadings of the additional items ranged from 0.31-0.76. The inclusion of the additional items worsened the fit, with fit statistics below the cut off for sufficient fit (Table 4.8).

Self-report sample

The two-factor self-report model was examined with four additional respectful maternity care items that were available from the exit questionnaire. This resulted in a Heywood case (factor loading greater than 1.00) for DA4, demonstrating that the model was overfit.[16] The modification indices indicated that the model fit would improve by loading RMC15 on the disrespect and abuse factor (MI=95.12). Interestingly, RMC15, "providers came quickly when called for them" is worded as the reverse of DA4 (health providers ignored or abandoned woman when in need or when she called for help). Compared with DA4, RMC15 was not a strong indicator. Thus, RMC15 was removed from the model, which resulted in a well-fitting model (Table 4.8). The additional items had low to moderate factor loadings (RMC12: 0.57, RMC13: 0.27, RMC14: 0.45). The fit of the original self-report model did not change substantially in the more saturated model, indicating that the two-factor structure were robust to the additional items.

4.5. Discussion

The aim of this study was to test the factor structure and construct validity of a measure of interpersonal quality of care during childbirth, the Maternal Health Interpersonal Quality Scale, using confirmatory factor analysis. In data collected in two different ways – observation of laboring and delivering women and women's self-report of their delivery experience – the confirmatory factor analyses demonstrated that a two-factor structure better represented the construct of interpersonal quality of care during childbirth. The two-factor structure supports the hypothesis that respectful maternity care and disrespect and abuse are not opposite ends of a continuum but two separate, correlated dimensions. Particularly in the self-report two-factor model, there was a high correlation between the factors (0.75), representing that disrespect and abuse and respectful maternity care are part of the same construct, measured in two different ways. Measuring either positive or negative aspects of interpersonal quality of care is insufficient to understand the full experience of women during childbirth. These findings illustrate that the absence of disrespectful and abusive treatment does not indicate the presence of respectful maternity care. Thus, intervening to address disrespect and abuse, as some studies did,[11,24] may reduce disrespectful and abusive practices but will not necessarily result in respectful maternity care. Moreover, each factor may have different effects on maternal health outcomes, and thus needs to be considered separately when measuring interpersonal quality of care during childbirth and when designing interventions.

For the self-report sample, while the two-factor model fit better than the one-factor model based on the chi-square difference test, the factor loadings and fit statistics of the models were very similar. Thus, deciding when to use a one-factor vs. two-factor model when measuring

interpersonal quality of care via self-report may be outcome- or intervention-specific. However, as stated above, a limitation of the one-factor scale is that it may miss important differences that are specific to the effects of either respectful maternity care or disrespect and abuse.

This study contributes to the growing body of validated measures for interpersonal quality of care during childbirth, specifically in low-resource settings. Two recent studies, in Ethiopia and Kenya, proposed scales for measuring interpersonal quality of care during childbirth from women's self-report using exploratory factor analysis.[25,26] The Ethiopia study extracted four factors from 15 items—friendly care, abuse-free care, timely care, and discrimination-free care—two of which overlap with the factors proposed in this analysis.[26] In the Kenya study, there were three factors from 30 items: dignified and respectful care, communication and autonomy, and supportive care.[25] The dignified and respectful care factor included both positive and negative aspects of care. However, direct comparison of the measures is limited by the differing items included in each measure. Further, these studies used data-driven, rather than hypothesis-driven approaches to assess the factor structure. Most similar to this study's findings, one study in the United Kingdom, using a confirmatory factor analysis, described a two-dimensional correlated scale for women's perceptions of childbirth, with one factor consisting of positive adjectives (e.g. considerate, supportive, polite) and one of negative adjectives (e.g. rude, unhelpful, insensitive) to describe the birth experience.[27]

In the present study, the two-factor models performed differently when measured via observation compared to self-report. When comparing the two models, there was a difference in the importance of the items, represented by the strength of the item loading, and in the correlation of the two factors. For example, items that loaded highly in the self-report model, such as those

related to neglect and privacy, had low factor loadings in the observation model. This may reflect which method of data collection is better able to measure the individual items, as further discussed below. The self-report model also showed higher internal consistency than the observation model. In the self-report model, adding items demonstrated that the fit was consistent with and robust to the two-factor structure. In contrast, for the observation model, adding items to the model worsened the fit, indicating that the model was not robust to the two-factor structure. Overall, this may indicate that the observation model may need further refinement and that more work is needed to understand the construct when measured via observation methods.

Several reasons may explain the differences between the two models. First, some of the items, such as physical privacy or being greeted respectfully when arriving in the maternity ward, are more subjective in nature and may be interpreted differently by the person experiencing it than by an observer. Second, there were some statistically significant differences in the characteristics of the women in the samples. The observation sample had a higher proportion of participants who had a Caesarean section or had any complication during delivery. One study found that women who have complications during childbirth are more likely to be treated poorly during childbirth, while women who have a Caesarean section are less likely to be treated poorly.[2] However, it is unclear how these factors would influence the factor structure. Third, observer bias may influence the observation measure. While all observers received the same training, their assessments may have been colored by their own experiences or interpretations. Fourth, for the self-report measure, women may be prone to report the socially desirable choice (social desirability bias) or, since they were interviewed soon after childbirth, their responses may be

influenced by the halo effect, with initial reflections of the birth experience being more positive.[28] These would result in an underestimate of the true prevalence of the negative items and an overestimate of the true prevalence of the positive items. In addition, poor treatment of women during childbirth is often normalized by both patients and providers.[29] This may also lead to an underestimate of reports of the negative items by the women. Indeed, in this study, the prevalence of disrespect and abuse when measured via observation was higher than on self-report, indicating possible evidence of normalization and social desirability bias. While it is possible that this difference is a result of the intervention that was implemented to decrease disrespect and abuse between the two data collection periods, this finding is consistent with another study that found a higher prevalence of disrespect and abuse when measured via observation than when measured via self-report in the same women.[30] However, it is not exactly clear how this difference in reporting would affect the factor loadings. Exploration of the factor structure using observation and self-report measures in the same women would provide a more direct comparison of the measurement methods to understand how the instrument performs and the importance of the individual items to the underlying construct.

My hypotheses for single-item questions to correlate with the scales as evidence of convergent and divergent validity were not supported by the data. Correlation patterns emerged by latent factor, rather than by hypothesized measures of convergent or divergent validity, with single-item questions assessing the respect providers showed women for delivery and satisfaction with delivery care correlating highly with the disrespect and abuse factor, but moderately with the respectful maternity care factor. These results were surprising but may indicate that the

disrespect and abuse factor was more reflective of women's global rating of their delivery experience than was the respectful maternity care factor.

This study had several limitations. First, time effects and data quality concerns resulted in using a smaller observation sample size than initially intended. Preliminary analyses using all of the observation data yielded a poorly fitting model. This may be due to the intervention that occurred to reduce disrespect and abuse during the two time periods. Thus, results of the observation confirmatory factor analysis should be interpreted with caution, and larger sample sizes measuring interpersonal quality of care during childbirth via observation should be used to test the findings. Second, in comparing the observation and self-report confirmatory factor analyses, I was limited by the items that overlapped between the two instruments. Third, convergent and divergent validity were tested only for the self-report model and with single-item variables, as opposed to using several items or a scale. Using more robust measures of respect and satisfaction to test convergent and divergent validity for both data collection methods would provide more confidence in the construct validity of the scales. Fourth, as mentioned, the factor analysis would have been strengthened by comparing the same women in the observation and the self-report measures to eliminate any variability due to demographic characteristics.

Conclusion

This study identified two correlated, but separate, dimensions of interpersonal quality of care during childbirth: respectful maternity care and disrespect and abuse. Thus, studies to date that measured either respectful maternity care or disrespect and abuse have laid the groundwork for measurement and intervention but may be too narrow in focus. This context-specific, full

measure of the construct expands the boundaries of interpersonal quality of care during childbirth and has the potential to appropriately guide, test, and evaluate interventions.

4.6. References

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

2. Kruk ME, Kujawski S, Mbaruku G, Ramsey K, Moyo W, Freedman L. Disrespectful and abusive treatment during facility delivery in Tanzania: a facility and community survey. *Health Policy Plan.* 2014:czu079.

3. Abuya T, Warren CE, Miller N, et al. Exploring the prevalence of disrespect and abuse during childbirth in Kenya. *PLoS One.* 2015;10(4):e0123606.

4. White Ribbon Alliance. *Respectful Maternity Care Charter: The Universal Rights of Childbearing Women.* Washington, DC: White Ribbon Alliance; 2011. http://whiteribbonalliance.org/wp-content/uploads/2013/10/Final_RMC_Charter.pdf

5. Rosen HE, Lynam PF, Carr C, et al. Direct observation of respectful maternity care in five countries: a cross-sectional study of health facilities in East and Southern Africa. *BMC Pregnancy Childbirth.* 2015;15:306.

6. Freedman LP, Ramsey K, Abuya T, et al. Defining disrespect and abuse of women in childbirth: A research, Policy and rights agenda. *Bull World Health Organ.* 2014;92:915-917.

7. Asefa A, Bekele D. Status of respectful and non-abusive care during facility-based childbirth in a hospital and health centers in Addis Ababa, Ethiopia. *Reproductive health.* 2015;12:33.

8. Okafor II, Ugwu EO, Obi SN. Disrespect and abuse during facility-based childbirth in a low-income country. *Int J Gynaecol Obstet.* 2014;128:110-113.

9. Sando D, Kendall T, Lyatuu G, et al. Disrespect and abuse during childbirth in Tanzania: are women living with HIV more vulnerable? *J Acquir Immune Defic Syndr Hum Retrovirol.* 2014;67 Suppl 4:S228-234.

10. Sheferaw ED, Bazant E, Gibson H, et al. Respectful maternity care in Ethiopian public health facilities. *Reprod Health.* 2017;14:60.

11. Kujawski S, Freedman L, Ramsey K, et al. Community and health system intervention to reduce disrespect and abuse during childbirth in Tanga Region, Tanzania: A comparative before-and-after study. *PLoS Med.* 2017;14.

12. Maternal and Child Health Integrated Program, USAID. Maternal and Newborn Quality of Care Survey: Labor and Delivery Observation Checklist. http://www.mchip.net/sites/default/files/mchipfiles/QoCLDObservation_0.pdf.

13. Bowser D, Hill K. *Exploring Evidence for Disrespect and Abuse in Facility-Based Childbirth: Report of a Landscape Analysis.* Washington, DC: Harvard School of Public Health, University Research Co., LLC; 2010

14. Filmer D, Pritchett L. Estimating wealth dffects without expenditure data--or tears: An application to educational enrollments in states of India. *Demography.* 2001;38(1):115-132.

15. Leonard K, Masatu MC. Outpatient process quality evaluation and the Hawthorne Effect. *Soc Sci Med.* 2006;63:2330-2340.

16. Brown TA. *Confirmatory Factor Analysis for Applied Research.* 2nd ed. New York: The Guilford Press; 2015.

17. Byrne BM. Factor analytic models: viewing the structure of an assessment instrument from three perspectives. *J Pers Assess.* 2005;85(1):17-32.

18. Furr RM, Bacharach VR. *Psychometrics: An Introduction.* 2nd ed. Thousand Oaks, California: Sage Publications, Inc; 2013.

19. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling.* 1999;6:1-55.

20. Revelle W. *Package 'psych': Procedures for Psychological, Psychometric, and Personality Research.* Evanston, Illinois: Northwestern University; 2017.https://CRAN.R-project.org/package=psych

21. Kline RB. *Principles and Practice of Structural Equation Modeling.* 3rd ed. New York: The Guilford Press; 2011.

22. Raykov T, Marcoulides GA. *Introduction to Psychometric Theory.* New York: Routledge; 2011.

23. Muthén LK, Muthén BO. How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Struct Equ Modeling.* 2002;9:599-620.

24. Abuya T, Ndwiga C, Ritter J, et al. The effect of a multi-component intervention on disrespect and abuse during childbirth in Kenya. *BMC Pregnancy Childbirth.* 2015;15:224.

25. Afulani PA, Diamond-Smith N, Golub G, Sudhinaraset M. Development of a tool to measure person-centered maternity care in developing settings: validation in a rural and urban Kenyan population. *Reprod Health.* 2017;14:118.

26. Sheferaw ED, Mengesha TZ, Wase SB. Development of a tool to measure women's perception of respectful maternity care in public health facilities. *BMC Pregnancy Childbirth.* 2016;16:67.

27. Redshaw M, Martin CR. Validation of a perceptions of care adjective checklist. *J Eval Clin Pract.* 2009;15:281-288.

28. Simkin P. Just another Day in a woman's life? Part II: Nature and consistency of women's long-term memories of their first birth experiences. *Birth.* 1992;19:64-81.

29. Freedman LP, Kruk ME. Disrespect and abuse of women in childbirth: Challenging the global quality and accountability agendas. *Lancet.* 2014;384:e42-e44.

30. Banks KP, Karim AM, Ratcliffe HL, Betemariam W, Langer A. Jeopardizing quality at the frontline of healthcare: prevalence and risk factors for disrespect and abuse during facility-based childbirth in Ethiopia. *Health Policy Plan.* 2017.

4.7. Tables and figures

Table 4.1. Sample sizes for observation and self-report data collection from two health facilities in Tanga Region, Tanzania, Staha Study, 2012-2015

	Dataset 1: Sept-Oct 2012			**Dataset 2: March-Sept 2015**			**Dataset 3: Nov-Dec 2015**		
	N	Positive items?	Negative items?	N	Positive items?	Negative items?	N	Positive items?	Negative items?
Observation	**317**	**Yes**	**Yes**	N/A	N/A	N/A	357	Yes	Yes
Self-report	240	No	Yes	**1680**	**Yes**	**Yes**	299	Yes	Yes

*Bolded numbers are those used in the final analyses

Table 4.2. Interpersonal quality of care items collected in the observation instrument and self-report instrument

	Observation instrument	Self-report instrument
Respectful maternity care questions		
Respectfully greeted the pregnant woman	x	x
Encouraged the woman to have a support person present throughout labor and birth	x	
Asked woman if she had any questions	x	x
Responded to woman's questions and concerns	x	
Informed the woman of preliminary exam findings	x	
At least once, explained what will happen in labor to the woman	x	
At least once, encouraged the woman to consume liquids/food throughout labor	x	x
At least once, encouraged to walk or assume different positions during labor	x	x
Supported the woman during labor in a friendly way	x	x
Draped woman or gave woman something to cover her during 1st stage of labor	x	
Privacy assured during examination	x	
Providers came quickly when called for them		x
Assisted to use toilet facilities		x
Able to communicate with your relatives or other persons who accompanied you		x
Received any assistance to reduce pain		x
Disrespect and abuse questions		
Woman's body seen by other people (aside from health provider) during delivery	x	x
Health providers shouted at or scolded woman	x	x
Health providers threatened to withhold treatment because woman could not pay or did not have supplies	x	x
Health providers threatened woman for any reason or made negative or disparaging comments about the woman	x	x
Health providers ignored or abandoned woman when in need or when she called for help	x	x
Delivered without any assistance	x	x
Tubal ligation (tying of fallopian tubes) without patient's permission	x	x
Hysterectomy (getting your uterus removed) without patient's or her relatives' permission	x	x
Caesarean section without patient's or her relatives' permission	x	x
Hit, slapped, pushed, pinched or otherwise beat the woman	x	x
Health providers sexually harassed woman or made sexual advances (for example, inappropriate touching or sexual comments that made woman feel uncomfortable)	x	x
Rape (being forced to have intercourse or perform any other sexual acts against your will by someone other than your husband)	x	x
Woman or baby not allowed to leave the hospital due to failure to pay (detention)	x	x
Health providers suggested or asked for a bribe or informal payment for better care	x	x

Table 4.3. Characteristics of the observation sample and self-report sample of women who delivered at two facilities in Tanga Region, Tanzania, 2012-2015

	Observation data					Self-report data					Comparison between complete case samples
	Total sample (N=291)[a]			Complete case (N=269)		Total sample (N=1,680)			Complete case (N=1,638)		
	N	%	% missing	N	%	N	%	% missing	N	%	p-value
Demographics											
Age, mean (SD)	25.5	6.4	1.4	25.5	6.4	25.7	6.2	0.0	25.7	6.2	0.70
Age categories											
15-19	55	18.9		51	19.2	275	16.4		267	16.3	
20-34	198	68.0	1.4	183	69.1	1196	71.2	0.0	1168	71.3	0.54
35+	34	11.7		31	11.7	209	12.4		203	12.4	
Parity											
1st birth	113	38.8		105	39.0	723	43.0		707	43.2	
2-3 births	107	36.8	0.0	99	36.8	601	35.8	0.0	584	35.7	0.38
4 or more births	71	24.4		65	24.2	356	21.2		347	21.2	
Attended secondary education or greater	53	23.8	0.0	48	23.2	466	27.8	0.1	455	27.8	0.16
Married	169	75.8	0.0	157	75.9	1459	86.9	0.0	1421	86.8	<0.0001
Poor (lowest 40% of wealth index)	92	42.0	1.8	89	43.8	665	39.8	0.6	647	39.7	0.26
Household has electricity	58	26.0	0.0	53	25.6	676	40.2	0.0	661	40.4	<0.0001
Household has mobile phone	198	88.8	0.0	183	88.4	1575	93.8	0.0	1538	93.9	0.003
Reported low mood or depression in last 12 months	67	30.0	0.0	64	30.9	423	25.2	0.0	410	25.0	0.07
Reported ever being physically abused or raped	5	2.2		4	1.9	55	3.3	0.1	53	3.2	0.29
Delivery care experience											
Caesarean section	39	13.5	0.7	34	12.6	32	1.9	0.1	28	1.7	<0.0001
Reported any complications during childbirth	109	48.9	0.0	102	49.3	633	37.7	0.0	611	37.3	0.001
Length of stay for delivery ≤ 1 day	79	35.8	0.9	71	34.6	498	29.6	0.0	486	29.7	0.15

[a]Age, parity, and Caesarean section collected with the observation instrument (N=291); remaining variables collected with the exit questionnaire (N=223)

Table 4.4. Endorsements of the interpersonal quality of care items included in the confirmatory factor analyses by data collection method

A. Observation data

		Total sample (N=291)			Complete case (N=269)	
		N	%	% missing	N	%
Respectful maternity care items:						
RMC1	Respectfully greeted when you arrived at the maternity ward	220	75.9	0.3	205	76.2
RMC2	Asked if you had any questions during stay in maternity ward	39	13.5	0.7	39	14.5
RMC3	Encouraged to walk or assume different positions during labor	226	78.2	0.7	211	78.4
RMC4	Encouraged to consume liquids/food throughout labor	122	42.2	0.7	116	43.1
RMC5	Supported during labor in a friendly way	229	80.1	1.7	218	81.0
RMC6	Encouraged woman to have support person present	23	8.0	0.7	23	8.6
RMC7	Responds to woman's questions and concerns	232	80.6	1.0	217	81.7
RMC8	Informs woman of findings	132	46.5	2.4	126	46.8
RMC9	Explains what will happen during childbirth	83	28.6	0.3	77	28.6
RMC10	Drapes woman or gives woman something to cover her	173	60.5	1.7	162	60.2
RMC11	Privacy assured during examination	271	95.4	2.4	258	95.9
RMC12	Assisted to use toilet facilities	-	-	-	-	-
RMC13	Able to communicate with your relatives or other persons who accompanied you	-	-	-	-	-
RMC14	Received any assistance to reduce pain	-	-	-	-	-
RMC15	Providers came quickly when called for them	-	-	-	-	-
Disrespect and abuse items:						
DA1	Woman's body seen by other people (aside from health provider) during delivery	28	9.7	0.3	26	9.7
DA2	Health providers shouted at or scolded woman	134	46.4	0.7	121	45.0
DA3	Health providers threatened woman for any reason or made negative or disparaging comments about the woman	131	45.0	0	119	44.2
DA4	Health providers ignored or abandoned woman when in need or when she called for help	59	20.3	0.3	50	18.6
DA5	Delivered without any assistance	10	3.5	0.3	10	3.7
DA6	Hit, slapped, pushed, pinched or otherwise beat the woman	39	13.5	0.3	36	13.4
DA7	Health providers suggested or asked for a bribe or informal payment for better care	5	1.7	0.3	5	1.9
DA8	Health providers threatened to withhold treatment because woman could not pay or did not have supplies	17	5.9	0.3	15	5.6

B. Self-report data

		Total sample (N=1,680)			Complete case (N=1,638)	
		N	%	% missing	N	%
Respectful maternity care items:						
RMC1	Respectfully greeted when you arrived at the maternity ward	1084	65.6	1.6	1073	65.5
RMC2	Asked if you had any questions during stay in maternity ward	939	56.1	0.4	918	56.0
RMC3	Encouraged to walk or assume different positions during labor	1127	67.1	0.1	1104	67.4
RMC4	Encouraged to consume liquids/food throughout labor	1176	70.0	0.1	1151	70.3
RMC5	Supported during labor in a friendly way	1302	77.6	0.1	1276	77.9
RMC6	Encouraged woman to have support person present	-	-	-	-	-
RMC7	Responds to woman's questions and concerns	-	-	-	-	-
RMC8	Informs woman of findings	-	-	-	-	-
RMC9	Explains what will happen during childbirth	-	-	-	-	-
RMC10	Drapes woman or gives woman something to cover her	-	-	-	-	-
RMC11	Privacy assured during examination	-	-	-	-	-
RMC12	Assisted to use toilet facilities	442	26.3	0.1	432	26.4
RMC13	Able to communicate with your relatives or other persons who accompanied you	341	20.3	0.1	338	20.6
RMC14	Received any assistance to reduce pain	146	8.7	0.0	143	8.7
RMC15	Providers came quickly when called for them	1440	86.1	0.4	1414	86.3
Disrespect and abuse items:						
DA1	Woman's body seen by other people (aside from health provider) during delivery	27	1.6	0.1	27	1.7
DA2	Health providers shouted at or scolded woman	88	5.2	0.1	88	5.4
DA3	Health providers threatened woman for any reason or made negative or disparaging comments about the woman	51	3.0	0.0	51	3.1
DA4	Health providers ignored or abandoned woman when in need or when she called for help	86	5.1	0.0	84	5.1
DA5	Delivered without any assistance	35	2.1	0.4	35	2.1
DA6	Hit, slapped, pushed, pinched or otherwise beat the woman	14	0.8	0.1	14	0.9
DA7	Health providers suggested or asked for a bribe or informal payment for better care	-	-	-	-	-
DA8	Health providers threatened to withhold treatment because woman could not pay or did not have supplies	-	-	-	-	-

Table 4.5. Factor loadings and fit statistics for confirmatory factor analyses for the one-factor vs. two-factor models for each data collection method fit with overlapping items

		Observation data (N=269)		Self-report data (N=1,638)	
		One-factor model (Model 1)	Two-factor model (Model 2)	One-factor model (Model 3)	Two-factor model (Model 4)
		Loadings[a]	Loadings[a]	Loadings[a]	Loadings[a]
Respectful maternity care items					
RMC1	Respectfully greeted when you arrived at the maternity ward	0.73	0.77	0.31	0.32
RMC2	Asked if you had any questions during stay in maternity ward	0.31[b]	0.37	0.68	0.71
RMC3	Encouraged to walk or assume different positions during labor	0.75	0.79	0.83	0.86
RMC4	Encouraged to consume liquids/food throughout labor	0.30	0.34	0.80	0.83
RMC5	Supported during labor in a friendly way	0.89	0.93	0.76	0.83
Disrespect and abuse items (reverse coded)					
DA1	Woman's body seen by other people (aside from health provider) during delivery	-	-	0.66	0.71
DA2	Health providers shouted at or scolded woman	0.66	0.89	0.93	0.95
DA3	Health providers threatened woman for any reason or made negative or disparaging comments about the woman	0.48	0.61	0.90	0.91
DA4	Health providers ignored or abandoned woman when in need or when she called for help			0.97	0.99
DA5	Delivered without any assistance	0.19[c]	0.27[b]	0.89	0.91
DA6	Hit, slapped, pushed, pinched or otherwise beat the woman	0.55	0.69	0.70	0.75
Correlation between the factors		-	0.50	-	0.75
Fit Statistics		**Value**	**Value**	**Value**	**Value**
Root Mean Square Error of Approximation (RMSEA) (90% CI)		0.09 (0.06, 0.11)	0.03 (0.00, 0.06)	0.07 (0.07, 0.08)	0.06 (0.05, 0.06)
Comparative Fit Index (CFI)		0.87	0.98	0.96	0.97
Tucker Lewis Index (TLI)		0.83	0.98	0.94	0.96
Chi-square difference test (two-factor vs. one-factor model)		$x^2 = 24.42$, df = 1, p<0.0001		$x^2 = 65.33$, df = 1, p<0.0001	

[a]p-value for all items <0.0001 unless otherwise noted; [b]p-value<0.01 [c]p-value<0.05

Table 4.6. Reliability (ordinal alpha) of the latent factors in the two-factor models for each data collection method

	Model 2: Observation two-factor model	Model 4: Self-report two-factor model
Factor 1: respectful maternity care	0.77	0.81
Factor 2: disrespect and abuse	0.70	0.92

Table 4.7. Convergent and divergent validity: distribution of the variables and correlation between the latent factors and validity variables for Model 4: self-report two-factor model

	Total sample (N=1,680)			Complete case (N=1,638)	
	N	%	% missing	N	%
Validity items					
Respect providers showed for delivery					
Excellent	24	1.4	0.1	24	1.5
Very good	269	16.0		254	15.5
Good	1259	75.0		1234	75.4
Fair	115	6.9		113	6.9
Poor	12	0.7		12	0.7
Satisfaction with experience during delivery					
Very satisfied	1435	85.4	0.1	1396	85.3
Somewhat satisfied	214	12.7		210	12.9
Somewhat dissatisfied	23	1.4		23	1.4
Very dissatisfied	7	0.4		7	0.4

	Convergent validity	Divergent validity
	Respect providers showed for delivery	Satisfaction with delivery
	Correlation	Correlation
Factor 1: respectful maternity care	0.48	0.50
Factor 2: disrespect and abuse	0.79	0.79

Table 4.8. Factor loadings and fit statistics for confirmatory factor analyses for the one-factor vs. two-factor models for each data collection method fit with additional items

		Observation (N=269)	Self-report (N=1,638)
		Model 5: Two-factor model with additional items	Model 6: Two-factor model with additional items
		Loadings[a]	Loadings[a]
Respectful maternity care items			
RMC1	Respectfully greeted when you arrived at the maternity ward	0.66	0.35
RMC2	Asked if you had any questions during stay in maternity ward	0.55	0.70
RMC3	Encouraged to walk or assume different positions during labor	0.89	0.84
RMC4	Encouraged to consume liquids/food throughout labor	0.36	0.83
RMC5	Supported during labor in a friendly way	0.85	0.84
RMC6	Encouraged woman to have support person present	0.37	-
RMC7	Responds to woman's questions and concerns	0.75	-
RMC8	Informs woman of findings	0.31	-
RMC9	Explains what will happen during childbirth	0.56	-
RMC10	Drapes woman or gives woman something to cover her	0.63	-
RMC11	Privacy assured during examination	0.31[b]	
RMC12	Assisted to use toilet facilities	-	0.57
RMC13	Able to communicate with your relatives or other persons who accompanied you	-	0.27
RMC14	Received any assistance to reduce pain	-	0.45
Disrespect and abuse items (reverse coded)			
DA1	Woman's body seen by other people (aside from health provider) during delivery	-	0.72
DA2	Health providers shouted at or scolded woman	0.90	0.95
DA3	Health providers threatened woman for any reason or made negative or disparaging comments about the woman	0.57	0.91
DA4	Health providers ignored or abandoned woman when in need or when she called for help		0.99
DA5	Delivered without any assistance	0.26[b]	0.91
DA6	Hit, slapped, pushed, pinched or otherwise beat the woman	0.60	0.74
DA7	Health providers suggested or asked for a bribe or informal payment for better care	0.76	-
DA8	Health providers threatened to withhold treatment because woman could not pay or did not have supplies	0.51	-
Correlation between the factors		0.45	0.74
Fit Statistics		**Value**	**Value**
Root Mean Square Error of Approximation (RMSEA) (90% CI)		0.07 (0.06, 0.08)	0.06 (0.05, 0.06)
Comparative Fit Index (CFI)		0.80	0.96
Tucker Lewis Index (TLI)		0.77	0.95

[a]p-value for all items <0.0001 unless otherwise noted; [b]p-value<0.05

4.8. Appendices

Appendix 4.1. Rationale for using observation data from Dataset 1 only for analysis

I ran two separate CFAs by time period for the observation to determine if there were any time effects. There were differences between the two. Therefore, I ran a combined CFA controlling for time period. The combined model had poor fit (RMSEA: 0.07, CFI: 0.88, TFI: 0.83). I explored the distribution of the items by time period and further stratified by facility. I noticed significant differences in the endorsement of the respectful maternity care items at time 2 between the facilities which were in the opposite of the hypothesized direction. This caused me to question the quality of the data. To eliminate time effects and any doubts about data quality, I decided to use the observation data from time 1 only.

Appendix 4.2. Power calculations for confirmatory factor analyses for each data collection method

Observation data (N=269)			Self-report data (N=1,638)		
Items	**95% coverage***	**Power**		**95% coverage***	**Power**
One-factor			**One-factor**		
X1	0.930	0.801	X1	0.949	1.000
X2	0.918	0.800	X2	0.949	1.000
X3	0.928	0.820	X3	0.949	1.000
X4	0.929	0.811	X4	0.939	1.000
X5	0.934	0.801	X5	0.951	1.000
X6	0.940	0.816	X6	0.943	1.000
X7	0.926	0.810	X7	0.952	1.000
X8	0.936	0.814	X8	0.932	1.000
X9	0.920	0.808	X9	0.958	1.000
X10	0.924	0.826	X10	0.941	1.000
X11	0.911	0.799	X11	0.938	1.000
Two-factor			**Two-factor**		
Factor 1:			Factor 1:		
X1	0.935	0.896	X1	0.945	0.995
X2	0.923	0.873	X2	0.944	0.990
X3	0.936	0.890	X3	0.940	0.990
X4	0.926	0.891	X4	0.938	0.996
X5	0.926	0.893	X5	0.944	0.991
Factor 2:			Factor 2:		
X6	0.928	0.922	X6	0.933	0.999
X7	0.921	0.923	X7	0.949	0.997
X8	0.918	0.939	X8	0.948	0.995
X9	0.923	0.921	X9	0.936	1.000
X10	0.935	0.944	X10	0.938	1.000
X11	0.919	0.920	X11	0.946	0.998
Factor 1 with Factor 2	0.899	0.796	Factor 1 with Factor 2	0.922	0.877

*The proportion of replications for which the 95% confidence interval contains the true population value

Appendix 4.3. Factor loadings and fit statistics for confirmatory factor analyses for the observation two-factor model: overlapping items and elimination of DA1 (N=269)

		All overlapping items		Eliminated DA1	
		One factor model	Two-factor model	One factor model	Two-factor model
		Loadings[a]	Loadings[a]	Loadings[a]	Loadings[a]
Respectful maternity care items					
RMC1	Respectfully greeted when you arrived at the maternity ward	0.73	0.77	0.73	0.77
RMC2	Asked if you had any questions during stay in maternity ward	0.30[d]	0.37	0.31[c]	0.37
RMC3	Encouraged to walk or assume different positions during labor	0.74	0.79	0.74	0.80
RMC4	Encouraged to consume liquids/food throughout labor	0.31	0.35	0.30	0.34
RMC5	Supported during labor in a friendly way	0.88	0.92	0.88	0.92
Disrespect and abuse items (reverse coded)					
DA1	Woman's body seen by other people (aside from health provider) during delivery	0.13[e]	0.16[e]		
DA2	Health providers shouted at or scolded woman	0.65	0.87	0.65	0.86
DA3	Health providers threatened woman for any reason or made negative or disparaging comments about the woman	0.47	0.58	0.47	0.59
DA4	Health providers ignored or abandoned woman when in need or when she called for help	0.29[d]	0.44	0.29[c]	0.44
DA5	Delivered without any assistance	0.36[d]	0.48	0.36[c]	0.47
DA6	Hit, slapped, pushed, pinched or otherwise beat the woman	0.53	0.64	0.53	0.64
Correlation between the factors			0.49		0.49
Fit Statistics		**Values**	**Values**	**Values**	**Values**
Root Mean Square Error of Approximation (RMSEA) (90% CI)		0.09 (0.07, 0.11)	0.07 (0.05, 0.09)	0.10 (0.08, 0.12)	0.08 (0.09, 0.10)
Comparative Fit Index (CFI)		0.76	0.87	0.77	0.87
Tucker Lewis Index (TLI)		0.71	0.84	0.70	0.83

[a]p-value for all items <0.0001 unless otherwise noted; [b]p-value<0.001; [c]p-value <0.01; [d]p-value <0.05; [e]p-value ≥0.05 (not significant)

Chapter 5: Discussion

5.1. Overview

The goal of this dissertation was to bring an epidemiologic lens to an implementation problem of limited maternal health progress despite interventions to increase in facility-based deliveries. It used a quantitative toolkit to assess the evidence base for interpersonal quality of care during childbirth at health facilities, which is required for the development and scale up of interventions to address mistreatment during childbirth. First, it tested a prevailing quality of care framework by examining whether structural inputs, at both proximal (maternal health) and distal (HIV) levels, affect interpersonal quality of care during labor and delivery. Second, it reviewed the state of the literature on the construct of interpersonal quality of care. Third, it constructed a measure of interpersonal quality of care that appropriately reflected the full construct of both positive and negative aspects of care in order to provide a more complete picture of how women are treated during childbirth.

5.2. Summary of results

In Chapter 2, I used data from health facilities in Malawi to test the underlying assumptions of quality of care frameworks linking HIV and maternal health structural inputs and interpersonal quality of care during childbirth. As a secondary goal, I tested whether maternal health structural inputs were a mediator between HIV structural inputs and interpersonal quality of care. I did not find meaningful effects of structural inputs on interpersonal quality of care and found no evidence of mediation. The results do not support the quality of care frameworks or qualitative evidence which suggest a relationship between structure and interpersonal processes of care.

In the systematic review in Chapter 3, I assessed the current state of the literature for instruments measuring interpersonal quality of care to identify high-quality, reliable, and valid instruments of relevance to the maternal health context. Given the lack of consensus for the definition of interpersonal quality of care, it was not surprising to find variation in how the construct was defined and labeled across the measurement instruments and inconsistency in the factor structure. Overall, few instruments had strong reliability and validity, and, of those that did, the studies that assessed the instruments were generally of poor quality. The review also challenged my confidence in exploratory factor analytic methods, which, as data-driven approaches, are often void of any theory to guide the analysis. In order to advance the interpersonal quality of care agenda and to ensure that measures appropriately reflect patient-provider interactions, there is a need for a unified definition of the construct and validations that are of strong methodological quality and that preferably use theory-driven approaches.

In Chapter 4, I determined whether respectful maternity care, including support during labor and effective communication, implies the absence of disrespect and abuse, such as hitting, verbal abuse, and physical privacy violations. To test this hypothesis, I performed confirmatory factor analyses with data from two different data collection methods, observation of laboring women and women's self-report of their labor and delivery experience at health facilities in Tanzania. While the models using observation and self-report data performed differently, both demonstrated a two-factor structure, supporting the hypothesis that respectful maternity care and disrespect and abuse are separate, yet correlated, dimensions of interpersonal quality of care. These findings indicated that respectful maternity care and disrespect and abuse may have

different effects, which could affect the success of interventions targeted only one aspect of interpersonal quality of care.

5.3. Strengths and limitations

The dissertation has several important strengths. The analysis presented in Chapter 2 is, to my knowledge, the first to quantitatively test the links between structural inputs and interpersonal quality of care during childbirth that are proposed by quality of care frameworks. Additionally, it used a robust measure for HIV spillover effects, operationalizing the spillover as structural inputs, rather than presence of programs or funding as in prior research, to provide a closer approximation of the effect. The systematic review in Chapter 3 assesses instruments measuring interpersonal quality of care in a variety of health care settings, as opposed to specific health contexts. In Chapter 4, the confirmatory factor analyses used two different data collection methods and a theory-driven, rather than data-driven, approach to test the factor structure of interpersonal quality of care.

Despite these strengths, the dissertation has limitations as well. First, both Chapter 2 and Chapter 4 used observations of labor and delivery to measure the interpersonal quality of care that women experienced during childbirth, which provide an outside, potentially objective, perspective of the care experience, rather than a subjective one from the woman herself. Using women's own perspectives in the analysis in Chapter 2 may have resulted in different findings. As evidenced by the factor analysis results in Chapter 4, the models performed differently when using data measured via observation compared with self-report. When deciding between observation and self-report measures, future research needs to consider the research goals and the

implications of each method. Second, in Chapter 2, the majority of the items in the measure of interpersonal quality of care were positive aspects of care. In light of the findings from Chapter 4 that the measure is two-dimensional, the results of Chapter 2 should be validated with a fuller measure of interpersonal quality of care which captures both positive and negative aspects of care. Third, I focused on the inputs of the quality of care frameworks, rather than health outcomes. However, I tested parts of the frameworks that are necessary for the outcomes to occur. A natural next step would be to extend this work to maternal or newborn health outcomes.

5.4. Lessons learned and implications for future research

This dissertation resulted in several lessons learned that have practical implications for interventions to address interpersonal quality of care in the maternal health field and the future study of the issue.

First, the findings from Chapter 2 did not support the assumptions of the quality of care frameworks that link structure and interpersonal process. If these results accurately reflect the relationship between structure and interpersonal process, the emphasis on structure as a cause of interpersonal quality of care, and interventions that specifically target structure with the intention of improving interpersonal quality of care, may be misguided. While there are still glaring deficits in structural quality in low-resource settings,[1] which should not be ignored, and structural inputs are necessary for health system performance, these results indicate that there may be an overemphasis on structural inputs to improve interpersonal quality. Indeed, poor treatment also exists in high-resource settings,[2,3] where the structural deficits that are present in low-resource settings are uncommon, signaling that there are other factors at play. However, the

current findings require validation, particularly in consideration of the limitations of the data, as outlined in Chapter 2 and above. Other potential drivers of interpersonal quality also warrant exploration, such as power dynamics, organizational culture, and lack of accountability. The field is poised to repeat the implementation failures of the Millennium Development Goal era if it intervenes on interpersonal quality of care based on unverified assumptions of quality of care frameworks and without thorough consideration of its causes.

To successfully guide intervention, we also need a measure of the full construct of interpersonal quality of care that is validated and context-specific. The findings from the confirmatory factor analyses in Chapter 4 illustrate the complexity of the childbirth experience and suggest that it may not be singularly defined as positive or negative. It is possible that a woman may both be supported and physically abused during labor, and that each event may contribute differently to her overall perception of the birth experience, future utilization preferences, and confidence in the health system. An intervention to improve respectful maternity care may not necessarily guarantee that a woman's childbirth experience is free from abuse. As evidenced from the systematic review in Chapter 3, other instruments validated in the maternal health context also suggest separate factors for positive and negative items, but the factor structures of these instruments all differed, ranging from two to four factors.[4-6] Moving forward, as the field explores how best to measure the construct and to intervene, the dimensionality of interpersonal quality of care during childbirth requires further examination.

Finally, in addition to the conclusions from the results of this dissertation, it is necessary to consider interpersonal quality of care in relation to other strategies to reduce maternal morbidity

and mortality. In recent years, the global maternal health community has galvanized around interpersonal quality of care as a priority, but the extent of its effects on maternal morbidity and mortality is not fully known. Poor interpersonal treatment can deter women from going to facilities for delivery and can influence future utilization, satisfaction with health services, and quality ratings.[7-9] Thus, improving interpersonal quality of care has the potential to increase health facility utilization for childbirth, which can contribute to maternal mortality reductions. Once at the facility, interpersonal quality of care can have physiological and emotional effects, including shorter labors and reduced likelihood of having a Caesarean section,[10] but this will likely have limited impact on maternal mortality. As deliveries continue to shift from home to health facilities, improvements in technical quality, as studies have suggested, are also paramount to ensure that facilities are staffed with personnel that have the knowledge, skills, and capacity to provide lifesaving clinical interventions.[11-13]

In sum, this dissertation suggests that the research agenda for interpersonal quality of care during childbirth in low-resource settings requires reassessment. There is a need to address interpersonal quality of care during childbirth. However, to avoid implementation failure and to instigate change, it is critical to ensure that interventions are based on robust evidence.

5.4. References

1. Kruk ME, Leslie HH, Verguet S, Mbaruku GM, Adanu RMK, Langer A. Quality of basic maternal care functions in health facilities of five African countries: an analysis of national health system surveys. *Lancet Glob Health.* 2016;4:e845-e855.

2. Jomeen J, Redshaw M, J. J. Ethnic minority women's experience of maternity services in England. *Ethn Health.* 2013;18:280-296.

3. Vedam S, Stoll K, Rubashkin N, et al. The mothers on respect (MOR) index: Measuring quality, safety, and human rights in childbirth. *SSM - Population Health.* 2017.

4. Afulani PA, Diamond-Smith N, Golub G, Sudhinaraset M. Development of a tool to measure person-centered maternity care in developing settings: validation in a rural and urban Kenyan population. *Reprod Health.* 2017;14:118.

5. Redshaw M, Martin CR. Validation of a perceptions of care adjective checklist. *J Eval Clin Pract.* 2009;15:281-288.

6. Sheferaw ED, Mengesha TZ, Wase SB. Development of a tool to measure women's perception of respectful maternity care in public health facilities. *BMC Pregnancy Childbirth.* 2016;16:67.

7. Bohren MA, Hunter EC, Munthe-Kaas HM, Souza JP, Vogel JP, Gülmezoglu AM. Facilitators and barriers to facility-based delivery in low- and middle-income countries: a qualitative evidence synthesis. *Reprod Health.* 2014;11:71.

8. Kruk ME, Paczkowski M, Mbaruku G, de Pinho H, Galea S. Women's preferences for place of delivery in rural Tanzania: a population-based discrete choice experiment. *Am J Public Health.* 2009;99(9):1666-1672.

9. Kujawski S, Mbaruku G, Freedman LP, Ramsey K, Moyo W, Kruk ME. Association Between Disrespect and Abuse During Childbirth and Women's Confidence in Health Facilities in Tanzania. *Matern Child Health J.* 2015;19(10):2243-2250.

10. Hodnett ED, Gates S, Hofmeyr GJ, Sakala C. Continuous support for women during childbirth. *Cochrane Database Syst Rev.* 2013;32:72.

11. Campbell OMR, Calvert C, Testa A, et al. The scale, scope, coverage, and capability of childbirth care. *Lancet.* 2016;388:2193-2208.

12. Souza JP, Gulmezoglu AM, Vogel J, et al. Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): A cross-sectional study. *Lancet.* 2013;381:1747-1755.

13. Koblinsky M, Moyer CA, Calvert C, et al. Quality maternity care for every woman, everywhere: a call to action. *Lancet.* 2016;388:2307-2320.