Health System Quality Improvement Strategies in Sub-Saharan Africa: Implementation and Impact

HARVARD UNIVERSITY

Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Committee on Higher Degrees in Population Health Sciences,
have examined a dissertation entitled

"Health System Quality Improvement Interventions in Sub-Saharan Africa: Implementation and Impact"

presented by

ANNA GAGE

candidate for the degree of Doctor of Philosophy
and hereby certify that it is worthy of acceptance.

M Kruk

Dr. Margaret E. Kruk, M.D., Committee Chair, Harvard T.H. Chan School of Public Health

Margaret McConnell

Dr. Margaret Anne McConnell, Ph.D., Harvard T.H. Chan School of Public Health

Sebastian Bauhoff

Dr. Sebastian Bauhoff, Ph.D., Harvard T.H. Chan School of Public Health

Date: 14 April 2022

Health System Quality Improvement Strategies in Sub-Saharan Africa: Implementation and Impact

Anna Gage

A dissertation submitted to the faculty of

The Harvard T. H. Chan School of Public Health

in partial fulfillment of the requirements

for the degree of *Doctor of Philosophy*

in Population Health Sciences

in the Department of Global Health and Population

Harvard University

Cambridge, Massachusetts

April 2022

Dissertation Advisor: Margaret E. Kruk Anna Gage

Health system quality improvement strategies in sub-Saharan Africa: Implementation and impact

Abstract

Poor quality health systems are increasingly recognized as a major barrier to achieving universal health coverage and improved health outcomes in Sub-Saharan Africa. As the extent and degree of poor-quality care has been documented in recent years, improving health system quality is a growing priority. Quality improvement interventions may act in non-linear ways because they are implemented within the context of complex adaptive health systems, where interconnected components allow for feedback loops, learning, and adaptation. The three papers comprising this dissertation explore healthcare quality improvement in Sub-Saharan Africa through a complex adaptive system lens, examining interventions at the macro, meso and micro levels of the health system.

Maternal and newborn care has been a primary focus of performance-based financing (PBF) projects across Sub-Saharan Africa, however there is a lack of evidence of the effect of PBF on neonatal health outcomes. Chapter 2 uses a difference-in-differences study design with secondary data to assess the impact of PBF on early neonatal health outcomes and associated health care utilization and quality in Burundi, Lesotho, Senegal, Zambia, and Zimbabwe. PBF had no detectable impacts on neonatal mortality or low birthweight and had limited and variable effects on the utilization and quality of neonatal health care. This study highlights the necessity of assessing health impacts directly and suggests other strategies will be necessary to improve newborn health outcomes.

Chapter 3 explores a national primary care quality improvement intervention in Tanzania, the Star Rating Assessment, in which primary care facilities received data, feedback and guidance to improve their quality of care. Across two rounds of data collection, there was varied levels of improvement across facilities. This study identifies contextual factors associated with facility quality improvement, finding that improvement was associated with community demand, external policies, and baseline quality levels. Geographic clustering in improvement was not completely explained by administrative boundaries, suggesting that nearby facilities may also play a role in spurring improvement. The results highlight that the facility's setting can promote or inhibit quality improvement as much as internal facility management and organization.

In Ethiopia, the quality of routine maternal and neonatal care needs to be improved to address lingering mortality and morbidity. Multiple providers often attend a single delivery over the course of labor, intrapartum and postpartum periods, particularly in larger health facilities, with unknown consequences for the quality of care. Chapter 4 explores how multiple providers work together to provide quality care using detailed observations of deliveries collected in Dire Dawa Administration, Ethiopia. The number of providers attending a delivery was unassociated with quality of care but working with coworkers who provide higher quality of care was modestly associated with better adherence to routine care guidelines. This study suggests that quality improvement interventions should take account of team structures and leverage provider relationships to create positive spillovers for quality of care.

Together, these three studies show promising opportunities for addressing the enormous gaps in health system quality in sub-Saharan Africa. Findings can be used to harness the feedback loops and dynamic relationships inherent in health systems to magnify the potential impact of quality improvement and to harmonize improvement at the macro, meso and micro levels for better population health.

Table of Contents

List of tables

List of figures

Acronyms

AMTSL: Active management of the third stage of labor

ANC: Antenatal care

ART: Antiretroviral therapy

CEM: Coarsened exact matching

CEmONC: Comprehensive emergency obstetric and newborn care

CFIR: Consolidated Framework for Implementation Research

CHMT: Council health management team

CI: Confidence interval

DHS: Demographic and Health Survey

HIV: Human immunodeficiency virus

HMIS: Health management information system

HRITF: Health Results Innovation Trust Fund

IESO: Integrated emergency surgical officer

MICS: Multiple Indicator Cluster Survey

NICU: Neonatal intensive care unit

PBF: Performance based financing

QoPIIPC: Quality of processes of intrapartum and immediate postpartum care index

RBF: Results based financing

SPA: Service Provision Assessment

WHO: World Health Organization

Acknowledgements

This dissertation would not have been possible without the many contributions of many people around the globe. First, I would like to thank and acknowledge the many women and health providers in Dire Dawa, Ethiopia, across Tanzania, and in Burundi, Lesotho, Senegal, Zambia, and Zimbabwe. Their lived experiences made this research possible, and I am grateful for their time and willingness to share their lives for the hope of improving their health systems.

I am enormously indebted to the guidance and wisdom of my dissertation advisory committee, Drs. Margaret Kruk, Sebastian Bauhoff and Maggie McConnell. They each helped inspire questions, sharpen my thinking, and helped me get unstuck when needed. In particular, I would like to thank Margaret for her mentorship not just over this dissertation, but over the past seven years. Beyond the research guidance, technical support, and life advice, Margaret helped me discover my love of research, for which I am eternally grateful.

I have received guidance and support from so many others in the Harvard community. I thank Drs. Kevin Croke and Jessica Cohen for their guidance and mentorship outside of my dissertation. I am so grateful to so many members of the Kruk research team both past and present. Thank you to Drs. Hannah Leslie, Elysia Larson, Sanam Roder-DeWan, Catherine Arsenault, Kojo Nimako and Todd Lewis for enriching my work in so many ways and being the 'sticky' colleagues that continue to support me long after their formal roles end. From unexpectedly hiking Table Mountain to working on jigsaw puzzles in Dar es Salaam, you have made the work fun and I am so lucky to be able to call so many colleagues friends. I am also grateful to so many of my GHP PhD classmates and friends who have worked with me in the cubes, gave feedback on so many presentations, helped me work through problems and had fun with outside of school. Thank you in particular to Drs. Mia Blakstad, Sian Tsuei and Dina Goodman-Palmer for being with me from beginning to end. Thank you also to Barbara Heil and Allison Conary for their support.

I am so appreciative to my colleagues and coauthors in Ethiopia and Tanzania who made this research possible. A special thank you to Dr. Bereket Yakob, who amid a Chikungunya outbreak, locust swarms, Ethiopian elections, and the Covid-19 pandemic still found a window to lead the data collection effort in Dire Dawa. His determination, persistence and attention to detail gave me the confidence in the data while I unable to travel. Thank you as well to Dr. Tsinuel Girma, who has been such a wonderful mentor, and Dr. Talhiya Yahya, whose openness to learn from the Star Rating data has been so inspiring.

I am so grateful to the dear friends with whom I first found a community at Harvard, Stacy Griswold, Lauren Schaeffer, Bill Daengal, and Sujay and Yuga Kakarmath. Thank you for your love and encouragement through all the ups and downs of the past eight years. Thank you to my wonderful family for their support, including my parents, Jim and Gwen Gage, my sister and brother-in-law, Meredith Gage and Tim Murphy, and my favorite nephew and niece, Desi and Sloane. Thank you for the trips to the mountain house, the board games, all the good food and unwavering support. Finally, thank you to Colin Clapham. Thank you for turning the lights on when I was working in the dark, listening to infinite versions of presentations, cooking me dinner when I came home late, and for supporting me at my best and my worst. I couldn't have done this without you.

Chapter 1 : Introduction

Motivation

Throughout Sub-Saharan Africa, more people today have access to healthcare than ever before [1]. Countries have built record numbers of healthcare facilities [2], the costs to patients of life-saving drugs such as ART have fallen [3], and health insurance coverage is growing [4]. As a result of these successes, the percent of people who receive the care they need care is rising: for example, in 2005 only 5% of women in Ethiopia delivered their baby in a healthcare facility, while in 2019 this share rose to 48% [5]. Over ten years, access to malaria diagnostics nearly doubled in 26 low- and middle-income countries including those in Sub-Saharan Africa [6], and 73% of patients diagnosed with hypertension have received advice or medication [7]. Mortality and morbidity have correspondingly declined in many countries, especially for diseases such as HIV where treatment has expanded greatly [8,9]. However, progress on health outcomes will need to rapidly accelerate in order to meet the Sustainable Development Goals, particularly for neonatal and maternal mortality [6]. Furthermore, approximately 2 million excess deaths in Sub-Saharan Africa in 2016 were still amenable to health system intervention [10], and confidence and satisfaction with the health systems are low [11].

In 2018, three landmark reports identified poor quality of care as one of the key outstanding challenges facing health systems around the world, especially in Sub-Saharan Africa [12–14]. The quality gaps are numerous and pervasive: for example, health care providers in low- and middle-income countries often do less than half of recommended evidenced based action, and approximately one third of patients experience disrespectful care, short consultations, poor communication or long wait times [12]. The underlying causes of poor quality are also numerous. Following the colonial era in Sub-Saharan Africa, many countries were left with antiquated health systems that were designed to care for a small settler population in urban areas and a lack of health worker educational institutions [15]. Continued economic extraction, conflicts and economic instability compounded resource challenges in many countries, resulting in inadequate health sector investments. While health systems in many countries have made enormous progress despite these challenges, poor governance, inappropriate service delivery, inadequate health worker education, and an untapped population demand for quality persist [12].

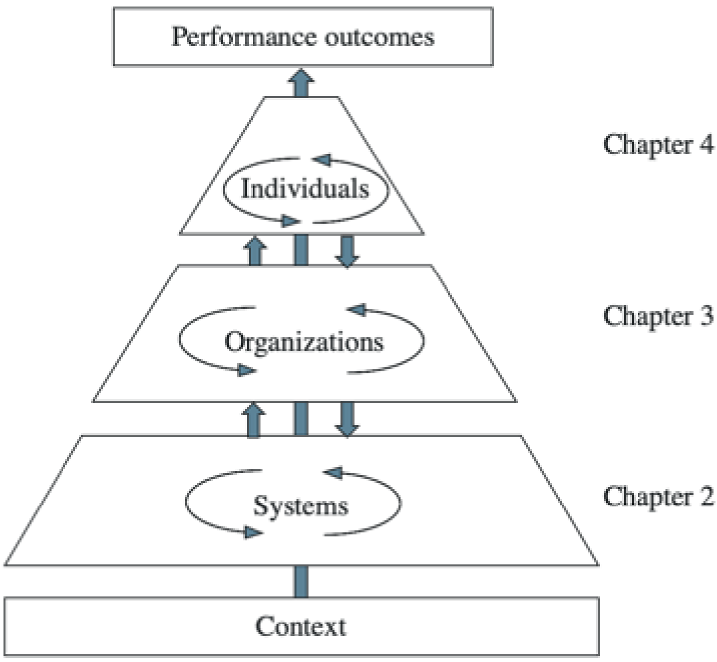
Since health systems have been diagnosed with poor quality, the global health community has increasingly turned to quality improvement as the treatment [16]. Systematic reviews on quality improvement have identified several key insights and gaps in the literature. First, quality improvement implementation and research tend to focus on micro-level approaches, or those focused on individual providers or facilities; however such approaches are often limited in their improvement and sustainability [12]. For example, in-service training and supportive supervision are common for health workers in Sub-Saharan Africa though these approaches are insufficient for closing the identified gaps [17]. Group problem solving approaches tend to have larger effects though the quality of evidence is often poor [18,19]. Conversely, less quality improvement research is focused on meso-level approaches, which focus on groups of facilities or districts, or macro-level approaches which address the foundations of the health system [20].

Second, quality improvement research often assesses the impact on process indicators such as the percent of actions completed but does not evaluate the impact of the improvement on health outcomes [19,21]. Consistent with the tendency toward micro-approaches, quality improvement research is often small-scale and short-term which does not lend itself to assessing eventual population health impact [22]. However, studies that examine interventions' impact on health outcomes have influenced policy discourse. For example, after an influential study of a community-accountability intervention found a substantial reduction in under-5 mortality in Uganda, there was an increased interest in this model [23]. Conversely, a large trial of WHO Better Birth checklist utilization in Uttar Pradesh, India found that while quality of maternal and immediate newborn care improved modestly, there was no impact on a composite health outcome [24]. This study prompted a closer examination of the foundations of maternal and newborn care and a shift to macro approaches, such as redesigning service delivery [25]. Studying eventual health outcomes is critical because health systems act as complex adaptive systems, such that there is not necessarily a linear path from intermediate outcomes to eventual population health [26,27].

Finally, there is a small but growing trend in quality improvement research to use implementation science approaches to understand not just whether quality improved but also how, and how to scale interventions [28]. Implementation science aims to close the 'know-do' gap between efficacious policies and implementation and understand the adoption, reach, and fidelity of interventions [29]. For example, implementation science studies on maternal care quality improvement have identified key enablers and inhibitors of quality improvement uptake, barriers to implementation and challenges to institutionalization [30]. However, reviews have found that implementation research often lacks adequate reporting or neglects the context of the intervention, which prevent successful translation of the findings into practice [28,31].

This dissertation addresses several of these gaps in the literature by examining health system quality and quality improvement within a complex adaptive systems framework. Systems thinking emphasizes shifting away from considering individual components of the health system to understanding the multifaceted and interconnected relationships between them [26]. In addition to these relationships, complex adaptive systems are furthermore able to self-organize, adapt or learn from experience [27]. Interconnected components of the health system and the external environment in which they are situated create feedback loops, spillovers, time-lags and non-linear relationships between quality improvement and outcomes [32]. Adapting a conceptual framework on performance management in complex adaptive systems [33], Figure 1.1 shows the dynamic relationships within and between levels of the health system that this dissertation will address.

Figure 1.1 Dissertation conceptual framework



Overview

In the three chapters that follow, I examine the impact, implementation, and mechanisms of existing or potential health care quality improvement interventions in Sub-Saharan Africa. Individually, each paper uses a complex adaptive systems lens to consider the relationships at play, for example between utilization and quality, between facilities and districts, and between providers caring for the same patient. Together, they span the macro, meso and micro levels of the health system, and thus in combination can further yield useful insights in how these levels interact with one another to produce high quality care and affect health outcomes. The interventions focus on maternal and neonatal health and primary health care, two priority areas of interest for health systems in Sub-Saharan Africa. Utilizing a complex adaptive systems framework, I employ methods from econometrics, implementation science and geographic analyses to inform decisions on health systems quality improvement.

Chapter 2 begins with a macro-level quality improvement intervention, which examines the impact of performance-based financing (PBF) interventions. PBF programs are financing reforms which incentivize performance on pre-defined quantity and quality indicators. Maternal and neonatal care are a key focus area for PBF in sub-Saharan Africa, where PBF projects have been implemented in 21 countries to-date. PBF has been studied extensively [34], with reported impacts on utilization and quality of maternal and neonatal care. However, more research is needed to understand the impact of PBF directly on neonatal health outcomes in sub-Saharan Africa. I sought to address this gap by assessing the impact of PBF programs on neonatal health outcomes in five countries: Burundi, Lesotho, Senegal, Zambia, and Zimbabwe. In addition, I also examined the mechanisms through which outcomes may be improved by evaluating the impact on maternal and neonatal care utilization and quality.

To conduct this analysis, I used secondary data from the Demographic and Health Surveys and Multiple Indicator Cluster Surveys that were conducted before and after PBF programs were implemented in each country. Leveraging the district level roll-out of PBF, I used a difference-in-differences approach that compared newborn mortality and low birthweight in districts that were and were not implementing PBF before and after roll-out. I further used coarsened-exact matching to ensure comparability between populations in implementing and non-implementing districts. I examined the impact pooled across all five countries, in each country individually, and among poor and high-risk mothers in order to understand whether there were potentially differential effects. Results from this study can inform policymakers about the potential benefits and limitations of PBF on neonatal health.

Chapter 3 shifts to examining a meso-level quality improvement intervention, Tanzania's Star Rating Assessment, using an implementation science approach. The Star Rating Assessment aims to improve the quality of primary care nationally through a data feedback approach, by which facilities are assessed on the quality of their care, rated on Star Rating scale, and then encouraged to develop strategies to address the gaps identified [35]. Over two rounds of Star Rating, some facilities improved their rating dramatically, while others failed to improve. This study seeks to identify the determinants of improvement both within the facility and in the surrounding environment.

In the analysis, I combined administrative data from the Star Rating Assessment with a variety of secondary data that measure aspects of the facilities, the populations they serve, and their surroundings. I first identified characteristics that are associated with facility improvement over the two rounds, then examined the contributions of administrative management and facility proximity to the variance in improvement. This study expands our understanding of how facility context shapes quality improvement, which can contribute to more targeted and effective improvement strategies in countries such as Tanzania.

Finally, in Chapter 4, I examined micro level determinants of maternal and newborn care quality in Dire Dawa, Ethiopia. Existing research on the determinants of care quality often assumes that one provider is responsible for all the care provided to a client, however in maternity care multiple providers often attend a single delivery. This study examines whether the single-provider model of analyzing delivery care is appropriate by exploring the dynamics of providers working together and how those dynamics are associated with the quality of care received by the mother and newborn. In this study we did not examine a specific quality improvement intervention but attempt to understand how providers influence one another, which may shed light on future quality improvement efforts.

Primary data was collected for this study though delivery observations in nine health care facilities in Dire Dawa. I first assessed the number of providers attending each delivery and the association with quality of care. Next, I examined how the seniority and performance of a provider's coworkers were associated with care quality, highlighting the role that coworker peers may play in improving quality.

Finally, Chapter 5 concludes by highlighting insights drawn from across these chapters and implications for future practice, policy, and research.

Chapter 2 : The effects of Performance-Based Financing on neonatal health outcomes in Burundi, Lesotho, Senegal, Zambia and Zimbabwe

Authors:

Anna Gage[1], Sebastian Bauhoff[1]

Affiliations:

[1]Harvard T.H. Chan School of Public Health, Boston, United States

Abstract

Maternal and newborn care has been a primary focus of performance-based financing (PBF) projects, which have been piloted or implemented in 21 countries in Sub-Saharan Africa since 2007. Several evaluations of PBF have demonstrated improvements to facility delivery or quality of care. However, no studies have measured the impact of PBF programs directly on neonatal health outcomes in Africa, nor compared PBF programs against another. We assessed the impact of PBF on early neonatal health outcomes and associated health care utilization and quality in Burundi, Lesotho, Senegal, Zambia and Zimbabwe.

We pooled Demographic and Health Surveys and Multiple Indicator Cluster Surveys and applied difference-in-differences analysis to estimate the effect of PBF projects on early neonatal mortality and low birthweight. We also assessed the effect of PBF on intermediate outputs that are frequently explicitly incentivized in PBF projects, including facility delivery and antenatal care utilization and quality, and cesarean section. Finally, we examined the impact among births to poor or high-risk women.

We found no statistically significant impact of PBF on early neonatal mortality, low birthweight, health care utilization or quality in a pooled sample. PBF was also not significantly associated with better health outcomes in each country individually, though in some countries and among poor women PBF improved facility delivery, antenatal care utilization, or antenatal care quality. There was no significant improvement on the health outcomes among poor or high-risk women in the five countries. PBF had no detectable impact on early neonatal health outcomes in the five African countries studied and had limited and variable effects on the utilization and quality of neonatal health care. These findings suggest that there is a need for both

a deeper assessment of PBF and for other strategies to make meaningful improvements to neonatal health outcomes.

Introduction

Despite decades of declining neonatal mortality rates, many countries in Sub-Saharan Africa are still not on track to reach the Sustainable Development Goal of 12 neonatal deaths per 1,000 live births by 2030. Maternal and newborn care services provided in health care facilities are viewed as critical to accelerate progress on neonatal health outcomes [36][37].

Improving the quantity and quality of maternal and newborn care services has been a primary focus of performance-based financing (PBF) projects in Africa in the past two decades, partly driven by support from the Health Results Innovation Trust Fund (HRITF) administered by the World Bank [38,39]. While there are many models of PBF, these projects generally entail a set of financing reforms that explicitly incentivize pre-defined quantity and quality indicators [40]. Through incentives, PBF aims to motivate providers to improve their performance, help attract more capable health workers, or provide additional funding that can support improvements [41]. In addition, it is hypothesized that PBF may enhance problem-solving and accountability separate from the incentive [42]. Commonly incentivized maternal and newborn service indicators include the volume of antenatal care visits and facility deliveries and quality measures, such as all deliveries being conducted by qualified personnel and presence of proper maternity equipment [38]. Health outcomes such as mortality reductions are typically not directly incentivized in PBF programs, but they are hypothesized to improve indirectly through improved availability and quality of service delivery [43].

A robust literature documents the variable impacts of African PBF projects on the quantity and quality of health care services. An influential early evaluation found that Rwanda's PBF raised the number of facility deliveries and the quality of antenatal care, among other intermediate outputs, but did not improve the number of antenatal care visits [44]. Recent reviews similarly found that incentivizing health facilities to provide deliveries can increase their number, but mixed evidence on quality of care and quantity of antenatal care with variation across projects and indicators within projects [39,41,45].

There is currently no empirical evidence on the direct impact of PBF on neonatal health outcomes in African countries. Several studies have modeled health impacts of PBF based on changes to utilization and quality [46,47], but direct evidence is critical for several reasons. First, changes in intermediate outputs may not always translate to better health. For example, increasing facility delivery may not improve neonatal health outcomes in the absence of high-quality care [48,49], and improved adherence to evidence-based checklists during delivery can fail to generate better maternal or newborn health outcomes [50]. Second, the evaluations to date have demonstrated mixed results, with improvements on some indicators, generally including facility delivery, but not on others, including delivery quality [51,52]. It is unclear how these inconsistent improvements may come together to affect health outcomes, and the modeling studies rely on strong assumptions about quality-adjusted coverage measures [46,47]. Third, PBF projects incentivize a particular set of indicators, and it remains largely unclear whether there are negative or positive spillovers. For example, PBF may inadvertently divert resources and attention but could also encourage closely associated beneficial behaviors that are not incentivized [41,53,54]. Finally, PBF projects generally pursue multiple strategies, so that focusing on intermediate outputs may miss other pathways to improved health outcomes. Examining the direct impact on health outcomes captures all pathways and spillovers that are otherwise difficult to model in the context of complex adaptive systems [27]. As improving maternal and child health outcomes, including neonatal health outcomes, is a primary objective of many PBF projects and explicitly noted as the final outcome in the program's conceptual framework, it is important to evaluate these impacts directly [55–57].

In this paper, we empirically evaluate the impact of five PBF projects in Africa on two important neonatal health outcomes, neonatal mortality, and low birthweight, as well as on intermediate outputs through which PBF may improve health outcomes: antenatal care utilization and quality, facility delivery utilization and quality, and cesarean section rates. We conduct both pooled and country-specific analyses, and also assess the impact of PBF for two vulnerable groups, poor women and women with high-risk births.

Our analysis offers three primary contributions. First, we provide direct evidence of the impact of African PBF projects on neonatal mortality, avoiding the challenges faced by modeling studies. Second, we compare the effectiveness of PBF projects in different countries against one another using the same methods and data. Most evaluations focus on just one project and because they use differing methodologies, they are not directly comparable [51,58]. As each project is implemented differently, a direct comparison can help to identify features of the health system context or project that may be more or less effective. Finally, our analysis represents a systematic replication of previous evaluations using alternative data sources [55].

Methods

Data and study countries

Our analysis focused on PBF projects in five African countries: Burundi, Lesotho, Senegal, Zambia and Zimbabwe. Countries were included into the study if they were in Sub-Saharan Africa, had implemented an PBF project supported by the World Bank's HRITF and for which the intervention provinces or districts are known, and had a publicly available nationally representative survey on health care and utilization both prior to and after implementation of the

PBF project. Although Burundi did not have a survey prior to its PBF implementation, we were able to include Burundi by using just the Demographic and Health (DHS) survey after implementation for a longer lookback of births. The DHS collects data on neonatal mortality for all births of the women respondents regardless of when the birth occurred. Burundi is excluded from the pooled analysis in a robustness check. Countries that assigned PBF to specific facilities or sub-districts within districts were further excluded from the study, as in this study the population's treatment status was assigned by their district residence rather than by facility catchment areas.

We used the Demographic and Health Surveys and Multiple Indicator Cluster Surveys (MICS) to assess the impact of PBF. Because there were differing amounts of time between the surveys and the PBF implementation in each country, we limited the analysis to births that occurred within three years before implementation and two years after. We also excluded all births from mothers outside of the defined treatment and control districts. In Zambia and Zimbabwe, data on the household's district was not available directly from the surveys. In these cases, we used the cluster geocodes to place households in districts. Although DHS geocodes are displaced to maintain privacy, the displacement is restricted so that clusters stay within the second administrative level, or the district, in these countries [59].

We assumed that a household was treated if it was located within a PBF implementation district, and therefore that all facilities within implementation districts were treated and that women would have gone to facilities within her district. Table 2.1 summarizes the data sources used for each country.

Table 2.1 PBF characteristics and data sources

	Burundi	Lesotho	Senegal	Zambia	Zimbabwe
First implementation date	December 2006	July 2016	April 2012	April 2012	March 2012
Second implementation date	October 2008	Oct 2016	N/A	N/A	N/A
Additional rollout	Expanded to control regions in April 2010	N/A	Expanded to control regions in May 2016	Expanded to 39 districts in October 2016	Expanded to 44 districts in 2015
Pre-implementation survey	DHS 2010[a]	DHS 2014	DHS 2011	DHS 2007	DHS 2010-2011
Post-implementation survey	DHS 2010 and DHS 2017	MICS 2018	Continuous DHS 2013-2017	DHS 2014 and DHS 2018	DHS 2015
District selection notes				Randomized treatment to districts. Additional unconditional financing arm in 10 districts.	Government selected implementation districts from pair-matched districts.
Major related concurrent interventions		Incentives for district teams for good quality of supervision and support to PBF project	Demand-side vouchers also provided for four ANC visits and skilled delivery	.	Introduced simultaneously with national elimination of user fees for targeted services.
Payment adjustment on other dimensions	Remoteness, poverty, staff and facility needs	Remoteness		Remoteness	Remoteness
Allocation of incentive					
Health facility	70%	50%	25%	40%	75%
Staff incentives	30%	50%	75%	60%	25%

[a]Given the absence of earlier data sources in Burundi, we used the birth recode file from 2010 for the pre-implementation survey through including births that occurred prior to implementation.

Performance-based Financing projects

The PBF projects differed in their design and implementation across the study countries (Table 2.1). In general, the projects were structured to provide healthcare facilities financial incentives conditional on reaching certain performance targets. Maternal and newborn care was a priority for all of the study countries, and targets included both quantity and quality of services. The volume of facility deliveries provided by a skilled birth attendant and antenatal care visits were rewarded in all study countries. Quality measures included structural quality items, such as water and soap available in delivery room (Lesotho), and process quality measures, such as correct use of the partograph (Senegal). The programs all had quantity-based formulas for determining the incentive, which were then inflated (or deflated, in Senegal) based on a quality score. None of the projects directly rewarded improvements on early neonatal death or low birthweight. Further details about the implementation and incentivized measures in each country are provided in Appendix Text A.1. While the program design and implementation was heterogeneous across the study countries, these differences were presumably made to bet suit the existing health system context and limitations, therefore giving the PBF programs the best chance of succeeding in each context.

Four of the five study countries employed purposive selection to select the districts for PBF implementation. For example, in Zimbabwe districts were pair-matched on baseline characteristics such as geographic accessibility and average catchment population and then government officials selected between the two districts for implementation. Implementation was randomized only in Zambia, where districts were also matched prior to randomization. In addition, Zambia also had a third treatment arm which gave facilities unconditional financing equivalent to the amount of the PBF arm. We use the pure control districts without unconditional financing as the controls in the primary analysis but conduct a sensitivity analysis which compares the conditional and unconditional arms in Zambia.

We selected control districts in Zambia and Zimbabwe to match those from the World Bank's impact evaluations [56,57]. Burundi and Senegal both implemented a phased rollout; consequently, we defined the control districts as those that later received PBF in those countries [55,60]. The additional rollouts did not occur within the time period considered in this study. Finally given the small size of Lesotho, we defined the control districts as all the remaining districts that had not received PBF. We excluded Quthing and Leribe districts in Lesotho because they had piloted PBF two years prior to the larger implementation of PBF [61]. Appendix Text A.1 lists of all the implementation and control districts for each country. In a sensitivity analysis, we use all non-implementation districts in all of the countries as controls, only excluding districts that had a prior pilot implementation.

Burundi and Lesotho rolled out the PBF project in two stages within the study period. In the primary analysis, we consider only the first set of implementation districts and the control districts; in a sensitivity analysis we separately compare the second set of implementation districts against the control districts.

Dependent variables

We examined the effect of the PBF projects on two primary neonatal health outcomes: early neonatal death and low birthweight. Early neonatal death, which is associated with facility delivery and quality [48,62], was defined as a death before or including 7 days of birth. Low birthweight, which is associated with ANC quantity and content [63], was defined as a birthweight below or including 2500 grams. If the baby was not weighed at birth, we used multiple imputation with five imputations to impute missing values based on the mother's report of the baby's size at birth and risk factors including multiple births, primipara, urban location, maternal age and primary education, wealth quintile, and district [64]. Although there may be measurement error in the mother's report of the baby's size, this measure is strongly correlated with related health outcomes such as prematurity and intrauterine growth restrictions [48,65]. As robustness checks, we also tested whether PBF impacted the likelihood of birthweight being recorded and the impact of PBF on birthweight among just the subset of observations where birthweight was recorded.

We also examined several pathways through which PBF might affect these health outcomes, including increased utilization or improved quality of antenatal or intrapartum care or increased caesarean sections. We defined antenatal utilization as at least four antenatal care visits and intrapartum utilization as delivering in a health facility. Antenatal and delivery quality were both defined as binary variables, where high quality care recipients received all of the recommended quality items while low quality care recipients received fewer items. Quality measures were alternatively defined as the percent of items received in a robustness check. Antenatal care quality items included the recommended number of Tetanus Toxoid shots, iron supplementation, a blood sample test, and antenatal care from a qualified provider. Iron supplementation was not measured in the 2018 Lesotho MICS, so quality in Lesotho during both waves was measured using the other three items. Delivery quality items included breastfeeding within an hour of delivery, postnatal check before discharge, and delivery with a trained provider. Finally, cesarean section was defined as the mother's report of a cesarean delivery.

The sample for each dependent variable varied based on data availability. Neonatal death data was available for all births, and we imputed birthweights for all births as described above. Antenatal care utilization and quality were only collected for the most recent birth; delivery quality was collected about the most recent birth if the woman had a facility delivery. Facility delivery and caesarean sections were collected about all births.

Analysis

We pooled data from all study countries and used a difference in differences specification to assess the impact of PBF on the study dependent variables:

$$Y_{idt} = \beta_0 + \beta_1(PBF_d * Post_t) + \beta_2\gamma_{idt} + \sum_{j=1}^{60}\beta_j * Month_t + \sum_{k=1}^{75}\beta_k * District_d + \epsilon_{idt}$$

Where Y is a dependent variable for an individual i in district d and month t, PBF is an indicator for whether the district was treated, Post is an indicator for whether the birth was after the date of implementation, γ is a set of covariates, Month is a set of fixed effects of the month of birth in relation to the date of implementation where PBF was implemented in month 37, and $District$ is a set of district fixed effects. As districts are collinear with countries, separate country fixed effects cannot be included in the main analytic model. We separately conducted a robustness check that includes country fixed effects instead of district fixed effects. We used multivariable linear probability models with standard errors clustered by district, as the level of implementation. We similarly tested for parallel pre-trends between implementation and control districts by interacting quarter fixed effects prior to and after the PBF implementation with the binary PBF indicator, excluding the quarter that PBF was implemented. This method can also be used to examine the effect of PBF over time.

Because PBF was not randomized to districts in most countries, we both matched on a set of covariates and controlled for them in our model to obtain a better balance on important characteristics and improve the precision of our estimates [66]. We used coarsened exact matching (CEM) to first match births on the set of covariates. CEM is a method that corrects for imbalances between composition of treatment and control districts by coarsening a set of covariates into bins, creating a stratum per bin and assigning observations to the strata, then dropping any births whose stratum does not contain at least one treated and one control unit [66,67]. We included covariates that are known to be associated with neonatal health outcomes, including multiple birth, primipara, maternal age, year of birth, mother's completion of primary education, urban vs. rural location, and whether the child's household is in the poorest two wealth quintiles in the country (wealth quintiles are constructed and provided by DHS). We included these covariates directly in the model in addition to using the CEM weights in order to further control potential residual confounding and improve precision [67].

We conducted several additional analyses to understand whether the effect differed among sub-populations of interest. First, we conducted the differences in differences model separately in each study country in addition to the pooled analysis. We did not further adjust the standard errors for the small number of clusters in some countries; doing so would result in even more conservative results. Second, we ran the pooled model among the subset of births in households that were in the poorest two wealth quintiles in the country and among the subset of high-risk births. We defined high-risk births as those to primipara women, to women younger than 18 years or older than 34, or multiple births. Because of some debate on how wealth quintiles are comparable within countries between urban and rural areas [68], we also examined the subset of poor and rural women as a robustness check.

Descriptive statistics are presented with the DHS and MICS sampling weights. Analyses were conducted in Stata 15. The original survey implementers obtained ethical approvals for data collection; the authors' institute approved this secondary analysis as exempt from human subject's review.

Results

A total of 30,200 births from DHS or MICS across the five study countries met the inclusion criteria for the study. These included 12,790 births born after the introduction of PBF in their respective countries and 12,700 births that occurred in districts that implemented PBF projects. After coarsened exact matching, 28,619 births were retained in the analysis, removing 1,016 births from control districts and 565 births from PBF districts that were not matched.

Table 2.2 Dependent variables and covariates in control and implementation districts prior to implementation among analytic sample

	Burundi		Lesotho		Senegal		Zambia		Zimbabwe	
	Control	**PBF**	**Control**	**PBF**	**Control**	**PBF**	**Control**	**PBF**	**Control**	**PBF**
Districts	6	3	4	4	4	2	10	10	16	16
Pre-implementation births	3,229	1,557	1,013	980	3,576	2,499	931	990	729	993
Post-implementation births	2,285	1,217	427	418	3,300	2,244	602	611	392	626
Pre-implementation dependent variables										
Early neonatal death	2.3%	3.2%	3.7%	3.2%	1.3%	1.7%	1.5%	1.6%	2.4%	3.0%
Low birthweight	18%	20%	13%	16%	21%	17%	15%	14%	13%	15%
Facility delivery	49%	50%	84%	76%	57%	46%	54%	59%	74%	64%
Delivery quality	86%	86%	56%	58%	55%	53%	71%	72%	54%	58%
C-section	1%	3%	13%	8%	3%	1%	2%	3%	5%	4%
4+ ANC visits	31%	38%	75%	69%	43%	32%	60%	59%	69%	63%
ANC quality	8%	0%	61%	60%	55%	49%	45%	44%	30%	30%
Pre-implementation covariates										
Mother's age at birth (mean)	26.7	27.0	25.2	25.5	26.6	26.2	26.3	26.2	25.9	25.5
Mother primary education	38%	45%	100%	100%	31%	18%	88%	90%	99%	99%
Primipara	20%	23%	44%	39%	20%	19%	20%	17%	27%	29%
Multiple birth	0%	1%	0%	1%	1%	2%	2%	1%	1%	2%
Urban	2%	3%	46%	19%	24%	13%	16%	11%	27%	23%
Poorest wealth quintile	21%	22%	13%	37%	41%	59%	28%	38%	28%	38%
Birthweight recorded	7%	8%	44%	46%	45%	30%	51%	61%	51%	61%

Continued on next page

Table 2.3 (Continued)

	Total	
	Control	**PBF**
Districts	40	35
Pre-implementation births	9,478	7,019
Post-implementation births	7,006	5,116
Pre-implementation dependent variables		
Early neonatal death	2.2%	2.5%
Low birthweight	17%	16%
Facility delivery	61%	57%
Delivery quality	61%	61%
C-section	4%	3%
4+ ANC visits	57%	52%
ANC quality	48%	43%
Pre-implementation covariates		
Mother's age at birth (mean)	26.3	26.2
Mother primary education	57%	62%
Primipara	24%	24%
Multiple birth	1%	1%
Urban	17%	13%
Poorest wealth quintile	25%	37%
Birthweight recorded	32%	39%

Table 2.2 displays the study outcomes and key covariates by treatment district prior to PBF implementation among the matched sample. Across the study countries, 658 (2.3%) births resulted in early neonatal death, ranging from 174 (1.5%) in Senegal to 99 (3.5%) in Lesotho. 4579 (16%) births were low birthweight. Facility delivery and antenatal care utilization rates were low in most countries prior to the intervention; only Lesotho had over 70% facility delivery rate and only 55% of births had four antenatal care visits. Birthweight was recorded on a card for less than half of births at baseline, though many fewer were recorded in Burundi (7%).

Treatment and control districts were not balanced on all covariates prior to PBF implementation even after matching. PBF was implemented more often in poorer districts, particularly in Lesotho, Senegal, and Zimbabwe, and in rural districts. Because most non-implementation districts in Lesotho were considered control districts given the country's size, the control areas are much more urban than the implementation areas. Despite these differing characteristics, the trends in most outcomes do not significantly differ between implementation and control districts prior to implementation (Appendix Figure A.2).

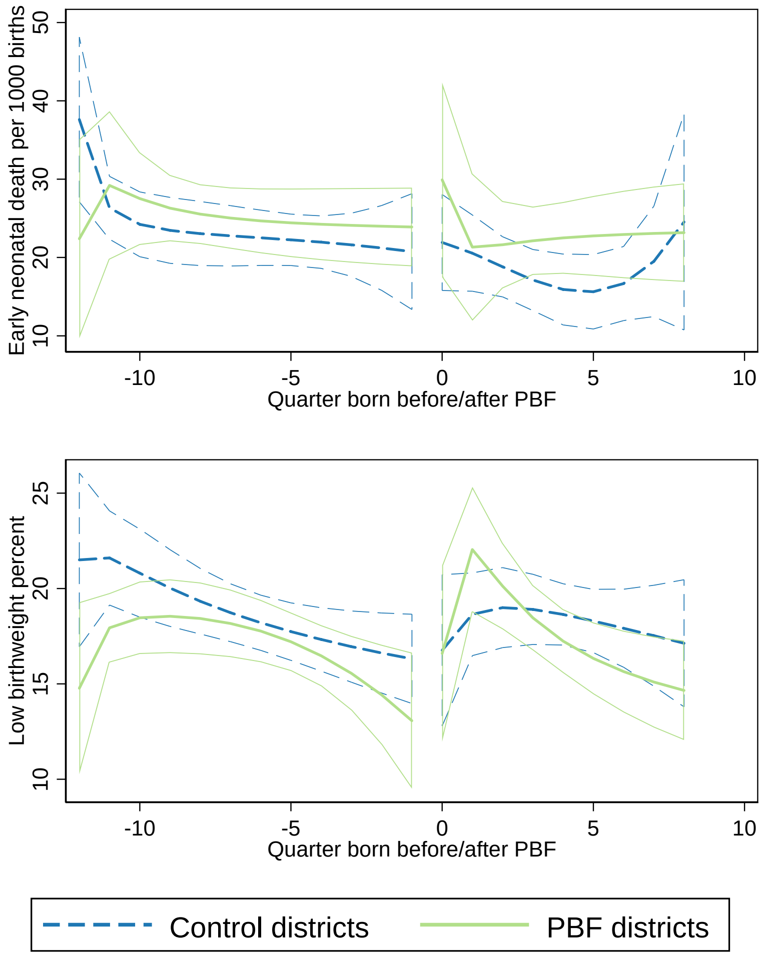
Table 2.3 presents the results from the difference in differences estimation pooling together births from all the study countries and stratified by country. We found no statistically or substantially significant effect of the PBF intervention on any of the health outcomes or intermediate outputs in the pooled analysis. The unadjusted trends for early neonatal death and low birthweight are shown in Figure 2.1, while the results for the intermediate outputs are shown in Appendix Figure A.1. These results were robust to excluding Burundi, to using all non-implementation districts as controls, to using the alternative definitions of the quality measures, to only including observations where birthweight was recorded, to using the second implementation date in Burundi and Lesotho, and to including country fixed effects instead of district fixed effects (Appendix Table A.1). There also do not appear to be delayed effects of PBF within the two-year period assessed (Appendix Figure A.1). The PBF interventions did not have an impact on whether birthweight was recorded (Appendix Table A.1).

Table 2.4 Effects of PBF on primary and secondary outcomes pooled and in all study countries

		Pooled	Burundi	Lesotho	Senegal	Zambia	Zimbabwe
Early neonatal death	Coef.	0.00	0.00	0.02	0.00	-0.01	0.00
	95% CI	(-0.01,0.01)	(-0.01,0.01)	(-0.01,0.05)	(-0.01,0.01)	(-0.04,0.01)	(-0.02,0.03)
	N	28619	8288	2838	11619	3134	2740
Low birthweight	Coef.	0.01	0.01	-0.05	0.03	0.00	-0.02
	95% CI	(-0.02,0.03)	(-0.12,0.13)	(-0.14,0.03)	(-0.01,0.08)	(-0.06,0.06)	(-0.1,0.06)
	N	28619	8288	2838	11619	3134	2740
Facility delivery	Coef.	0.03	**0.08**	0.03	0.03	0.03	-0.02
	95% CI	(-0.01,0.07)	**(0.02,0.14)**	(-0.04,0.09)	(-0.03,0.08)	(-0.06,0.12)	(-0.1,0.06)
	N	21471	**2140**	1849	11619	3123	2740
Delivery quality	Coef.	-0.05	-0.05	-0.09	-0.05	-0.05	-0.03
	95% CI	(-0.14,0.04)	(-0.16,0.06)	(-0.24,0.06)	(-0.11,0.01)	(-0.16,0.06)	(-0.14,0.09)
	N	13054	1219	1558	6275	2026	1976
C-section	Coef.	0.00	0.01	-0.01	0.00	0.00	-0.02
	95% CI	(-0.01,0.01)	(-0.03,0.05)	(-0.11,0.09)	(-0.01,0.01)	(-0.03,0.04)	(-0.05,0.01)
	N	21424	2145	1849	11564	3128	2738
ANC 4 visits	Coef.	0.04	-0.06	**0.12**	0.02	0.06	0.00
	95% CI	(-0.02,0.10)	(-0.18,0.07)	**(0.01,0.22)**	(-0.1,0.15)	(-0.02,0.13)	(-0.12,0.13)
	N	14383	793	**1840**	7383	2157	2210
ANC quality	Coef.	0.02	0.09	-0.03	0.03	**0.09**	-0.03
	95% CI	(-0.04,0.09)	(-0.05,0.24)	(-0.14,0.08)	(-0.1,0.16)	**(0.01,0.17)**	(-0.16,0.09)
	N	14510	796	1869	7445	**2172**	2228

Note: Bolded estimates signify confidence intervals that do not contain zero. Estimated coefficients for from multivariable difference-in-difference regressions representing the percentage point change in the outcome, with standard errors clustered at the district level.

Figure 2.1 Pooled unadjusted trends in early neonatal death and low birthweight before and after PBF implementation



Shaded areas show 95% CIs

Consistent with the pooled results, PBF did not have a significant effect on early neonatal death or low birthweight in any of the study countries. Zambia's PBF may have resulted in a slight decline in early neonatal death, but the 95% confidence interval contained zero. However, several countries did see some effect on intermediate outputs. Facility delivery rose 8 percentage points in Burundi (95% CI: 0.02-0.14), antenatal care visits rose by 12 percentage points in Lesotho (95% CI: 0.01,0.22), and antenatal visit quality improved by 9 percentage points in Zambia (95% CI: 0.01,0.17). There were no effects on delivery quality or cesarean sections in any country. In Zambia, there were no effects on the primary or secondary outcomes when comparing the PBF districts to the unconditional financing arm rather than the pure control arm (Appendix Table A.1).

Table 2.4 presents the results when the pooled sample is restricted to the two sub-populations of interest. PBF increased antenatal care utilization by 8 percentage points (95% CI: 0.00,0.17) among poor women. It did not have any effect on the health outcomes or any of the other intermediate outputs in either of the populations of interest, robust also to looking at the subset of poor rural women (Appendix Table A.1).

Table 2.5 Effects of PBF on primary and secondary outcomes among populations of interest

	Poor women			High risk births		
Outcome	Percentage point change	95% CI	N	Percentage point change	95% CI	N
Early neonatal death	0.00	(-0.01,0.02)	9680	0.00	(-0.01,0.02)	10887
Low birthweight	0.00	(-0.03,0.03)	9680	0.01	(-0.04,0.05)	10887
Facility delivery	0.02	(-0.04,0.09)	8051	0.03	(-0.02,0.09)	8222
Delivery quality	-0.05	(-0.16,0.06)	3476	-0.07	(-0.18,0.05)	5570
C-section	-0.01	(-0.02,0.01)	8034	-0.01	(-0.03,0.01)	8205
ANC 4 visits	**0.08**	**(0,0.17)**	5122	0.04	(-0.02,0.1)	5771
ANC quality	0.06	(-0.03,0.14)	5152	0.00	(-0.06,0.07)	5824

Note: Bolded estimates signify confidence intervals that do not contain zero. Estimated coefficients for from pooled multivariable difference-in-difference regression, with standard errors clustered at the district level.

Discussion

PBF is considered an innovative approach to tackle the challenges to improving neonatal health outcomes that persist in many African countries. This study used quasi-experimental methods and population representative secondary data to assess the effect of PBF projects on neonatal health outcomes, and the quantity and quality of care in five African countries. Despite the large sample sizes from pooling the data, we found no effect on any of the examined outputs or outcomes. Although there were several positive impacts on utilization and antenatal care quality among individual country projects and among poor women, no project had a statistically detectable impact on either neonatal mortality or low birthweight. Furthermore, the PBF projects did not have detectable impacts on the health outcomes for two vulnerable sub-groups, poor women and women with a high-risk birth.

There may be several reasons for our null findings. First, the potential of PBF may be constrained by the ability of health facilities or providers to adjust their behavior to improve performance. In practice, they may already be operating at capacity given their environmental, educational, and structural constraints. For example, chronic staff shortages limited sustained improvement in Zimbabwe [69]. Poorly functioning health systems may instead require greater foundational change than adjustments to provider performance [70]. Second, PBF may have both positive and negative effects on different aspects of provider motivation [71,72], and its effects on non-incentivized services can be ambiguous [53,54]. Although improving health outcomes is a stated primary goal of all PBF projects in this study, it is possible that the projects had positive impacts on important clinical and non-clinical areas that we did not consider. Third, the specific design and implementation of the projects could affect their impacts. For example, the incentives may be too low or not be tied to the most effective behaviors. This may be particularly relevant for quality of care: PBF predominantly incentivizes structural quality [38], which may be only weakly correlated with care processes [73]. Problems with implementation, for example in delayed bonus payments, may have also impeded improvements [56]. Fourth, given the differences in program design and implementation across countries, there may have been mixed effects between programs that cancel out in the pooled analysis. The lack of detectable country-specific effects challenges this hypothesis, however. For example, even in Senegal where there were demand-side vouchers for women to attend ANC and facility delivery, there was no detectable impact on these intermediate outputs.

Despite the large pooled sample size, the study may also still not be adequately powered to detect changes in early neonatal death. An ex-post power calculation (Appendix Table A.2) suggests that the minimum detectable effect is a 0.67 percentage point change in the probability of early neonatal death, with the available sample size, 80% power and a 5% significance level. Smaller changes may be policy relevant, however, the small coefficient size and lack of effect in any of the intermediate outputs suggests that an effect would still not be detectable even with a larger sample size.

Some of our results differ from those of earlier impact evaluations of these PBF projects, which are summarized in Table 2.5. While no prior study had directly assessed the impacts on health outcomes, several studies found positive impacts on utilization or quality, particularly on rates of facility delivery [55–57]. We found a positive impact on facility delivery in Burundi, though smaller effect size than in earlier studies [55], and no impact in Zambia or Zimbabwe. There may be a number of explanations for this divergence, including differences in the sampling strategy, timing of data and inclusion criteria; differences in the covariates used to control for baseline differences; and our use matching to reduce covariate imbalance. There are also differences in how quality is measured. Our quality measures rely on a relatively small number of process measures from self-reports, whereas the earlier studies tend to use more indicators and rely more heavily on structural measures. For example, the Burundi evaluation uses a composite facility-based measure constructed using 57 structural and process indicators [55], while the large impact on delivery quality in Zambia is driven by the availability of equipment, medicines and supplies in the delivery room [57].

Table 2.6 Summary of effects from previous impact evaluations

	Burundi	Lesotho	Senegal	Zambia	Zimbabwe
Early neonatal death	NA	NA	NA	NA	NA
Low birthweight	NA	NA	NA	NA	NA
Facility delivery	22 pp	NA	NA	13 pp	13 pp
Delivery quality	17 pp[d]	NA	NA	57 pp	No effect
C-section	NA	NA	NA	NA	7 pp
ANC visits	No effect	NA	NA	No effect	No effect
ANC quality	17 pp[d]	NA	NA	Mixed[e]	Mixed[f]

Note: Statistically significant effects reported; all reported effects were positive. NA: Not assessed. [a](Bonfrer et al., 2014) [b](Friedman et al., 2016b) [c](Friedman et al., 2016a) [d]Facility quality measured overall, rather than by service. [e]Found improvements in iron supplementation and malaria drugs, decrease in urine sample taken, and no change in other 5 ANC quality measures assessed. [f]Found improvements in urine sample taken and tetanus injections, and no change in other 6 ANC quality measures assessed.

This study has several limitations. First, a woman's treatment status may have been misclassified based on her district of residence at the time of the interview. This may be the case if the woman moved districts between the birth and the survey, sought care outside of her district, or visited a private facility which did not receive the PBF intervention within an PBF district. While these cases should affect a small percent of women and should not differentially affect women in intervention or comparison districts, a misclassified status would bias the results toward the null. Second, the quality measures available in the DHS and MICS data sets were limited. We selected indicators for process quality that may have a large impact on neonatal health outcomes but only partially capture routine delivery and antenatal care quality. Third, the mostly non-randomized implementation of the PBF projects could result in residual confounding that persists despite matching at baseline. Although we found pre-trends to be largely parallel, there could be unobserved time-variant factors that differentially impacted the districts during the study period. A further discussion of notable social and political events in the study countries around the time of implementation is included in Appendix Text A.2, but nothing is expected to differentially impact treatment and control areas. Fourth, we were unable to look at a longer time frame beyond two years because of PBF implementation in the control areas in some of the countries at that time. Although neonatal mortality can be responsive to changes in the health system [74], it may take longer than this period to see an effect particularly if there were delays in signing contracts or delivering payments [75,76]. Finally, we were unable to look at treatment heterogeneity at levels lower than the country because of limited sample sizes.

The mixed and variable effects we observed across countries indicate scopes for learning from comparative studies. Such comparisons and innovations in measurement (e.g., of quality) can also be used to adjust ongoing projects [77]. The large number of HRITF-supported PBF pilots provides an important opportunity for such further research.

Overall, our results indicate that PBF – as implemented in the five projects we examined – may have limited impacts on neonatal health outcomes, as well as the associated utilization and quality pathways. While this does not preclude PBF from having other effects, positive or negative, this finding suggests caution with designing and deploying PBF with the goal of improving neonatal health outcomes at the population level. PBF may have other benefits, e.g., arising from increased autonomy and supervision [40], but must also contend with other criticisms, such the lack of domestic ownership and the diversion of attention and resources away from broader health systems strategies [76,78]. Different strategies will likely be needed to make meaningful progress on improving neonatal health outcomes in Sub-Saharan Africa.

Chapter 3 : The role of context in facility quality improvement: Implementation research of Tanzania's Star Rating Assessment

Authors:

Anna D. Gage[1], Talhiya Yahya[2], Margaret E. Kruk[1], Eliudi Eliakimu[2], Mohamed Mohamed[3], Donat Shamba[4], Sanam Roder-DeWan[4,5]

Affiliations:

[1]Harvard T.H. Chan School of Public Health, Boston, United States

[2]Ministry of Health, Community Development, Gender, Elderly and Children, Dodoma, Tanzania

[3]East Central and Southern Africa Health Community, Arusha, Tanzania

[4]Ifakara Health Institute, Dar es Salaam, Tanzania

[5]UNICEF, Dar es Salaam, Tanzania

Abstract

Since 2015, Tanzania has sought to improve the quality of all primary care facilities through its Star Rating Assessment. Facilities are assessed on quality and rated on a scale from zero to five stars. A facility quality improvement plan is developed based on the results and the data are shared with council and regional administrations. We examined two rounds of Star Rating data to identify the contextual factors associated with facility quality improvement, and to quantify the contribution of administration and proximity to the variance in quality improvement.

The outcome was the change in number of stars at each facility between the first and second rounds of assessment. We adapted the Consolidated Framework for Implementation Research to identify contextual variables that may be associated with facility improvement from several secondary data sources. We estimated multilevel regression models and a hierarchal spatial autoregressive model to estimate the proportion of variance at the facility and council levels, and the influence of nearby facilities and councils on improvement.

72% of the 5,595 primary care facilities improved their Star Rating score at reassessment. Factors associated with improvement include baseline performance, facility type and level, affiliation with the Results Based Financing program, population density surrounding the facility and distance to a major road. 20% of the variance in facility improvement was at the council level. Geographic clustering of improvement was not completely explained by the council; the improvement of nearby facilities and councils was also associated with a facility's own improvement.

Although the majority of facilities were able to improve their Star Rating score, there was great variation in their capacity to improve. Both council administration and proximity play important roles in a facility's ability to improve. Quality improvement interventions should be

designed to take advantage of mechanisms such as peer learning and peer pressure that operate above the facility level.

Background

Following the Astana declaration of 2018, primary healthcare is once again high on the global health agenda, with many countries renewing commitments to strengthen their primary care systems [79]. However, poor quality of care often limits the potential of primary care. In Tanzania in 2016, approximately 45,000 deaths were estimated to be due to poor quality of care, many of which were due to conditions that are addressable in the primary care setting [10]. Improving the quality of care provided is therefore necessary to realize the health benefits of primary care.

Quality improvement in healthcare has traditionally focused on micro approaches that rely on changing practices of individual providers or facilities [12,19]. However, there is growing recognition that these strategies may have limited impact within complex adaptive health systems. Macro and meso strategies that work at the system or area levels are needed to address the underlying social, political, economic and organizational structures that produce poor quality of care [12,20,80].

This study examines meso level drivers to quality improvement in the context of Tanzania's flagship primary care quality improvement strategy, the Star Rating Assessment. The Health Quality Assurance Division of the Ministry of Health implemented the Star Rating Assessment in 2015 as part of a larger government initiative to improve delivery of services [35,81,82]. The strategy relies on a data feedback approach. First, assessment teams comprised of two independent health workers and one member of the council health management team (CHMT) collected data from all primary care facilities on quality of care [35].

Results were translated into a rating of zero to five stars and shared with health facility administrators to jointly developed a tailored quality improvement plan based on their specific gaps [35]. Ratings were also discussed with CHMTs and regional health management teams. Some, but not all, administrations took an active interest in the Assessment and further supported facilities; for example, some used the Star Rating tool as a supportive supervision tool or encouraged learning across facilities. However, this support was not uniform across councils and regions, nor was there a uniform quality improvement plan. The decision of whether and how to improve was determined locally. The Ministry of Health originally planned to close facilities with zero stars, though this did not occur after many facilities scored zero stars at baseline. There were no other national positive or negative financial incentives for improvement tied to the rating. The facilities were reassessed in 2017-2018, and there are plans to continue additional rounds of reassessment every few years. The program was hypothesized to act through increasing accountability at multiple levels, using data to identify gaps, and supporting locally driven improvement solutions.

While the Star Rating Assessment was Tanzania's flagship strategy for improving quality of primary care, it was implemented in the context of other health system changes that may have influenced quality. First, decentralization was a major priority of the larger government initiative to improve service delivery [81]. Fiscal responsibility was first delegated to the councils in 2014 and then decentralized even further to the frontline health facilities through direct facility financing in 2018 [83,84]. Second, a Results-Based Financing (RBF) program was implemented in public facilities in eight of Tanzania's 31 regions (purposively selected) in 2015 to address quality and utilization of healthcare [85]. Star Rating Assessment results were used to determine eligibility for enrolling in RBF, but the programs were otherwise independently implemented. To enroll in RBF, facilities in these regions needed to receive more than zero stars at baseline or reassessment. Facilities in RBF regions that did not initially meet these criteria were given a 10-

million-shilling (about USD 4,500) “starter fund” to improve their quality. Once enrolled in RBF, the criteria used to evaluate performance for incentives differed from the Star Rating tool with some overlap; facilities continued to be assessed through both programs. Third, three other regions and two councils within another region also implemented the 10-million-shilling starter fund in zero-star public facilities to improve their quality independently of the RBF program.

A better understanding of how facility context shapes quality improvement will contribute to more targeted and effective improvement strategies in countries such as Tanzania. This study examines how the context of the Star Rating Assessment implementation was related to the degree of facility quality improvement over two rounds of the assessment. First, we identified the factors that were associated with greater facility level quality improvement in order to understand in what contexts the program was most effective. Second, we quantified the contributions of council administration and geographic proximity to the variance in quality improvement to understand how quality improvement may be related across groups of facilities.

Methods

Conceptual framework

The Consolidated Framework for Implementation Research (CFIR) was adapted for this study, both to make it applicable for a low-income context and to fit a nation-wide assessment (Appendix Figure B.1) [86,87]. First, ‘structural environment’ was added to the outer setting domain, and the inner setting was limited to the constructs of structural characteristics, networks and communications and culture. We hypothesize that these factors may be related to the support and capacity that a facility may have to improve. Second, the modified framework conceptualizes two pathways through which the outer setting may influence quality improvement: membership in local administration and geographic proximity to other facilities

[88]. Policies, management, supervision, and funds flow through the council, which is the lowest level of government charged with health facility administration. Separate from the membership in the council, a facility's location and immediate surroundings may play a role in determining implementation uptake in a low income setting such as Tanzania, where poor communication and high transportation costs isolate facilities. For example, proximity to high performing facilities at baseline may facilitate peer comparison and learning. The role of proximity in contributing to improvement is thus examined independent of council boundaries. These two pathways may influence the hypothesized accountability mechanisms of Star Rating.

Study sample

The baseline Star Rating Assessment was conducted in 2015 and included 6,993 primary healthcare facilities in mainland Tanzania. The assessment excluded facilities in Zanzibar and Pemba; referral hospitals, including national, zonal and regional referral hospitals; and stand-alone clinics such as maternity homes, dental clinics and other specialized clinics. It attempted to assess all dispensaries, health centers, and primary level hospitals in the public and private sectors. The reassessment took place in 2017 and 2018 and covered 7,289 facilities.

Facilities with two rounds of Star Rating scores were included in the study. Although the Assessment was conducted in the Dar es Salaam region at baseline, this data was unavailable, and therefore that region was excluded from this analysis. Facilities without geocoordinates and institutional facilities, such as those managed by the military, prisons, police or had unknown management type (2.2% of facilities) were further excluded from our main analytic sample.

Study design and dependent variable

This cross-sectional study examined variation in facility improvement between the two rounds of the Star Rating Assessment. The primary dependent variable was the facility's change in the Star Rating between the baseline and the reassessment rounds. The assessment included four domains and twelve sub-domains with different scoring weights as shown in Table 3.1. It included measures of structural quality, e.g. medicines and equipment, as well as processes quality, e.g. adherence to clinical guidelines and patient experience assessed through a facility audit, record reviews, and interviews with providers and clients [89]. Dispensaries, health centers and primary hospitals each had their own tools, with additional items required as the level increases [89]. The overall score ranges from 0-100%; a zero-star facility has a minimum domain score below 20%, one star 20-39%, two stars 40-59%, three stars 60-79%, four stars 80-89%, and five stars 90-100%.[8] We examined whether the underlying domain scores had bunching [90] just above the star thresholds which may signal manipulation in order to achieve a higher Star Rating score. We conducted a sensitivity analysis using the change in the overall score as a secondary dependent variable. The dependent variable data comes from the two sets of Star Rating Assessment data, which was merged with the 2019 Tanzania Master Health Facility database containing the facility coordinates.

Table 3.1 Domains and scoring for the Star Rating Assessment

Domain	Assessment area	Scoring	Examples of indicators
Management of health facility and staff performance	Legality and licensing	0%	Valid license observed
	Health facility management	10%	Staff attendance register is observed as completed
	Use of facility data for planning and service improvement	5%	HMIS is observed as up to date
	Staff performance assessment	5%	Interviewed providers aware of performance targets
Fulfillment of service charters and accountability	Social accountability	10%	Records of meetings indicate community participation
	Client satisfaction	5%	Interviewed clients have high average satisfaction score
	Organization of services	5%	Observed schedules of facility outreach
	Handling of emergency cases and referral system	10%	Last documented referral took less than an hour to transport
Safe and conducive facilities	Health facility infrastructure	10%	Privacy ensured in consultation area
	Infection prevention and control	10%	All service areas observed have running water and soap
Quality of care	Clinical services	15%	Review of three ANC records indicate adherence to clinical guidelines (i.e. iron supplementation)
	Clinical support services	15%	Essential medicines are observed available

Independent variables

Contextual factors that may influence a facility's ability to improve were identified following the modified conceptual framework. The variables were compiled from a range of data sources, defined and described in Table 3.2. Where possible, data from 2015 was used to be temporally aligned with the baseline assessment.

Table 3.2 Covariates organized by framework domains

Construct	Measure	Definition	Data source	Mean	SD
Outer setting					
Patient needs and resources	Population density	Sum of people within a 5 km radius of facility (x 1000)[a]	World Pop 2015 est. 100m res [91]	24	64
	Population demand for coverage[b]	Percent of women in council who gave birth in the past 5 years in a facility	Demographic and Health Survey 2016	0.71	0.22
	Informed consumers[b]	Percent of women in council who have completed primary education	Demographic and Health Survey 2016	0.73	0.15
	Healthcare agency[b]	Percent of women in council who are involved in decisions about their own healthcare	Demographic and Health Survey 2016	0.74	0.13
Cosmopolitanism	Facility density[b]	Number of facilities in council per 100,000 people	Star Rating Assessment data	15.4	7.2
	Urban council[b]	Facility is in a town council or municipality	Star Rating Assessment data	0.12	0.33
Structural environment	Accessibility	Distance in km to major road (bilinear interpolation)[a]	Open Street Map 2016	2.32	4.54
	Remoteness	Distance in 10 km to city of at least 50,000 population	Natural Earth version 2	7.1	5.1
Peer pressure	Facility rank at baseline	Percentile rank of star rating baseline score in comparison to other facilities in the council	Star Rating Assessment data	0.42	0.33
External policies and incentives	RBF participation	Public facilities from RBF regions with baseline scores above zero	Star Rating Assessment data	15%	
	RBF ineligibility	Public facilities from RBF regions with baseline scores equal to zero	Star Rating Assessment data	11%	

Table 3.2 (Continued)

External policies and incentives	Starter fund	Public facilities from starter fund areas with baseline scores equal to zero	Star Rating Assessment data	4%	
	External supervision	Supervisor visited facility in past six months and used checklist, discussed facility performance and helped facility make data-based decisions	Service Provision Assessment 2015	76%	
Inner setting					
Structural characteristics	Ownership	Public Private Non-profit	Star Rating Assessment data	81% 9% 10%	
	Level	Dispensary Health center Primary hospital	Star Rating Assessment data	85% 12% 3%	
	Baseline performance	Star rating at baseline	Star Rating Assessment data	0.81	0.71
	Human resources	Total number of full-time health providers[a]	Service Provision Assessment 2015	8.6	21.9
Culture (sub-analysis)	Routine data use	Facility has record of routine quality assurance system (measured at baseline)	Service Provision Assessment 2015	15%	
	Client responsiveness	Facility has procedure for reviewing patient feedback	Service Provision Assessment 2015	9%	
	Community engagement	Facility had staff-community meeting within six months	Service Provision Assessment 2015	64%	
	Management functions	Facility took an action following a recent management meeting	Service Provision Assessment 2015	46%	

[a]Natural log of variable used in models
[b]Council level variables

Variables from the Demographic and Health Surveys (DHS) were calculated at the council level and then applied to all facilities in the council. The DHS is only representative to the regional level, consequently there were some councils with very small sample sizes. To smooth the variation that arose from the small sample sizes, predicted values were used from a null three-level random intercept model with the households nested within councils and regions for these variables [92].

Given the limited data on the inner setting characteristics, a secondary analysis was conducted on a sub-sample of facilities that had data available from the 2014-2015 Service Provision Assessment (SPA). The SPA is a nationally representative facility survey that includes modules on facility management and provider motivation. SPA facilities were joined to the Star Rating data using the geocoordinates; the nearest facility within 1 km of the same level and management was considered a match in order to account for GPS error. Descriptive statistics from the SPA are presented with sampling weights (Table 3.2).

Additional covariates came from WorldPop, OpenStreetMap and Natural Earth (Table 3.2). In Tanzania, urban administrations are town councils or municipalities, while rural administrations are called district councils [93]. 'Councils' is used throughout this study to refer to both rural and urban administrations.

Analysis

Improvement in Star Rating scores across facility characteristics are described and mapped, interpolating using inverse distance weighting with a weight of 2 and clipped to 10 km around the facility in order to visualize trends in densely populated areas.

Guided by the conceptual framework, the contribution of the council level to facility improvement was examined by estimating a two-level random intercept model with facilities nested within councils. The percent of variation in improvement explained by the covariates was calculated as the difference in variance between the adjusted model and null model divided by the null model variance. A full random intercept model was run with all of the variables available from the SPA data as a sub-sample.

Second, the contribution of proximity in addition to council membership was examined through spatial analyses. Moran's I of the change in star rating was first calculated, testing for the presence of spatial autocorrelation against the null hypothesis of complete spatial randomness [94]. The residuals of the two-level random intercept models were also tested for residual spatial autocorrelation not explained by the included covariates or council. The Moran's I used an inverse distance weight matrix ($1/x^2$) for facilities within 50 km.

Finally, a hierarchical spatial autoregressive model [95] was fitted using Bayesian Markov Chain Monte Carlo algorithms to explicitly model the spatial processes at both the facility and council levels:

$$y_{ij} = \rho W_i y + \beta X_{ij} + \theta_j + e_{ij}$$

$$\theta_j = \lambda M_j \theta + u_j$$

Where y is the change in Star Rating in facility i and council j, ρ is a spatial lag term at the facility level, W_i is the facility inverse distance weight matrix for facilities within 50 km, X is the vector of covariates included in the full model, λ is a spatial lag term at the council level and M_j is a queen's contiguity spatial weight matrix for councils. This model accounts for the persistent autocorrelation in the facility and council error terms through two levels of spatial lag terms, which can be interpreted as the association of a facility's Star Rating improvement with improvement in nearby facilities (ρ) and adjacent councils (λ). As a sensitivity analysis, we also constrained λ to be zero, assuming that the spatial patterning only occurs at the facility level rather than the council level. Additional detail about the hierarchal spatial models is available in Appendix Text B.1.

The National Institute for Medical Research of Tanzania and the Ifakara Health Institute Institutional Review Board approved the original study; the Harvard Institutional Review Board determined this secondary analysis was not human subject's research. Analyses were conducted in R and maps were drawn in QGIS.

Results

5,595 facilities had two rounds of Star Rating scores available and met the inclusion criteria. 81% of facilities were public and 85% were dispensaries (Table 3.2). Facilities performed poorly on average at baseline: 34% of facilities scored zero stars at baseline and 52% scored 1 star. 15% of facilities participated in the RBF program, and a further 11% of facilities were public facilities in RBF regions but were ineligible because they did not meet the Star Rating criteria. There were on average 47 facilities per council or 15 facilities per 100,000 people. Among facilities with a SPA, 76% had external supervision in the past six months and 15% had a record of regular quality assurance activities before the Star Rating was implemented in 2015.

Figure 3.1 shows the change in number of stars from baseline to reassessment by facility type. 181 (3%) facilities received a lower star rating at reassessment in comparison to baseline, 1,386 (25%) received the same score, 2,531 (45%) improved by one star, and 1,497 (27%) improved by two or more stars. There was no difference in improvement between dispensaries, health centers and primary hospitals. Public facilities improved more than private for-profit and non-profit facilities. There was a strong baseline effect, with facilities that scored zero stars at baseline having the largest improvement. The fraction of facilities that had lower scores at reassessment were more likely to have scores of two or higher at baseline and be a private for-profit or non-profit facility. While scores in all domains decreased in these facilities, the largest declines in the fulfillment of service charter and accountability domain. There is evidence of bunching in the underlying domain scores just above the thresholds for the star cutoffs in both the baseline and reassessment (Appendix Table B.1), suggesting that there may be some manipulation to achieve higher star ratings.

Figure 3.1 Improvement in Star Rating by facility type, ownership, and baseline score

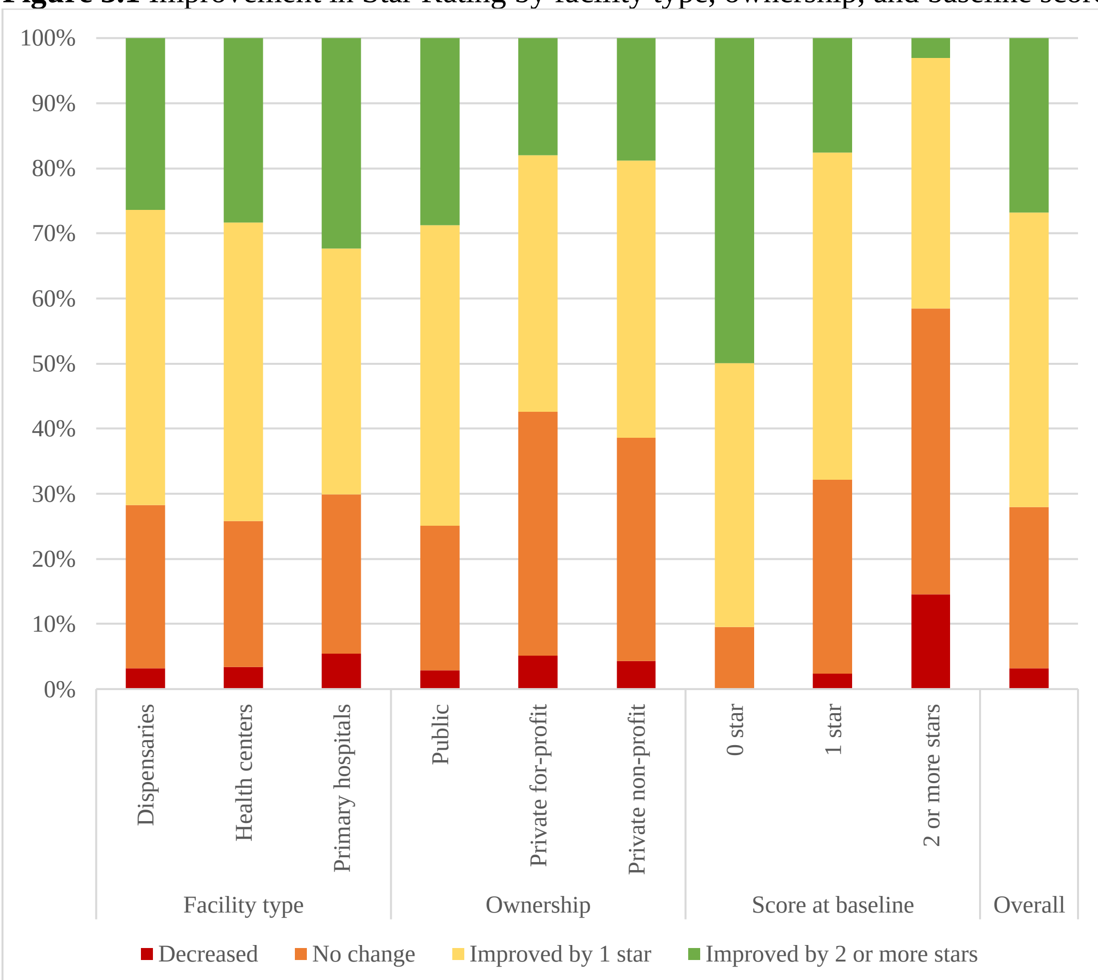


Figure 3.2 maps the baseline scores and unadjusted improvement. At baseline, Star Rating performance was best in Arusha and Kilimanjaro regions and poorest in Kigoma and Mtwara regions. Facilities improved by the most in Pwani region and in regions surrounding Lake Victoria with the exception of Mara region. Facilities improved the least in Mara, Tanga,

and Ruvuma. There was strong geographic clustering of both baseline performance and improvement, with Moran's I of 0.17 ($p<0.01$) and 0.18 ($p<0.01$) respectively.

Figure 3.2 Baseline and change in Star Rating scores

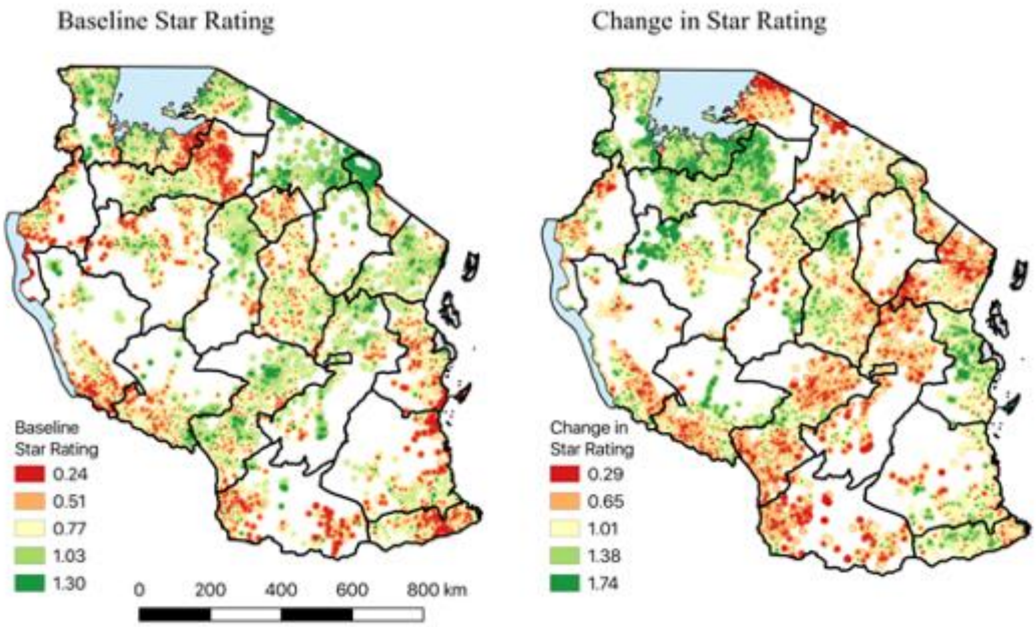


Table 3.3 shows the results of the random intercept models on the full study sample. In the null model, 20% of the variance in facility improvement was from variance between councils, while 80% was from facilities within councils. In the outer settings model, 29% of the total variance was explained by the covariates. In the full model which also includes the inner settings variables, 33% of the total variance was explained by the covariates. Holding baseline performance constant, primary hospitals and health centers improved by more than dispensaries, and public facilities improved by more than private or non-profit facilities. Greater population density around the facility and being closer to a major road were associated with greater improvement. RBF participation was associated with a 0.37-star improvement (95% CI: 0.27, 0.46), while being ineligible for RBF was associated with a 0.60-star improvement (95% CI: 0.50, 0.70). The starter fund outside of RBF regions was not associated with improvement. Every one-star increment in the facility's baseline score was associated with a 0.69 star decrease in

improvement (95% CI: -0.78, -0.6), indicating that low performing facilities had much greater improvement even after holding the other contextual factors constant.

Table 3.3 Random effects and HSAR models of Star Rating improvement in 5,595 facilities

	Null model		Outer settings model		Full model		HSAR model	
	Coef.	95% CI	Coef.	95% CI	Coef.	95% CI	Coef.	95% CI
Sum of people within 5 km radius of facility (ln)			0.07	(0.05,0.09)	0.07	(0.05,0.1)	0.07	(0.05,0.09)
Institutional delivery percent in council			0.20	(-0.22,0.62)	0.34	(-0.09,0.78)	0.30	(0.11,0.49)
Female primary education percent in council			0.04	(-0.56,0.63)	0.24	(-0.37,0.85)	0.23	(-0.06,0.51)
Healthcare decisions percent in council			-0.35	(-1.08,0.38)	-0.46	(-1.2,0.29)	-0.17	(-0.51,0.18)
Facilities in council per population			0.00	(-0.01,0.01)	-0.01	(-0.02,0.00)	0.00	(-0.01,0.00)
Urban council			-0.01	(-0.22,0.2.0)	-0.01	(-0.23,0.21)	0.06	(-0.04,0.16)
Distance to major road (ln)			-0.01	(-0.02,0.00)	-0.01	(-0.02,-0.00)	-0.01	(-0.02,-0.00)
Distance to large city (10 kms)			0.01	(-0.00,0.01)	0.00	(-0.00,0.01)	0.00	(-0.01,0.00)
Percentile rank at baseline			-1.14	(-1.2,-1.07)	0.03	(-0.15,0.2)	-0.18	(-0.34,-0.01)
RBF Participation			0.52	(0.44,0.61)	0.37	(0.27,0.46)	0.29	(0.20,0.38)
RBF ineligibility due to low baseline			0.65	(0.56,0.75)	0.60	(0.50,0.70)	0.50	(0.40,0.59)
Starter fund			0.09	(-0.03,0.2)	0.10	(-0.01,0.21)	0.08	(-0.03,0.19)
Ownership (Public ref.)								
Private					-0.29	(-0.37,-0.22)	-0.32	(-0.39,-0.24)
Non-profit					-0.10	(-0.16,-0.03)	-0.11	(-0.17,-0.04)
Level (Dispensary ref.)								
Health center					0.36	(0.31,0.42)	0.34	(0.28,0.4)
Primary hospital					0.76	(0.65,0.87)	0.74	(0.63,0.85)
Baseline performance					-0.69	(-0.78,-0.6)	-0.58	(-0.66,-0.5)
First level (facility) lag term							0.34	(0.27,0.4)
Second level (council) lag term							0.36	(0.1,0.61)
Constant	1.01	(0.93,1.08)	0.75	(0.18,1.31)	0.72	(0.15,1.29)	0.31	(-0.01,0.62)
Council variance	0.16		0.10		0.11		0.05	
Individual variance	0.63		0.46		0.42		0.44	
Moran's I of residuals (p value)	0.05	0.00	0.04	0.00	0.05	0.00		

Models with the sub-sample of SPA facilities are presented in Table 3.4. 18% of the variance in the null model is at the council level. In the second model, which includes all the variables from the full model in Table 3.2, the covariates explain 40% of the total variance, while the addition of the new variables from the SPA data in the final model only contributes one percentage point more of explained variance. In addition to the previously identified variables, availability of human resources and routine data use at baseline are also both associated with greater improvement on the Star Rating scale. The models which use the change in the overall score (0-100) are robust to these findings (Appendix Table B.2).

Table 3.4 Random effects models of Star Rating improvement in sub-sample of 672 facilities

	Null model		Prior full model		Additional inner settings model	
	Coef.	95% CI	Coef.	95% CI	Coef.	95% CI
Sum of people within 5 km radius of facility (ln)			0.03	(-0.03,0.1)	0.00	(0.00,0.00)
Institutional delivery percent in council			0.29	(-0.25,0.83)	0.29	(-0.26,0.84)
Female literacy percent in council			0.23	(-0.49,0.95)	0.35	(-0.38,1.08)
Healthcare decisions percent in council			-0.69	(-1.58,0.20)	-0.72	(-1.62,0.19)
Facilities in council per population			-0.01	(-0.02,0.00)	-0.01	(-0.02,0.00)
Urban council			0.21	(-0.07,0.49)	0.20	(-0.08,0.48)
Distance to major road (ln)			0.02	(0.00,0.05)	0.01	(0.00,0.02)
Distance to large city (10 kms)			0.00	(-0.01,0.02)	0.01	(-0.01,0.02)
Percentile rank at baseline			0.03	(-0.39,0.44)	-0.05	(-0.46,0.35)
RBF Participation			0.16	(-0.03,0.34)	0.17	(-0.01,0.36)
RBF ineligibility due to low baseline			0.51	(0.25,0.76)	0.49	(0.24,0.75)
Starter fund			-0.13	(-0.54,0.27)	-0.04	(-0.44,0.36)
External supervision					0.12	(-0.02,0.25)
Ownership (Public ref.)						
Private			-0.36	(-0.6,-0.13)	-0.37	(-0.61,-0.13)
Non-profit			-0.10	(-0.26,0.05)	-0.07	(-0.23,0.08)
Level (Dispensary ref.)						
Health center			0.46	(0.34,0.58)	0.15	(-0.01,0.31)
Primary hospital			0.86	(0.68,1.03)	0.19	(-0.12,0.49)
Baseline performance			-0.68	(-0.86,-0.51)	-0.68	(-0.85,-0.51)
Number of full-time health providers (ln)					0.18	(0.10,0.26)
Record of routine quality assurance					0.19	(0.06,0.32)
Procedure for reviewing patient feedback					-0.04	(-0.16,0.08)
Recent staff-community meeting					-0.06	(-0.21,0.09)
Follow-up on management meeting					0.02	(-0.10,0.14)
Constant	1.01	(0.92,1.11)	1.37	(0.53,2.22)	1.31	(0.62,1.99)
Council variance	0.14		0.07		0.08	
Individual variance	0.65		0.41		0.38	

In the random intercept models, Moran's I of the residuals dropped from its unadjusted 0.18 to 0.05 ($p<0.05$), indicating that a large portion of the spatial autocorrelation arises from facilities' membership in councils. However, the still positive clustering of the residuals indicates the presence of spatial patterns associated with improvement beyond council membership. The hierarchical spatial model which includes spatial lag terms at both levels is also presented in Table 3.3. Both lag terms are large and significantly associated with quality improvement. The covariates in this model account for 39% of the total variance. Two additional covariates, the council's institutional delivery and percentile rank at baseline, are significantly associated with improvement in the Star Rating. Controlling for a facility's baseline performance, facilities that ranked lower than other facility's in their district had greater improvement.

Discussion

The success of quality improvement interventions is dependent not only on processes within the facility, but also on the context in which they operate. This study examined how the context of the Star Rating Assessment implementation was related to the degree of facility quality improvement in Tanzania. Both council administration and proximity to other high-performing facilities played important roles in a facility's ability to improve its quality of care as measured by the Star Rating system. Councils accounted for approximately 20% of variance in improvement, and facilities improved more when they were near to other facilities that also improved.

Among the factors examined, RBF participation, RBF ineligibility, baseline performance and facility type were consistently among the largest predictors of Star Rating improvement. Facilities ineligible for RBF participation because they did not meet the Star Rating criteria improved by about 60% more than facilities that actually received funding from RBF. The incentive to become eligible to enroll in RBF was thus potentially greater than the incentives that the RBF put into place. Councils may have also put significant pressure on the ineligible facilities in order to receive the additional funding. Primary hospitals and public facilities were much more likely to improve than dispensaries or privately managed facilities. Private facilities may have felt less pressure to improve given their separate funding streams, or councils may have given them less support to improve. Dispensaries may have fewer financial and human resources for improvement than primary hospitals.

More research is needed to understand the causal mechanisms behind these associations. Forthcoming qualitative work with the facility managers supports some of the hypotheses around context from this study. For example, managers noted that the Star Rating stimulated competition between neighboring facilities, while others felt jealous of surrounding facilities that had higher scores at baseline. They also supported the idea that the council administration was critical to facilities that improved in clarifying and strengthening accountability streams, both between the facility and the council and also between the community and the facility. Council proximity may have acted through similar mechanisms to facility proximity, or the associations may be more reflective of the general geographic context in which the councils were operating.

This work has several implications for policy. First, the results suggest that the data and feedback strategy was not just implemented at a micro level where only facility-level characteristics were important to the improvement process. Rather, council administration and peer learning and pressure from neighboring facilities functioning at the meso and macro levels may have been important pathways for improvement. Just as health systems interventions have been designed more explicitly with the district in mind, interventions could also consider more explicitly how to take advantage of neighborhoods of health facilities and networks across and within district borders [96]. In the next round of assessment, certificates will be made for each facility to post publicly to further encourage social accountability and peer advocacy among facilities. Second, the strong association with baseline performance indicates that there was a strong floor effect or regression to the mean; most of the improvement came from low performing facilities. Now that most facilities are now at one or two stars, new strategies may be required in order to see continued improvement. Finally, the difference in improvement between facilities enrolled in and ineligible for RBF indicates a need for reexamining the design of RBF incentives [78].

This study has several limitations. First, there was limited data availability particularly in mapping the CFIR to a national level program in a low-income setting. Ideally, more data on areas such as council interactions with the facilities, culture at the facilities and readiness for change, and characteristics of the health workers such as self-efficacy and knowledge of the intervention, would be available. While the SPA supplemented national data in a sub-sample, this dataset was still limited on these constructs. Second, the Star Rating tool is limited in the way that it assessed quality. Data on health outcomes was unavailable, and it is unclear how the assessed quality relate to health outcomes. There are also many more items on inputs which could overemphasize their importance. The available data on user experience may be influenced by user's differential and growing expectations of healthcare systems [97]. The Ministry of Health is also revising the tool in subsequent rounds to put less emphasis on inputs. Finally, the Star Rating data was collected by health workers who were affiliated with the CHMTs. This has two possible consequences: first, we were unable to disentangle the variance from councils from the variance from the data collectors. Second, some CHMTs may have been motivated to inflate the ratings that were just under the threshold to just above the threshold during reassessment to show greater improvement, which would have contributed to an overestimate of the amount of improvement. While there was evidence of bunching in domain scores just above the thresholds, the analysis was robust to using the overall score rather than the star rating.

The majority of primary care facilities improved on the Star Rating scale in Tanzania between the first two rounds of assessment, however this improvement was not uniform. Identifying the contextual factors which facilitated or inhibited improvement can help to design better quality improvement interventions that take this context into account, through for example supporting peer learning. Such considerations will be critical as quality improvement interventions shift from the micro to the meso or macro scales.

Chapter 4 : Do clinical colleagues matter? The quality of delivery care in Dire Dawa, Ethiopia

Authors:

Anna Gage[1], Bereket Yakob[2], Margaret McConnell[1], Tsinuel Girma Nigatu[2], Brook Damtachew[3], Sebastian Bauhoff[1], Margaret Kruk[1]

Affiliations:

[1]Harvard T.H. Chan School of Public Health, Department of Global Health and Population, Boston, United States

[2]University of British Columbia School of Population and Public Health, Fenot Project, Addis Ababa, Ethiopia

[3]Dil Chora Hospital, Department of Obstetrics and Gynecology, Dire Dawa, Ethiopia

Abstract

Poor quality of delivery care in Ethiopia is impeding progress on reducing maternal and neonatal mortality. Previous studies examining the quality of maternity frequently attribute the care provided to a single provider; however multiple providers are often involved. We examine how characteristics of provider groups influence quality of care.

Observations of deliveries and the immediate postpartum period were conducted in two hospitals and seven health centers in Dire Dawa Administration, Ethiopia. Five phases of the delivery were identified (first exam, first stage of labor, third stage of labor, immediate newborn care, and immediate maternal postpartum care), and a provider was assigned responsibility for the quality of routine care provided during each stage. Defining a provider's coworkers as the other providers who cared for the same client during a different stage, we examined three key group characteristics: the typical performance of the index provider's coworkers, the seniority of the coworkers' cadre relative to the index provider and the years of experience of the coworkers relative to the index provider. We estimated the associations between quality of care and these group characteristics, controlling for individual provider's characteristics and contextual factors.

824 clients and 95 unique providers were observed across the nine health facilities. For the average client, 50% of the recommended routine clinical actions were completed during the delivery overall, with immediate maternal postpartum care being the least well performed (17% of recommended actions). More than one healthcare provider was involved in 55% of deliveries. The number of providers was unassociated with the quality of care, but a one standard deviation increase in the coworkers' performance was associated with a two-percentage point increase in quality of care ($p<0.01$) or 4% relative to the mean; this association was largest among providers

in the middle quartiles of performance. Junior cadres on average also provided more routine care actions than their senior counterparts.

A provider's typical performance had a modest positive association with quality of delivery care given by their coworker. As delivery care is often provided by multiple healthcare providers, examining the dynamics of how they influence one another to produce quality care can provide important insights for improvement interventions.

Introduction

Ethiopia's health system, like those in other low-resource settings, has gaps in maternal and newborn care quality that pose a barrier to improved health outcomes. Although the facility delivery rate has increased dramatically in the past 10 years from 10% in 2011 to 48% in 2019, maternal and newborn mortalities remain high, with 401 maternal deaths per 100,000 live births and 33 neonatal deaths per 1000 live births [5],[98]. The majority of such deaths could be avoided with high quality neonatal and maternal care [99,100]. However, in 2018, just 20% of mothers who delivered in a health facility were estimated to receive a high quality of routine maternal care in the Tigray region in northern Ethiopia [101].

Examinations of poor maternal care quality in similar contexts have found that there is often wide variation in the quality that a single provider will provide across different deliveries. For example, in Uganda the quality of routine actions during labor and delivery varied as much as sixty percentage points across the deliveries for which a single healthcare provider was responsible [102]. In Kenya and Malawi, two studies found that the provider contributed very little to the explained variance in the technical and interpersonal quality of maternity care [103,104]. Delivery level factors such as the time of day or mother's characteristics were unable to explain this variation [103]. The importance of the provider to quality may vary over the course of a delivery, however: another study found that healthcare providers were more important in accounting for the quality of the intrapartum period rather than the assessment or postpartum phases [105]. While these studies have also found that facility, region, and country-level factors can play important roles in determining the quality of care, it is worth further considering why provider quality varies in order to shed light on potential improvement strategies.

One limitation of the above studies in their examination of quality is the focus on attribution of quality to a single healthcare provider over the course of a labor and delivery. While many mothers in Ethiopia and globally receive care from a single healthcare provider, this is not the only model. Rather, multiple healthcare providers may attend a single delivery over the course of the labor, delivery, and immediate postpartum period, particularly during more complex cases that lead to higher morbidity and mortality. When groups of providers attend a single patient, their interactions with one another may affect the quality of care [106]. Indeed, the disciplines of complexity science and team science suggest that provider groups are more than the collection of independently acting individuals, but rather they may influence one another in nonlinear and dynamic ways [106,107]. The culmination of these interactions between providers creates the informal group norms and culture around the quality of care that may affect realized quality and health outcomes. Formal hierarchical relationships and channels of provider supervision may also affect quality of care.

Groups or teams of providers have been examined in healthcare contexts in high-income countries. For example, studies have found that the quality of HIV care is influenced by the performance of a provider's peers on quality [108] and that a provider's patterns of prescribing medications are affected by the presence of a specialist in a provider's network [109]. Similarly, studies have found moderate effects of group characteristics such as professional composition and aspects of teamwork such as coordination on quality of care [110–112]. However, these dynamics are less examined in healthcare in low- and middle-income countries. Using observations of delivery care in Ethiopia and adopting a complex systems lens, this study seeks to understand how the number of providers and group dynamics are associated with the quality of delivery care.

Methods

Conceptual framework

We draw on the theory of small groups as complex systems to inform the conceptual framework for this analysis [107]. Characteristics of groups include connections between members and to the projects of the group, coordinating their behavior in pursuing collective projects and using a shared set of tools, knowledge, and other resources. Groups can be considered open systems: that is, members may be part of more than one group and they interact with the environment in which they work [106,107]. In this analysis, we defined a group of healthcare providers to be all providers who care for a specific woman and newborn over the course of the delivery. Their group identity was thus focused on the shared objective of providing high quality care, and they drew on one another as well as the broader facility environment in achieving this objective.

Arrow et al argue that group behavior involves interactions at three levels [107]. First, behavior is influenced by the constituent elements of the groups, in this case the individual providers who bring different experiences, training, and roles to the group. Second, there are interactions of the group as an entity, which will include feedback loops between group members. Third, there are interactions with the context in which the group is embedded; this context includes the characteristics of the facility as well as the characteristics of the patient and specific delivery for which the group is providing care.

We focused our analysis on the interactions of the group as an entity, with an emphasis on potential feedback loops on quality of care. We also accounted for care context. Drawing on the peer effects literature, we hypothesized several mechanisms by which group members may influence one another in their provision of quality care. First, members may have influenced one another through hierarchal relationships defined between providers by their cadre rank or years of experience. For example, a provider working with their superior may have performed better due to their better supervision or social pressure. Second, there may be informal influence between providers of different abilities. When working with a high performing group member, there may have been positive spillovers because of social pressure, knowledge spillovers or social comparison [113]. For example, a high-quality provider may remind his or her fellow providers to complete some particular action for the delivery. Conversely, there may be negative spillovers due to free riding [113], for example if a provider decides not to monitor the woman's vitals because she assumes that her high performing colleague would have already detected any problems with the labor. While our data does not permit investigation into these specific mechanisms, we assessed whether working with a high performing colleague is associated with higher or lower performance.

Setting

Dire Dawa Administration is a city administration in Eastern Ethiopia with a population in 2019 of approximately 493,000 people, with 63% living in urban Dire Dawa City and the remaining population residing in rural areas surrounding the City [114]. In the five years preceding 2019, 84% of pregnant women received at least one antenatal care visit and 69% delivered in a health facility [114]. Despite rising utilization rates, neonatal mortality increased between 2011 and 2016 to 36 deaths per 1000 live births, signaling potential problems with the quality of care received [18]. In 2021, the administration had 53 public health facilities: 2 hospitals, 15 health centers and 35 health posts. In addition, there were 5 private hospitals and 7 private clinics.

Data

This analysis was part of a broader study to understand the quality of maternal and immediate postpartum care in Dire Dawa Administration's public health system and how women's past experiences with services impact her delivery decisions. Cross-sectional primary data was collected in 9 facilities. Public facilities with the highest volumes of deliveries using the 2019 health information system data were included in the study; these facilities were collectively responsible for 83% of the facility deliveries in the region in 2019. Private facilities were excluded from the study. The 9 selected facilities consisted of two public hospitals, four health centers in Dire Dawa city and three rural health centers.

Data tools relevant to this analysis include observations of deliveries, interviews with observed clients upon discharge from the facility, and interviews with all providers who provided care. First, all providers who provided intrapartum or immediate postpartum care in the study facilities were invited to take part in a provider survey that asked about their training, perceptions of the working environment and quality of care, and knowledge of complications diagnoses and management.

Second, quality of care was assessed through observations of deliveries by trained health workers. All clients presenting for delivery during the observation period were invited to participate in the study and their care was observed from the time of arrival at the facility until 6 hours postpartum or discharge from the facility. Data collectors identified which items providers completed on a checklist adapted from the Maternal and Child Health Integrated Program tool from USAID. For groups of actions (e.g., first exam), the collector also indicated which provider conducted the actions. For this analysis, only deliveries that were observed from admission to discharge were included, so Cesarean section deliveries and intrapartum referrals were excluded. The observation checklist was organized into thirteen discrete modules; the ones relevant to this analysis are those for the first exam, checks on the client during the first stage of labor, third stage of labor, immediate newborn care and immediate maternal postpartum care. Finally, all participants whose care was observed were invited to participate in an exit interview upon their discharge from the facility.

Observations of care were conducted in most facilities from December 22, 2020 to February 20, 2021. However, until January 31, 2021, one of the hospitals (Sabian Primary Hospital) was not accepting maternity patients because it was a designated Covid-19 treatment facility. Observations of delivery care in Sabian, therefore, occurred between February 13 and March 21, 2021.

Outcome

The primary outcome of interest for this analysis was the quality of routine maternal care, adapted from the quality of processes of intrapartum and immediate postpartum care index (QoPIIPC) [115]. The original index contains 20 indicators of routine actions that should be conducted during every delivery. To attribute the performance of the indicators to a particular provider, we separated this index into five sub-indices associated with the module of the checklist (first exam, first stage of labor, third stage of labor, immediate newborn care, and immediate maternal postpartum care). For example, the first exam sub-index consists of seven items that should be completed during the first exam such as taking pulse and asking whether the client experienced vaginal bleeding. The items making up each sub-index are presented in Table 4.1. While the second stage of labor was also observed by the team, the QoPIIPC index does not include any indicators particular to this stage. Performance on each sub-index, which we refer to as the quality of a stage, is calculated as the percent of actions that were completed so it ranges from 0 to 100%. We attributed the quality of each stage for every delivery to the provider who was noted as providing care during that stage; see Appendix Figure C.1 for an example of this structure.

Table 4.1 Components of quality of intrapartum and immediate postpartum care processes index by delivery stage

	Percent complete	N
Average of first exam actions	47%	823
Checks woman's HIV status	68%	809
Asks whether woman has experienced headaches or blurred vision	6%	823
Asks whether woman has experienced vaginal bleeding	7%	823
Takes blood pressure during initial client assessment	73%	822
Takes pulse during initial client assessment	67%	822
Washes hands before initial examination	13%	823
Wears gloves before vaginal examination	99%	781
Average of first stage of labor actions	48%	795
At least once, explains what will happen in labor	42%	824
Prepares uterotonic drug to use for AMTSL	81%	791
Uses partograph during labor	53%	793
Prepares bags and masks for neonatal resuscitation	14%	783
Average of third stage of labor actions	72%	822
Correctly administers uterotonic	56%	820
Assesses completeness of placenta and membranes	74%	819
Assesses for perineal and vaginal membranes	90%	819
Ties or clamps cord when pulsations stop, or by 2-3 minutes after birth	54%	821
Average of immediate newborn care items	81%	694
Immediately dries baby with towel	98%	694
Places newborn on mother's abdomen skin-to-skin	64%	690
Average of immediate maternal postpartum actions	17%	824
Takes mother's vital signs 15 minutes after birth	0.2%	823
Palpates uterus 15 minutes after birth	19%	824
Assists mother to initiate breastfeeding	37%	695
Quality of Intrapartum and Immediate Postpartum Care Processes index	51%	828

Notes: AMTSL: Active management of the third stage of labor

Several of the indicators included a timeliness component, for example, whether the provider checked the mother’s vital signs 15 minutes after birth. To credit providers who completed these actions outside of the designated timeframe, we also constructed an alternate index without any time limits on the actions as a sensitivity analysis. This affects the overall index as well as the intrapartum stage and immediate maternal postpartum stage sub-indices.

13% of delivery observations did not include the module on checks during the first stage of labor. We deemed 30% of these missing observations as "valid missing" if labor was induced directly after the first exam or if there was less than 60 minutes between the end of the first exam and delivery, as there may not have been time for the enumerator to complete the module. For the remaining observations, non-performance of the actions under the first stage (e.g., missing actions imputed as 0) was attributed to the provider who conducted the first exam. Among observations where this stage was not missing, the same provider conducted the first exam and the first stage 65% of the time, while 33% of observations had the same provider for the first and third stages of labor.

Independent variables

As described above, each stage of a delivery was assigned to a single responsible provider. Using this structure of the data, we defined a group as the providers that cared for a single client at different stages of her delivery. For deliveries with more than one provider, we were interested in how quality for a particular delivery stage completed by the index provider is associated with the characteristics of other providers in the group. While there may be many other providers working at a facility that are unassociated with a delivery, we define these specific provider groups as the index provider's coworkers for the delivery.

We defined three independent variables of interest. First, we examined the cadre of the coworkers relative to the index provider. We created a binary variable for whether the coworkers had a cadre superior to the index provider. There are five categories of cadres, here ranked from low to high: midwife or nurse (diploma), midwife or nurse (Bsc.), health officer, general practitioner, and integrated emergency surgical officer (IESO). Second, we examined the years of professional experience of the coworkers relative to the index provider. Similar to cadre, we created a binary variable for whether coworkers were more experienced than the index provider.

Third, we defined a measure of coworker performance following a two-step approach [113]. In the first step, we created an individual provider performance score for every provider observed, which captures the index provider's own capabilities. We specified a fixed effects model to estimate the provider's ability across all deliveries and delivery stages where they were the responsible provider:

$$Qual_{ij} = \beta_0 + \beta_1 Del\_stage_i + \theta_j + \epsilon_{ij}$$

Where $Qual_{ij}$ is the quality of the index delivery stage i for provider j, Del_stage is a fixed effect for the four stages of delivery, and θ_{ij} is the provider fixed effect. We controlled for delivery stage because of the differences in quality between each stage. The provider fixed effect θ_j thus became a measure of an individual provider's capabilities.

In the second step, we took the average of the provider capabilities measure for all other providers who cared for the same client besides the index provider. For example, suppose three providers (ID 11, 12, and 13) cared for a single client (01) at different stages of her delivery; their respective capability scores are -0.04, 0.11, and 0.02. For a stage that provider 11 completed, the average coworker performance for this stage is thus the average of the capability scores of providers 12 and 13: (0.11+0.02)/2=0.065. As a sensitivity analysis for deliveries with three or more providers, we also compared the average coworker performance with the best coworker's performance. The coworker performance measure was standardized for the analysis.

Covariates

Drawing on the conceptual framework, we defined several covariates for the individual providers and the contextual environment. First, we included the index provider's cadre and number of years of experience. Second, the context environment included characteristics of the client, the delivery, and the facility. These include whether the birth had a complication (neonatal resuscitation initiated, newborn referred to the NICU, or mother treated for post-partum hemorrhage or eclampsia); whether the delivery was at higher risk for a complication (grand multiparity, mother younger than 18 or older than 35, or multiple births); time of delivery (morning 8 am-6 pm, or night 6 pm-8 am); and the client's wealth. Client wealth was defined by quintiles within the analytic sample based on a principal components analysis of household assets in the client exit interview. Items in the asset index include ownership of car, motorbike, bicycle, refrigerator, phone, television, radio and bank account as well as the type of roof, wall, floor, cooking fuel and presence of electricity and toilet. Finally, we also controlled for delivery stage as defined above and facility fixed effects.

Analysis

We first described the quality of care provided to the study sample on the QoPIIPC index and the characteristics of the clients and providers. We examined whether quality differed by the number of providers caring for the client and examined what factors were associated with group-based care.

We then fit a linear mixed-effects model to assess the contribution of group dynamics to quality of care, with observations of delivery stages nested within the index provider:

$$Qual_{ij} = \beta_0 + \sum_{k=1}^{3} \beta_k \, Group_{ik} + \sum \beta_4 Covar_{ij} + \delta_j + \epsilon_{ij}$$

Where $Qual_{ij}$ is the quality of delivery stage i for index provider j; $Group_{ik}$ are the three independent measures of interest and $Covar_{ij}$ is a vector of the provider and contextual covariates listed above. Standard errors were clustered at the level of the index provider. Covariates were missing for a small number of providers and clients; we used multiple imputation such that all observed deliveries meeting the criteria could be included in the analysis. This analysis was conducted among all group deliveries. In addition, we conducted sub-analyses separately for each stage of delivery.

We further investigated how the coworker performance measure interacts with the other group characteristics. First, we categorized the index provider's performance into quintiles and interacted it with coworker performance to understand how relative performance may be associated with quality. Second, we also ran interaction models between coworker performance and the two measures of provider rank (cadre and years of experience) respectively. We used these interaction models to predict quality across the range of coworker performance holding all other covariates at their means and graphed these marginal models.

Results

Over the data collection period, 983 clients in the nine facilities were invited to participate in the study. 979 clients (99.6%) agreed to participate, and 828 (84%) were observed throughout the whole delivery and thus met the inclusion criteria for this analysis. The observed clients were cared for by 95 unique providers; 84 (88%) of them were interviewed for the study.

Client and provider characteristics are shown in Table 4.2. 452 (54%) of the observed deliveries took place at one of the two study hospitals, while the remaining deliveries were at the seven health centers. 213 (26%) of clients experienced a complication during the delivery and 87 (11%) had a higher risk pregnancy. 457 (55%) of deliveries had more than one provider attending over the course of their delivery: 72% of hospital deliveries and 35% of health center deliveries. Among the deliveries with more than one provider, the mean number of providers was 2.4. Two thirds of providers were midwives or nurses (BSc.); providers had an average of 6.2 years of cumulative professional experience.

Table 4.2 Delivery and provider characteristics

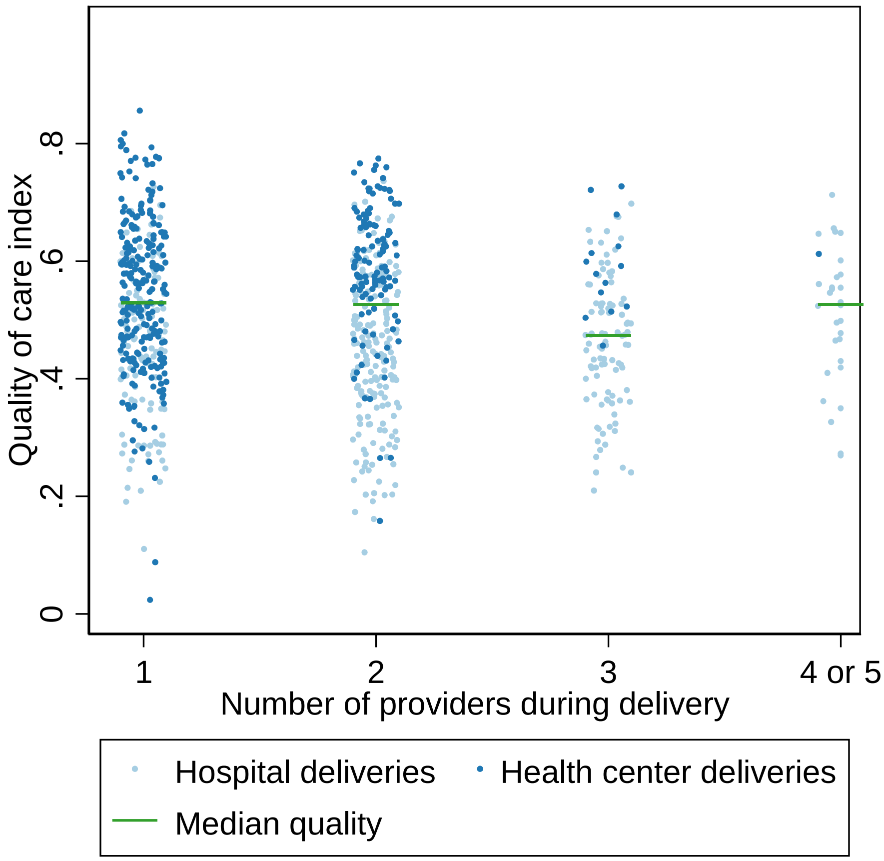
	All deliveries		Hospitals (N=2)		Health centers (N=7)	
	N	%	N	%	N	%
Client and delivery characteristics						
N deliveries observed	828		452		376	
Client's primary language						
Oromiffa	484	63%	196	49%	288	79%
Amharic	178	23%	129	32%	49	13%
Somali	81	11%	52	13%	29	8%
Other	23	3%	23	6%	0	0%
Poorest wealth quintile	150	18%	52	12%	98	26%
Experienced complication	213	26%	69	15%	144	38%
Higher risk pregnancy	87	11%	45	1%	42	11%
Time of delivery						
Day (8 am-6 pm)	341	41%	186	41%	155	41%
Night (6 pm-8 am)	487	59%	266	59%	221	59%
Provider characteristics						
N providers interviewed	84		50		34	
Years of experience (mean/sd)	6.2	4.5	5.9	3.6	6.5	5.7
Female	54	64%	31	62%	23	68%
Cadre						
Midwife or Nurse (diploma)	13	15%	7	14%	6	18%
Midwife or nurse (BSc)	55	65%	33	66%	22	65%
Health officer	3	4%	0	0%	3	9%
General practitioner	7	8%	4	8%	3	9%
IESO	6	7%	6	12%	0	0%
Provider group characteristics						
Deliveries with >1 provider	457	55%	326	72%	131	35%
N providers among group deliveries (mean/sd)	2.38	0.63	2.48	0.70	2.12	0.35
At least one health officer, GP or IESO among providers in group	76	16%	65	20%	11	8%
At least one provider with over median years of experience	198	43%	119	37%	79	60%

Notes: Delivery complication is neonatal resuscitation, newborn referred to the NICU, or mother treated for post-partum hemorrhage or eclampsia. Higher risk is mother grand multiparous (5 or more births), younger than 18 or older than 35, or has multiple births. IESO: Integrated Emergency Surgical Officers.

The quality of intrapartum and immediate postpartum care processes provided to the clients in this study was poor (Table 4.1). During the average delivery, only half of the recommended routine actions were done. Actions ranged from 0.2% of mothers whose vitals were checked fifteen minutes after birth to 98% of babies that were dried immediately with a towel. The recommended actions were most often completed for the immediate newborn care stage, while they were least often done during the immediate maternal postpartum care stage. When the index was defined without any time constraints on the actions, average quality of care rose to 58% of recommended actions across the whole delivery (Appendix Table C.1). However, still only 3% of mothers had their vitals checked after delivery. Over 90% of the variance in quality of care was at the client level, rather than the facility or index provider levels (Appendix Figure C.2).

Between 1 and 5 providers cared for a single client and baby over the course of the delivery. The number of providers was unassociated with quality of care (Figure 4.1 and Appendix Table C.2). The primary determinant of having more than one provider care for a client was the place of delivery: the odds of a group delivery were 4.5 times higher in a hospital than health center (Appendix Figure C.3). Time of delivery, client wealth and language were slightly associated with a group delivery, but neither higher risk clients nor complicated deliveries was associated with more providers. The remaining results are only among deliveries with more than one provider.

Figure 4.1 Quality of intrapartum and immediate postpartum care processes by number of providers and facility type



In the bivariate associations between group characteristics and quality of delivery stages (Appendix Table C.4), coworker's performance and seniority by cadre are associated with quality of care. Scatterplots of the index provider and coworker performance (Appendix Figures C.4 and C.5) further show that this is in part due to quality clustering by facility. Table 4.3 shows the associations of coworker characteristics with quality of care adjusted for all covariates. After controlling for the facility fixed effect, index provider and client characteristics, a one standard deviation increase in coworker performance was associated with a 2-percentage point increase in quality of care during a given delivery stage, or 4% relative to the mean performance. This association was larger during the first stage of labor (4.2 percentage point increase) (Appendix Table C.5). Working with coworkers that were superior in rank to the index provider was associated with 3.5 percentage point lower quality, while the coworker's relative experience was unassociated with quality. In deliveries with three or more providers, the average coworker performance had a higher association with quality than the maximum coworker performance (Appendix Table C.6).

Table 4.3 Group dynamics associations with quality of care (Outcome range 0-1)

	Coef.	p value	95% CI
Group characteristics			
Peers' performance across all deliveries	0.023	0.00	(0.01,0.04)
Peers are more senior cadre than index	-0.035	0.02	(-0.07,-0.01)
Peers are more experienced than index	-0.009	0.46	(-0.03,0.02)
Number of providers	0.000	0.96	(-0.02,0.02)
Index provider characteristics			
Years of experience	-0.002	0.41	(-0.01,0.00)
Provider cadre (Midwife or nurse diploma ref.)			
Midwife or nurse Bsc.	-0.043	0.04	(-0.09,-0.00)
Health officer	-0.141	0.02	(-0.25,-0.03)
GP	-0.064	0.00	(-0.11,-0.02)
IESO	-0.066	0.48	(-0.25,0.12)
Context and environment			
Delivery complication	0.004	0.81	(-0.03,0.04)
Higher risk pregnancy	0.003	0.88	(-0.04,0.04)
Night delivery (morning reference)	0.007	0.51	(-0.01,0.03)
Client wealth index (poorest reference)			
Wealth 2	0.037	0.06	(-0.00,0.07)
Wealth 3	0.071	0.00	(0.03,0.11)
Wealth 4	0.039	0.05	(-0.00,0.08)
Wealth 5 (wealthiest)	0.050	0.01	(0.01,0.09)
Delivery stage (first exam ref.)			
First stage of labor	0.001	0.96	(-0.03,0.04)
Delivery care	0.233	0.00	(0.20,0.26)
Immediate newborn care	0.300	0.00	(0.26,0.34)
Immediate postpartum care	-0.320	0.00	(-0.35,-0.29)
Facility (Dil Chorra Hospital ref.)			
Sabien Primary Hospital	0.017	0.35	(-0.02,0.05)
Biyowale Health Center	0.127	0.01	(0.04,0.22)
Legeharae Health Center	0.129	0.00	(0.07,0.19)
Melka Jebdu Health Center	0.117	0.00	(0.06,0.18)
Wahil Health Center	-0.049	0.09	(-0.11,0.01)
Gende Gerada Health Center	0.083	0.05	(0.00,0.16)
Goro Health Center	0.148	0.00	(0.09,0.21)
Jelobelina Health Center	0.160	0.00	(0.09,0.23)
Constant	0.438	0.00	(0.35,0.53)
N obs	2189		

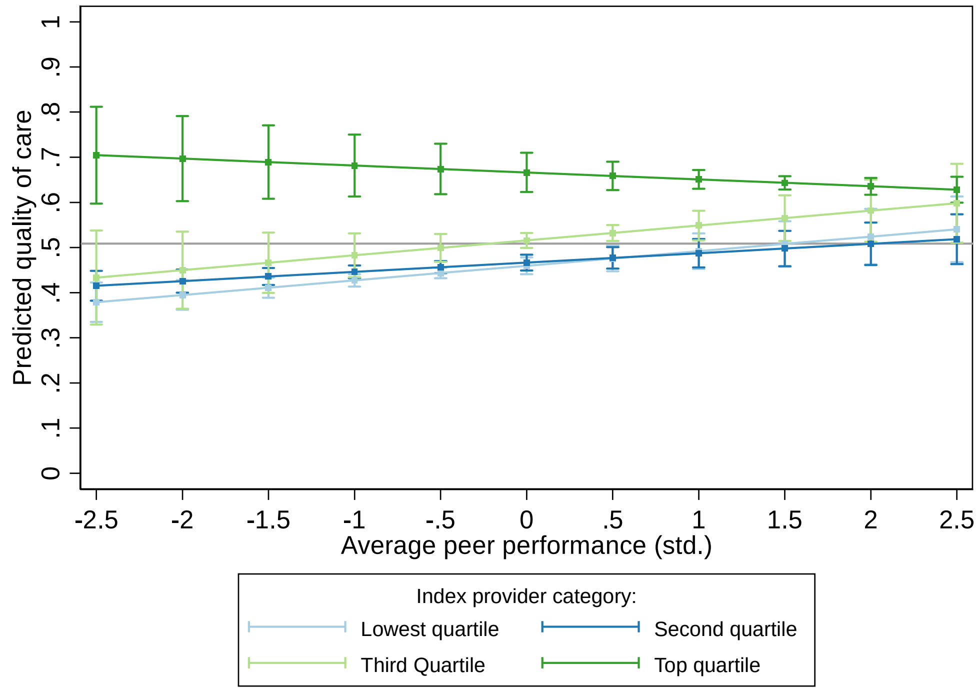
Table 4.3 (Continued)

Notes: Regressions are at delivery-stage level, with stages nested within providers. Delivery complication is neonatal resuscitation, newborn referred to the NICU, or mother treated for post-partum hemorrhage or eclampsia. Higher risk is mother grand multiparous (5 or more births), younger than 18 or older than 35, or has multiple births

Given that each subindex contains between 3 and 7 items, these coefficients translate to less than one additional action completed. However, the coefficients are large in comparison to the risk characteristics: deliveries that had a complication or were from higher risk pregnancies were not more likely to have these routine actions completed than less risky deliveries. The quality of care that women in the wealthiest quintile received was on average 5 percentage points higher than women in the poorest wealth quintile.

The marginal associations of coworker performance by the index provider's capabilities are shown in Figure 4.2. There is no association between coworker performance and quality among the top performing providers: they perform consistently well regardless of their coworkers' performance. However, providers in the middle and low quartiles have large improvement when surrounded by better coworkers. For example, the predicted performance of a third quartile provider when their coworkers' performance is one standard deviation above average is 53% in comparison to 48% when their coworkers are 1 standard deviation below average. Graphs of coworker performance by seniority are included in Appendix Figures C.6 and C.7; there are not strong differences in the associations by either cadre rank or years of experience relative to the index provider.

Figure 4.2 Predicted quality of care by coworker performance and index provider performance



Notes: Points show the predicted estimate of quality by their co-workers' performance for the given level of index provider performance; bars show 95% confidence intervals. Gray line shows mean performance (0.51).

Discussion

Deliveries are often attended by multiple providers who must work together to provide high quality care for the mother and newborn, particularly in larger health facilities. Adopting a small groups as complex systems framework, this study examined the provider group dynamics and their associations with quality of care in Dire Dawa, Ethiopia. This study had four key findings. First, the observed quality of routine labor and delivery was poor, especially for postpartum maternal care that is vital in timely diagnosis of potentially fatal conditions such as postpartum hemorrhage. Second, women in the poorest wealth quintiles received worse care and women with risks or emerging complications did not receive higher quality, even after controlling for facility and provider characteristics. Third, we found that approximately 70% of hospital deliveries and 35% of health center deliveries had more than one provider involved, but more providers were not associated with higher quality of care. Fourth, providers performed better when working with someone who delivered high quality care, though the differences were small.

We hypothesize several mechanisms by which these provider dynamics may be acting, though our study cannot tease these apart directly. First, the positive association between coworker performance and quality indicates that the potential positive spillovers such as knowledge sharing or social pressure, outweigh the negative spillovers such as free riding. Concretely, we found that if a provider is working with a competent colleague, they do not consequently do less for the client. This is consistent with other peer performance literature, which has shown benefits from working with high-performing colleagues [108,113]. The association with coworker performance was strongest in the first stage of labor, a stage which may require more communication and coordination between providers to share information after the first exam and before delivery. Greater interaction and interdependence between providers during the delivery may therefore create stronger associations. This aligns with recent work demonstrating that increasing the duration and intensity of collaboration between providers can reduce patient mortality [116]. Contrary to our hypothesis, quality of care was lower when coworkers were of a superior cadre to the index provider. Given that the lower cadres provide the routine actions more often than their senior counterparts, there may be some reverse accountability in working with a more diligent junior colleague. Other work has also shown that junior cadres in Ethiopia have smaller gaps between their knowledge and practice [117].

This work can inform quality improvement in several ways. First, interventions should emphasize improving immediate postpartum maternal care, which has the largest identified gaps and the greatest potential for reducing severe maternal morbidity and mortality. While the gaps in monitoring vitals and palpating the uterus that we found in Dire Dawa are larger than those seen in other areas of Ethiopia, the immediate postpartum period has previously been identified as a neglected area [118,119]. Indeed, in evaluations of quality improvement interventions in Ethiopia, the mother's postpartum care is often not even included as an outcome in favor of immediate newborn care [120],[121]. The poor quality of care in this study was observed even under the potential biases of the Hawthorne effect which suggests that providers would perform better while others watching. While evidence suggest these effects dissipate quickly, without observers in the room, the providers may provide even lower quality care [122]. Second, the propensity for multiple providers during delivery suggests that health worker education should emphasize team-based models of care. Knowing how to work as a team, assign roles and responsibilities and communicate findings may improve the performance of the team [123]. While our study assessed mainly normal births, highly optimized team-based care will also be required to reduce mortality in the case of obstetric emergencies [124].

Third, improvement interventions may consider management or supervision approaches that pair providers of mixed competency levels. Our analysis shows that only the very top performers do not benefit from their coworkers' performance, while most providers may benefit from working with a high performer. Arranging staffing schedules to intentionally pair higher and lower providers may increase the potential for these positive spillovers and strengthen accountability. Supervision approaches typically use traditional cadre hierarchies to define who may act as a supervisor [125]; however, this work suggests that junior cadres may be able to effectively supervise more senior coworkers. Alternatively, actual performance could be used to identify potential supervisors.

Finally, the culmination of these interactions between providers over many deliveries creates the facility culture and norms for quality of care. Interventions that aim to change facility norms on quality of care, such as group problem solving or quality improvement collaboratives have shown moderate to high effects on health worker practices in meta-analyses [18,19]. For example, an improvement intervention in India that used an integrated training, mentoring and a plan-do-study-act cycle with an emphasis on peer learning improved postpartum monitoring of mothers from 52% to 94% [126]. However, this study also shows the limits of micro-focused interventions. The potential for a two-percentage point improvement from pairing providers will not overcome the enormous quality deficit seen in labor and delivery care in Dire Dawa. Rather, other macro level strategies such as improved pre-service education or redesigning maternal and newborn care service to be provided at CEmONC capable facilities will likely be necessary [12,25].

This study is among the first to examine how multiple providers influence one another to provide quality of care in a low-income healthcare setting. The detailed observation data with information on who attended each stage of delivery is a clear strength of the analysis. There are, however, several limitations to address in future research. First, missing provider attribution for the first stage of delivery required us to make assumptions about who was responsible for the care that was not provided. Second, although we controlled for many potential confounders, the associations may not be interpreted causally. The formation of the provider groups may in some cases be endogenous, for example if a proactive provider actively seeks support from a high-quality provider to assist with a complex delivery. Third, this study only focused on routine actions that should be done for every mother and newborn. However, the quality of complications management is likely both more impactful for consequent health outcomes and often requires multiple providers to work together concurrently rather than sequentially. Future research should examine the coordination and communications of providers working to address complications. Fourth, as the quality scales ranged from 0-100%, there may be floor or ceiling effects. Finally, generalizability of the study may be limited. This study took place during the Covid-19 pandemic in Ethiopia and for a period during the study, one of the two public hospitals in Dira Dawa was closed to deliveries because it was serving as a Covid treatment facility. This resulted in unusually high utilization in the other hospital which may have altered provider dynamics. The high proportion of variance due to client level may have been smaller if more providers or facilities were included in the study.

Over the past decade, approximately 31% of deliveries in Sub-Saharan Africa took place in hospitals and there are calls to increase this proportion [127,128]. As more women deliver in hospitals, they are more likely to be attended by multiple healthcare providers, which could have consequences for the quality of care they receive. Unpacking the provider dynamics in how they work together to deliver care quality can yield useful insights for quality improvement in the future.

Chapter 5 : Conclusion

Key Findings and Implications

This dissertation examined health system quality improvement at the macro, meso, and micro levels through a complex adaptive system lens. The papers demonstrated the dynamic relationships that are involved in producing high quality care, including those between providers caring for the same patient, between nearby facilities and council health teams, and at the system level between utilization and quality in producing health outcomes. Together, they also suggest the importance of relationships between these levels of the health system. The results of the three papers show concerningly low levels of quality and no easy pathways for improvement. However, two key insights the design and implementation of health system quality improvement interventions emerge from this work. First, quality interventions can be purposefully designed to take advantage of feedback loops and dynamic relationships to magnify their potential impact. Second, there are opportunities to leverage different levels of the system to work together to create high quality care.

Harnessing spillovers

In Chapter 3, facilities that were geographically proximate to facilities where the Star Rating score improved—and districts neighboring improving districts—were both more likely to improve. In Chapter 4, providers that worked with higher performing providers were more likely to provide higher quality care to delivering women and their newborns. Both findings were observational rather than causal, yet many potential confounders in the environment were controlled in the analyses. This suggests that positive spillovers may support the provision of high-quality care. Hypothesized mechanisms for these spillovers include social pressure, where facilities or providers are concerned for their reputation; knowledge transfer, where information about evidence-based care or quality improvement may be shared; social comparison, where low-performers attempt to minimize the disparity between themselves and high performers; or explicit supervisory relationships between providers [113]. Explicitly acknowledging and harnessing these positive spillovers into quality improvement interventions could magnify the potential impact of interventions.

There are several ways in which quality improvement interventions could harness these spillovers. For example, in Tanzania, the next round of the Star Rating Assessment is planned to be shared publicly, which may increase the social pressure facilities feel to be seen as high performers. However, some research has found negative unintended consequences of publicly sharing information on quality of care; interventions should take care to minimize negative externalities [129]. In Dire Dawa, interventions such as acknowledging or rewarding high performers may help to leverage social pressure. The district health management teams in Tanzania play a critical role in transferring knowledge and sharing best practices between facilities, yet the findings from Chapter 3 suggest that learning may occur both within and across district boundaries. Interventions could further formalize or strengthen informal connections between facilities through learning collaboratives or similar networks [18]. At the provider level, nursing and medical education curricula should emphasize team-based care models, team communication, and working across different cadres to provide high quality care [123]. Common quality improvement interventions plan-do-study acts could also emphasize peer-learning in order to minimize within-facility variation [126].

Along with the positive spillovers, there is also the potential for negative spillovers, where concentrated groups of poor-performing providers, facilities or districts do not have the pressure, knowledge transfer, or comparison to spur improvement [130]. For example, in Tanzania there were some councils in Ruvuma region where almost no facilities improved on the Star Rating scale. Quality improvement interventions may need to target these groups specifically to prevent self-reinforcing cycles of underachievement, for example by shifting provider schedules to pair high and low performers, bringing in additional support or expertise, or redefining peer reference groups to contain a global best performer rather than a local one.

Improvement across the system

The gaps in the quality of care found in each of the Chapters were formidable. Only half of the recommended evidence-based actions were conducted for delivering mothers and newborns in Dire Dawa; in Tanzania, the Ministry of Health abandoned its original plan to shut down zero-star facilities after the baseline Star Rating Assessment when it found that one third of facilities would need to be closed; and less than half of pregnant women across Burundi, Lesotho, Senegal, Zambia and Zimbabwe received a minimum standard of antenatal care quality prior to the introduction of PBF. These estimates are not necessarily surprising; they broadly align with the findings from the three global reports on quality in 2018 [12–14]. However, they do represent an onerous challenge for improvement. The size of these gaps suggests that small tweaks around the edges, such as rescheduling providers to facilitate collaboration between low performers and high performers, may be beneficial but will not be nearly sufficient to provide high quality care across a health system.

Macro-level interventions that address the foundations of the health system will likely be needed to address these gaps, yet Chapter 2 shows that even large changes such as performance-based financing may fall short. Together, the chapters show that there may be synergies to acting at multiple levels of the health system simultaneously. For example, in Tanzania, participation and eligibility for the results-based financing program was associated with greater improvement on the Star Rating Assessment, suggesting that these two interventions may have complementarities that supported one another. In Dire Dawa, adjustments to provider schedules and team communication interventions may have a greater impact if underlying pre-service education was reformed to place a greater emphasis on postpartum monitoring for mothers. Complex adaptive system theory suggests that more feedback loops and particularly more positive reinforcing loops than balancing loops is important to jointly create the conditions for high performance. In order to enact new quality improvement interventions such as redesigning where maternal and neonatal care services are delivered in the system [128], changes at the macro, meso and micro levels will be required.

Future Research

Several areas of future research emerge from this work. First, the goal of quality improvement interventions should be not just to improve process measures, but to have impacts on population health, satisfaction, or efficiency. This dissertation reveals the necessity for researchers to measure these ultimate outcomes directly whenever possible rather than relying on intermediate outcomes given the non-linearity of causal mechanisms for valued outcomes. Chapter 2 found that while there were some mixed impacts of performance-based financing on quality and utilization of care, together these did not produce better health outcomes. While Chapters 3 and 4 showed some evidence of positive spillovers, there may also negative spillovers or unintended consequences that affect the eventual outcomes in ways that intermediate outcomes are not able to completely explain. A priority research question following Chapter 3, for example, is whether improvement on the Star Rating Assessment scale is associated with better health outcomes in the community surrounding a facility. Studying downstream outcomes, particularly health outcomes, can be costly for researchers because they often require large samples and long follow up periods. However, secondary data sources including those from routine health management information systems and DHS could be used with quasi-experimental study designs such as difference-in-differences or interrupted time-series designs to assess these outcomes.

Second, to further support health system improvement, researchers should move beyond measuring utilization or structural quality and emphasize health system function—i.e., how health systems treat patients. In the DHS used in Chapter 2, the quality measures assessed were limited to a small set of routine care indicators. While DHS is limited by self-report of quality and long recall periods, incorporating additional questions on receipt of evidence-based care and user experience would be useful, for example on skin-to-skin contact with the newborn and the provider's communication during delivery. The Star Rating Assessment in Chapter 3 has many more items related to inputs such as equipment and supplies, which may overemphasize its importance even with the relative weighting for each domain. In data feedback interventions, more weight could be placed on processes of care and client perspectives, which are likely more related to health outcomes. And, while Chapter 4 utilizes very detailed data on processes of care from observations of deliveries, this type of primary data collection is very intensive and can only be done for a short period in small geographic areas. Efforts to include more quality-of-care measures in routine health information systems and eventually moving to patient-level electronic health records would support population-wide quality measures at a lower cost [25].

Finally, future healthcare quality improvement research would benefit from further implementation science methods as well as mixed quantitative and qualitative methods to better understand the dynamic relationships at play. For example, qualitative research is forthcoming on the Star Rating Assessment which explores how facility managers interpreted their baseline Star Rating score, and whether and how they decided whether to act on the information provided. This will provide further insights for how to support facilities in improvement in future rounds of the Star Rating Assessment. As many countries continue with performance-based financing programs, more implementation research is also needed to understand the specific barriers to effective implementation of the program, particularly in the barriers to improved quality of care. Finally, as communities, Ministries of Health, and global organizations explore new interventions for improving quality such as service delivery redesign, this combination of health outcome evaluation and implementation research will be needed to understand the consequences for better quality care and better health.

References

1. Asma S, Jakab Z. Tracking universal health coverage 2021 Global Monitoring Report [Internet]. World Health Organization, World Bank; 2021. Available from: https://www.who.int/data/monitoring-universal-health-coverage

2. Croke K, Mengistu AT, O'Connell SD, Tafere K. The impact of a health facility construction campaign on health service utilisation and outcomes: analysis of spatially linked survey and facility location data in Ethiopia. BMJ global health. 2020;5(8):e002430.

3. Berger J, Calmy A, Moon S. Driving a decade of change: HIV/AIDS, patents and access to medicines for all. Journal of the International AIDS Society. 2011;14(1):1–12.

4. Ifeagwu SC, Yang JC, Parkes-Ratanshi R, Brayne C. Health financing for universal health coverage in Sub-Saharan Africa: a systematic review. Global health research and policy. 2021;6(1):1–9.

5. Ethiopian Public Health Institute. Ethiopia Mini Demographic and Health Survey [Internet]. Addis Ababa; 2021 [cited 2021 Nov 4]. Available from: https://dhsprogram.com/pubs/pdf/FR363/FR363.pdf

6. Boerma T, Requejo J, Victora CG, Amouzou A, George A, Agyepong I, et al. Countdown to 2030: tracking progress towards universal coverage for reproductive, maternal, newborn, and child health. The Lancet. 2018;391(10129):1538–48.

7. Geldsetzer P, Manne-Goehler J, Marcus M-E, Ebert C, Zhumadilov Z, Wesseh CS, et al. The state of hypertension care in 44 low-income and middle-income countries: a cross-sectional study of nationally representative individual-level data from 1· 1 million adults. The Lancet. 2019;394(10199):652–62.

8. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1204–22.

9. Wang H, Abbas KM, Abbasifard M, Abbasi-Kangevari M, Abbastabar H, Abd-Allah F, et al. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1160–203.

10. Kruk ME, Gage AD, Joseph N, Danei G, Garcia-Saiso S, Salomon JA. Mortality due to low quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. The Lancet. 2018;392(10160):2203–12.

11. Roder-DeWan S, Gage A, Hirschhorn LR, Twum-Danso NA, Liljestrand J, Asante-Shongwe K, et al. Level of confidence in and endorsement of the health system among

internet users in 12 low-income and middle-income countries. BMJ global health. 2020;5(8):e002205.

12. Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. The Lancet Global Health. 2018;6(11):e1196–252.

13. National Academies of Sciences Engineering and Medicine. Crossing the global quality chasm: Improving health care worldwide. Washington, DC; 2018.

14. World Health Organization O for EC and D and The World Bank,. Delivering quality health services: a global imperative for universal health coverage. Geneva; 2018.

15. Packard RM. Post-colonial medicine. In: Medicine in the twentieth century. Taylor & Francis; 2020. p. 97–112.

16. Sifrim ZK, Barker PM, Mate KS. What gets published: the characteristics of quality improvement research articles from low-and middle-income countries. BMJ quality & safety. 2012;21(5):423–31.

17. Leslie HH, Gage A, Nsona H, Hirschhorn LR, Kruk ME. Training And Supervision Did Not Meaningfully Improve Quality Of Care For Pregnant Women Or Sick Children In Sub-Saharan Africa. Health affairs (Project Hope). 2016;35(9):1716.

18. Garcia-Elorrio E, Rowe SY, Teijeiro ME, Ciapponi A, Rowe AK. The effectiveness of the quality improvement collaborative strategy in low- and middle-income countries: A systematic review and meta-analysis. PloS one. 2019;14(10):e0221919.

19. Rowe AK, Rowe SY, Peters DH, Holloway KA, Chalker J, Ross-Degnan D. Effectiveness of strategies to improve health-care provider practices in low-income and middle-income countries: a systematic review. The Lancet Global Health. 2018;6(11):e1163–75.

20. Edward A, Peters DH, Daniels A, Rang Y, Matsubayashi T. Review of strategies to strengthen the performance of health organizations. Improving health service delivery in developing countries: from evidence to action. 2009;69–99.

21. Althabe F, Bergel E, Cafferata ML, Gibbons L, Ciapponi A, Alemán A, et al. Strategies for improving the quality of health care in maternal and child health in low- and middle-income countries: an overview of systematic reviews. Paediatr Perinat Epidemiol [Internet]. 2008;22. Available from: https://doi.org/10.1111/j.1365-3016.2007.00912.x http://dx.doi.org/10.1111/j.1365-3016.2007.00912.x

22. Arsenault C, Rowe SY, Ross-Degnan D, Peters DH, Roder-DeWan S, Kruk ME, et al. How does the effectiveness of strategies to improve healthcare provider practices in low-income and middle-income countries change after implementation? Secondary analysis of a systematic review. BMJ Quality & Safety. 2021;

23. Björkman M, Svensson J. Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda. Q J Econ. 2009;124(2):735–69.

24. Semrau KEA, Hirschhorn LR, Marx Delaney M, Singh VP, Saurastri R, Sharma N, et al. Outcomes of a Coaching-Based WHO Safe Childbirth Checklist Program in India. New England Journal of Medicine. 2017 Dec 14;377(24):2313–24.

25. Nimako K, Kruk ME. Seizing the moment to rethink health systems. The Lancet Global Health. 2021;

26. De Savigny D, Adam T. Systems Thinking For Health Systems Strengthening. Geneva: World Health Organization; 2009.

27. Paina L, Peters DH. Understanding pathways for scaling up health services through the lens of complex adaptive systems. Health policy and planning. 2012;27(5):365–73.

28. Alonge O, Rodriguez DC, Brandes N, Geng E, Reveiz L, Peters DH. How is implementation research applied to advance health in low-income and middle-income countries? BMJ Global Health. 2019;4(2):e001257.

29. Dintrans PV, Bossert TJ, Sherry J, Kruk ME. A synthesis of implementation science frameworks and application to global health gaps. Global health research and policy. 2019;4(1):1–11.

30. Stokes T, Shaw EJ, Camosso-Stefinovic J, Imamura M, Kanguru L, Hussein J. Barriers and enablers to guideline implementation strategies to improve obstetric care practice in low- and middle-income countries: a systematic review of qualitative evidence. Implementation Science. 2016;11(1):1–10.

31. Kruk ME. More health for the money—toward a more rigorous implementation science. Vol. 6, Science translational medicine. American Association for the Advancement of Science; 2014. p. 245ed17-245ed17.

32. Atun R. Health systems, systems thinking and innovation. Health policy and planning. 2012;27(suppl_4):iv4–8.

33. Newton-Lewis T, Munar W, Chanturidze T. Performance management in complex adaptive systems: a conceptual framework for health systems. BMJ Global Health. 2021;6(7):e005582.

34. Diaconu K, Falconer J, Verbel A, Fretheim A, Witter S. Paying for performance to improve the delivery of health interventions in low-and middle-income countries. Cochrane Database of Systematic Reviews. 2021;(5).

35. Yahya T, Mohamed M. Raising a mirror to quality of care in Tanzania: the five-star assessment. The Lancet Global Health. 2018;

36. Gülmezoglu AM, Lawrie TA, Hezelgrave N, Oladapo OT, Souza JP, Gielen M, et al. Interventions to reduce maternal and newborn morbidity and mortality. In: Disease Control Priorities 3. 2016.

37. World Health Organization. Newborns: reducing mortality [Internet]. World Health Organization; 2019 [cited 2020 May 28]. (Fact sheets). Available from: https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality

38. Gergen J, Josephson E, Coe M, Ski S, Madhavan S, Bauhoff S. Quality of Care in Performance-Based Financing: How It Is Incorporated in 32 Programs Across 28 Countries. Glob Health Sci Pract. 2017 Mar 24;5(1):90–107.

39. Kandpal E. Completed Impact Evaluations and Emerging Lessons from the Health Results Innovation Trust Fund Learning Portfolio [Internet]. 2016 [cited 2020 Jun 4]. Available from: https://www.rbfhealth.org/resource/completed-impact-evaluations-and-emerging-lessons-health-results-innovation-trust-fund

40. Renmans D, Holvoet N, Criel B, Meessen B. Performance-based financing: the same is different. Health Policy Plan. 2017;32(6):860–8.

41. Lemière C, Torsvik G, Mæstad O, Herbst CH, Leonard KL. Evaluating the Impact of Results-Based Financing on Health Worker Performance. The World Bank; 2013. (HNP Discussion Paper).

42. Engineer CY, Dale E, Agarwal A, Agarwal A, Alonge O, Edward A, et al. Effectiveness of a pay-for-performance intervention to improve maternal and child health services in Afghanistan: a cluster-randomized trial. International journal of epidemiology. 2016;45(2):451–9.

43. Health Results Innovation Trust Fund. Learning Agenda for Results-Based Financing in the Health Sector The Health Results Innovation Trust Fund Learning Strategy [Internet]. World Bank; 2016 [cited 2017 Mar 26]. Available from: https://www.rbfhealth.org/sites/rbf/files/The%20Health%20Results%20Innovation%20Trust%20Fund%20Learning%20Strategy.pdf

44. Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. Lancet. 2011;377(9775):1421–8.

45. Witter S, Fretheim A, Kessy FL, Lindahl AK. Paying for performance to improve the delivery of health interventions in low and middleincome countries. Cochrane Effective Practice and Organisation of Care Group. 2012;(2).

46. Zeng W, Shepard DS, Nguyen H, Chansa C, Das AK, Qamruddin J, et al. Cost–effectiveness of results-based financing, Zambia: a cluster randomized trial. Bulletin of the World Health Organization. 2018;96(11):760.

47. Chinkhumba J, De Allegri M, Brenner S, Muula A, Robberstad B. The cost-effectiveness of using results-based financing to reduce maternal and perinatal mortality in Malawi. BMJ Global Health. 2020;5(5):e002260.

48. Fink G, Ross R, Hill K. Institutional deliveries weakly associated with improved neonatal survival in developing countries: evidence from 192 Demographic and Health Surveys. International journal of epidemiology. 2015;44(6):1879–88.

49. Lim SS, Dandona L, Hoisington JA, James SL, Hogan MC, Gakidou E. India's Janani Suraksha Yojana, a conditional cash transfer programme to increase births in health facilities: an impact evaluation. Lancet. 2010;375(9730):2009–23.

50. Semrau M, Evans-Lacko S, Alem A, Ayuso-Mateos JL, Chisholm D, Gureje O, et al. Strengthening mental health systems in low- and middle-income countries: the Emerald programme. BMC Med. 2015 Apr 10;13:79.

51. Eichler R, Agarwal K, Askew I, Iriarte E, Morgan L, Watson J. Performance-based incentives to improve health status of mothers and newborns: what does the evidence show? Journal of health, population, and nutrition. 2013;31(4 Suppl 2):S36.

52. Turcotte-Tremblay A-M, Spagnolo J, De Allegri M, Ridde V. Does performance-based financing increase value for money in low- and middle- income countries? A systematic review. Health Econ Rev. 2016;6(1):1–18.

53. Sherry TB. A Note on the Comparative Statics of Pay-for-Performance in Health Care. Health economics. 2016;25(5):637–44.

54. Sherry TB, Bauhoff S, Mohanan M. Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda. American Journal of Health Economics. 2017;

55. Bonfrer I, Van de Poel E, Van Doorslaer E. The effects of performance incentives on the utilization and quality of maternal and child care in Burundi. Soc Sci Med. 2014 Dec 1;123:96–104.

56. Friedman J, Das A, Mutasa R. Rewarding Provider Performance to Improve Quality and Coverage of Maternal and Child Health Outcomes: Zimbabwe Results-Based Financing Pilot Program. The World Bank; 2016.

57. Friedman J, Qamruddin J, Chansa C, Das AK. Impact evaluation of Zambia's health results-based financing pilot project. Washington, DC: World Bank Group. 2016;

58. Oxman AD, Fretheim A. Can paying for results help to achieve the Millennium Development Goals? Overview of the effectiveness of results-based financing. Journal of Evidence-Based Medicine. 2009;2(2):70–83.

59. Burgert CR, Colston J, Roy T, Zachary B. Geographic Displacement Procedure and Georeferenced Data Release Policy fro the Demographic and Health Surveys [Internet].

Calverton, Maryland: ICF International; 2013 [cited 2020 Mar 16]. Available from: https://dhsprogram.com/pubs/pdf/SAR7/SAR7.pdf

60. Falisse J-B, Ndayishimiye J, Kamenyero V, Bossuyt M. Performance-based financing in the context of selective free health-care: an evaluation of its effects on the use of primary health-care services in Burundi using routine data. Health policy and planning. 2015;30(10):1251–60.

61. The World Bank. Baseline Study for the Maternal and Newborn Health Performance Based Financing Project in Lesotho [Internet]. 2017 [cited 2020 Mar 16]. Available from: http://documents.worldbank.org/curated/en/661841492598398927/Baseline-study-for-the-maternal-and-newborn-health-performance-based-financing-project-in-Lesotho

62. Leslie HH, Fink G, Nsona H, Kruk ME. Obstetric facility quality and newborn mortality in Malawi: a cross-sectional study. PLoS Medicine. 2016;13(10):e1002151.

63. Coria-Soto IL, Bobadilla JL, Notzon F. The effectiveness of antenatal care in preventing intrauterine growth retardation and low birth weight due to preterm delivery. Oxford University Press; 1996.

64. Katz J, Lee AC, Kozuki N, Lawn JE, Cousens S, Blencowe H, et al. Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. The Lancet. 2013;382(9890):417–25.

65. McClure CK, Bodnar LM, Ness R, Catov JM. Accuracy of maternal recall of gestational weight gain 4 to 12 years after delivery. Obesity. 2011;19(5):1047–53.

66. Chen B, Halliday TJ, Fan VY. The impact of internal displacement on child mortality in post-earthquake Haiti: a difference-in-differences analysis. International journal for equity in health. 2016;15(1):114.

67. Blackwell M, Iacus S, King G, Porro G. cem: Coarsened exact matching in Stata. The Stata Journal. 2009;9(4):524–46.

68. Poirier MJ, Grépin KA, Grignon M. Approaches and alternatives to the wealth index to measure socioeconomic status using survey data: a critical interpretive synthesis. Social Indicators Research. 2020;148(1):1–46.

69. Moyo I, Gandidzanwa C, Tsikiria T, Mabhena T, Dieleman M, Kane S. Learning from implementation. Process Monitoring and Evaluation II of Zimbabwe's Results-Based Financing Project: The case of Mutoko, Chiredzi, Nkayi and Kariba Districts [Internet]. The World Bank; 2015 Mar [cited 2020 Oct 28]. Available from: https://www.rbfhealth.org/sites/rbf/files/Learning%20from%20Implementation%20-%20Process%20Monitoring%20and%20Evaluation%20II%20of%20Zimbabwe%E2%80%99s%20Results-Based%20Financing%20Project.pdf

70. Kruk M, Gage A, Arsenault C, Jordan K, Leslie H, Roder-DeWan S, et al. High quality health systems—time for a revolution: Report of the Lancet Global Health Commission on High Quality Health Systems in the SDG Era. Lancet Global Health. 2018;

71. Lohmann J, Muula AS, Houlfort N, De Allegri M. How does performance-based financing affect health workers' intrinsic motivation? A Self-Determination Theory-based mixed-methods study in Malawi. Social Science & Medicine. 2018;208:1–8.

72. Shen GC, Nguyen HTH, Das A, Sachingongu N, Chansa C, Qamruddin J, et al. Incentives to change: effects of performance-based financing on health workers in Zambia. Human resources for health. 2017;15(1):20.

73. Leslie HH, Sun Z, Kruk ME. Association between infrastructure and observed quality of care in 4 healthcare services: A cross-sectional study of 4,300 facilities in 8 countries. PLoS Med. 2017;14(12):e1002464.

74. Magge H, Nahimana E, Mugunga JC, Nkikabahizi F, Tadiri E, Sayinzoga F, et al. The All Babies Count Initiative: Impact of a Health System Improvement Approach on Neonatal Care and Outcomes in Rwanda. Global Health: Science and Practice. 2020;8(3):505–17.

75. Rajkotia Y, Zang O, Nguimkeu P, Gergen J, Djurovic I, Vaz P, et al. The effect of a performance-based financing program on HIV and maternal/child health services in Mozambique—an impact evaluation. Health Policy Plan. 2017;32(10):1386–96.

76. Ridde V, Yaogo M, Zongo S, Somé P-A, Turcotte-Tremblay A-M. Twelve months of implementation of health care performance-based financing in B urkina F aso: A qualitative multiple case study. The International journal of health planning and management. 2018;33(1):e153–67.

77. Fritsche G, Peabody J. Methods to improve quality performance at scale in lower-and middle-income countries. Journal of global health. 2018;8(2).

78. Paul E, Albert L, Bisala BN, Bodson O, Bonnet E, Bossyns P, et al. Performance-based financing in low-income and middle-income countries: isn't it time for a rethink? BMJ global health. 2018;3(1):e000664.

79. World Health Organization and UNICEF. Declaration of Astana [Internet]. Global Conference on Primary Health Care; 2018 [cited 2020 Feb 10]. Available from: https://www.who.int/docs/default-source/primary-health/declaration/gcphc-declaration.pdf

80. Adam T, de Savigny D. Systems thinking for strengthening health systems in LMICs: need for a paradigm shift. Health Policy Plan. 2012;27(suppl4):iv1–3.

81. United Republic of Tanzania. Tanzania Development Vision 2025 - Big Results Now: BRN Healthcare NKRA Lab Report – Part I [Internet]. 2014. Available from: http://www.tzdpg.or.tz/fileadmin/documents/dpg_internal/dpg_working_groups_clusters/cluster_2/health/Sub_Sector_Group/BRN_documents/Tanz_Healthcare_Lab_Report_Part_1_0212_RH_-_v21__Final_Lab_Report_.pdf

82. Ministry of Health and Social Welfare. Health Sector Strategic Plan July 2015 – June 2020 (HSSP IV). Reaching all Households with Quality Health Care [Internet]. Dar Es Salaam; 2015 [cited 2020 Feb 28]. Available from: http://www.tzdpg.or.tz/fileadmin/documents/dpg_internal/dpg_working_groups_clusters/cluster_2/health/Key_Sector_Documents/Induction_Pack/Final_HSSP_IV_Vs1.0_260815.pdf

83. Boex J, Fuller F, Malik A. Decentralized Local Health Services inTanzania [Internet]. Urban Institute; 2015 [cited 2020 Feb 28]. Available from: https://www.urban.org/sites/default/files/publication/51206/2000215-Decentralized-Local-Health-Services-in-Tanzania.pdf

84. Kapologwe NA, Kalolo A, Kibusi SM, Chaula Z, Nswilla A, Teuscher T, et al. Understanding the implementation of Direct Health Facility Financing and its effect on health system performance in Tanzania: a non-controlled before and after mixed method study protocol. Health research policy and systems. 2019;17(1):11.

85. Chimhutu V, Tjomsland M, Mrisho M. Experiences of care in the context of payment for performance (P4P) in Tanzania. Globalization and Health. 2019;15(1):59.

86. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation science. 2009;4(1):50.

87. Nilsen P. Making sense of implementation theories, models and frameworks. Implementation science. 2015;10(1):53.

88. Arcaya M, Brewster M, Zigler CM, Subramanian S. Area variations in health: A spatial multilevel modeling approach. Health & place. 2012;18(4):824–31.

89. Yahya T, Eliakimu E, Mohamed M. Tanzania Star Rating Assessment Tools [Internet]. Health Quality Assurance Division, Tanzania Ministry of Health; 2020 [cited 2020 Jul 1]. Available from: https://figshare.com/s/cb90b0c9c30dae9587ac

90. McCrary J. Manipulation of the running variable in the regression discontinuity design: A density test. Journal of econometrics. 2008;142(2):698–714.

91. WorldPop. The spatial distribution of population in 2015, Tanzania [Internet]. Project GHRPD, editor. 2018. Available from: www.worldpop.org

92. Datta GS. Model-based approach to small area estimation. In: Handbook of Statistics. Elsevier; 2009. p. 251–88.

93. Commonwealth Local Government Forum. The Local Government System in Tanzania [Internet]. 2018 [cited 2020 Jan 23]. Available from: http://www.clgf.org.uk/default/assets/File/Country_profiles/Tanzania.pdf

94. Griffith D. Spatial Autocorrelation: A Primer. Resource Publications in Geograph, Association of American geographers; 1987.

95. Dong G, Harris R. Spatial autoregressive models for geographically hierarchical data structures. Geographical Analysis. 2015;47(2):173–91.

96. Sochas L. The predictive power of health system environments: a novel approach for explaining inequalities in access to maternal healthcare. BMJ Global Health. 2020;4(Suppl 5).

97. Roder-DeWan S, Gage AD, Hirschhorn LR, Twum-Danso NA, Liljestrand J, Asante-Shongwe K, et al. Expectations of healthcare quality: A cross-sectional study of internet users in 12 low-and middle-income countries. PLoS medicine. 2019;16(8):e1002879.

98. World Bank. World Development Indicators [Internet]. Bank TW, editor. 2017 [cited 2018 Feb 10]. Available from: http://databank.worldbank.org/data/reports.aspx?source=World-Development-Indicators#

99. Tessema GA, Laurence CO, Melaku YA, Misganaw A, Woldie SA, Hiruye A, et al. Trends and causes of maternal mortality in Ethiopia during 1990-2013: findings from the Global Burden of Diseases study 2013. BMC public health. 2017 Feb 2;17(1):160.

100. Berhan Y, Berhan A. Perinatal mortality trends in Ethiopia. Ethiopian journal of health sciences. 2014;24:29–40.

101. Weldearegay HG, Kahsay AB, Medhanyie AA, Godefay H, Petrucka P. Quality of and barriers to routine childbirth care signal functions in primary level facilities of Tigray, Northern Ethiopia: Mixed method study. PloS one. 2020;15(6):e0234318.

102. Rokicki S, Mwesigwa B, Cohen JL. Know-do gaps in obstetric and newborn care quality in Uganda: a cross-sectional study in rural health facilities. Tropical Medicine & International Health. 2021;26(5):535–45.

103. Arsenault C, English M, Gathara D, Malata A, Mandala W, Kruk ME. Variation in competent and respectful delivery care in Kenya and Malawi: a retrospective analysis of national facility surveys. Tropical Medicine & International Health. 2020;25(4):442–53.

104. Helfinstein S, Jain M, Ramesh BM, Blanchard J, Kemp H, Gothalwal V, et al. Facilities are substantially more influential than care providers in the quality of delivery care received: a variance decomposition and clustering analysis in Kenya, Malawi and India. BMJ global health. 2020;5(8):e002437.

105. Owili PO, Muga MA, Mendez BR, Chen B. Quality of maternity care and its determinants along the continuum in Kenya: A structural equation modeling analysis. PLoS One. 2017;12(5):e0177756.

106. Mach M, Abrantes AC, Soler C. Teamwork in Healthcare Management. Teamwork in Healthcare. 2021;23.

107. Arrow H, McGrath JE, Berdahl JL. Small groups as complex systems: Formation, coordination, development, and adaptation. Sage Publications; 2000.

108. Stecher C. Physician network connections to specialists and HIV quality of care. Health Services Research. 2021;

109. Nair HS, Manchanda P, Bhatia T. Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. Journal of Marketing Research. 2010;47(5):883–95.

110. Schmutz JB, Meier LL, Manser T. How effective is teamwork really? The relationship between teamwork and performance in healthcare teams: a systematic review and meta-analysis. BMJ open. 2019;9(9):e028280.

111. Rosen MA, DiazGranados D, Dietz AS, Benishek LE, Thompson D, Pronovost PJ, et al. Teamwork in healthcare: Key discoveries enabling safer, high-quality care. American Psychologist. 2018;73(4):433.

112. Gharaveis A, Hamilton DK, Pati D. The impact of environmental design on teamwork and communication in healthcare facilities: a systematic literature review. HERD: Health Environments Research & Design Journal. 2018;11(1):119–37.

113. Tan TF, Netessine S. When you work with a superman, will you also fly? an empirical study of the impact of coworkers on performance. Management Science. 2019;65(8):3495–517.

114. UNICEF. Situation Analysis of Children and Women: Dire Dawa Administration [Internet]. 2019 Dec [cited 2021 Sep 29]. Available from: https://www.unicef.org/ethiopia/media/2556/file/Dire%20Dawa.pdf

115. Tripathi V, Stanton C, Strobino D, Bartlett L. Development and Validation of an Index to Measure the Quality of Facility-Based Labor and Delivery Care Processes in Sub-Saharan Africa. PLoS One. 2015;10(6):e0129491.

116. Chen Y. Team-specific human capital and team performance: Evidence from doctors. American Economic Review. 2019;

117. Gage AD, Kruk ME, Girma T, Lemango ET. The know-do gap in sick child care in Ethiopia. PLoS One. 2018;13(12):e0208898.

118. Kebede AA, Taye BT, Wondie KY, Tiguh AE, Eriku GA, Mihret MS. Healthcare provider's adherence to immediate postpartum care guidelines in Gondar province hospitals, northwest Ethiopia: A multicenter study. Plos one. 2021;16(10):e0259263.

119. Fisseha G, Berhane Y, Worku A, Terefe W. Quality of the delivery services in health facilities in Northern Ethiopia. BMC health services research. 2017;17(1):1–7.

120. Magge H, Kiflie A, Nimako K, Brooks K, Sodzi-Tettey S, Mobisson-Etuk N, et al. The Ethiopia healthcare quality initiative: design and initial lessons learned. International Journal for Quality in Health Care. 2019;31(10):G180–6.

121. Hagaman AK, Singh K, Abate M, Alemu H, Kefale AB, Bitewulign B, et al. The impacts of quality improvement on maternal and newborn health: preliminary findings from a health system integrated intervention in four Ethiopian regions. BMC health services research. 2020;20:1–12.

122. Leonard K, Masatu MC. Outpatient process quality evaluation and the Hawthorne Effect. Social science & medicine. 2006;63(9):2330–40.

123. Fransen AF, de Ven J, Banga FR, Mol BWJ, Oei SG. Multi-professional simulation-based team training in obstetric emergencies for improving patient outcomes and trainees' performance. Cochrane Database of Systematic Reviews. 2020;(12).

124. Dalby P, Gosman GG, Stein K, Streitman D, Wise N. Crisis teams for obstetric patients. Textbook of Rapid Response Systems. 2017;229–39.

125. Austin A, Gulema H, Belizan M, Colaci DS, Kendall T, Tebeka M, et al. Barriers to providing quality emergency obstetric care in Addis Ababa, Ethiopia: healthcare providers' perspectives on training, referrals and supervision, a mixed methods study. BMC pregnancy and childbirth. 2015;15(1):1–10.

126. Taneja G, Sarin E, Bajpayee D, Chaudhuri S, Verma G, Parashar R, et al. Care Around Birth Approach: A Training, Mentoring, and Quality Improvement Model to Optimize Intrapartum and Immediate Postpartum Quality of Care in India. Global Health: Science and Practice. 2021;9(3):590–610.

127. Gage AD, Fink G, Ataguba JE, Kruk ME. Hospital delivery and neonatal mortality in 37 countries in sub-Saharan Africa and South Asia: An ecological study. PLOS Medicine. 2021 Dec 1;18(12):e1003843.

128. Roder-DeWan S, Nimako K, Twum-Danso NA, Amatya A, Langer A, Kruk M. Health system redesign for maternal and newborn survival: rethinking care models to close the global equity gap. BMJ Global Health. 2020;5(10):e002539.

129. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. Jama. 2005;293(10):1239–44.

130. Cumbler E, Pierce R. The search for perpetual motion: fatigue, friction, and drag in quality improvement. Vol. 26, Quality Management in Healthcare. LWW; 2017. p. 129–30.

131. Josephson E, Gergen J, Coe M, Ski S, Madhavan S, Bauhoff S. How do performance-based financing programmes measure quality of care? A descriptive analysis of 68 quality checklists from 28 low- and middle-income countries. Health Policy Plan. 2017 May 26;32(8):1120–6.

132. Manuel des Procedures Pour la Mise en Oeuvre du Financement Base sur la Performance Seconde Generation [Internet]. Ministere de la Sante Publique et de la Lutte Contre le Sida, Republique du Burundi; 2017 Jan [cited 2020 Nov 3]. Report No.: Version 3. Available from: http://www.fbpsanteburundi.bi/cside/contents/docs/Manuel_des_procedures_FBP_seconde_generation.pdf

133. Lesotho Ministry of Health. Maternal and Newborn Health Performance Based Financing Project User Guide Manual [Internet]. 2016. Available from: https://www.rbfhealth.org/sites/rbf/files/documents/Lesotho-Maternal-Newborn-Health-PBF-Project-User-Manual.pdf

134. RBF Health. Senegal [Internet]. 2017. Available from: https://www.rbfhealth.org/rbfhealth/country/senegal

135. El-Khoury M, Faye S, Baruwa E. Results-Based Financing, Senegal: A look inside the ‘black box.’ In Rabat, Morocco; 2016. Available from: https://afhea.org/docs/presetationspdfs/Sophie%20-%20Results-Based%20Financing,%20Senegal-%20%20A%20look%20inside%20the%20%E2%80%98black%20box%E2%80%99.pdf

Appendix

Appendix A: Chapter 2 Supplementary materials

Text A.1. PBF implementation details

The PBF projects differed in their design and implementation across the five study countries. This appendix provides more detailed information about the design and implementation in each country. First, we provide an overview of the types of indicators that were incentivized by the PBF programs. Second, we include descriptions of each project sourced from published documents.

Incentivized indicators

We examined all incentivized indicators in the PBF study countries from the TRAction database [131] and categorized them into structural quality, process quality, or coverage indicators. Indicators were included if they were directly related to antenatal, labor and delivery, PMTCT or postnatal services, or if they were service agnostic but were applicable to maternal services (i.e., availability of electricity or handwashing). A total of 259 unique indicators were classified.

Structural quality indicators were further categorized into infection prevention items (i.e., sterile gloves); key equipment (i.e. baby weighing scale); medications (i.e. local anesthesia available); patient amenities (i.e. curtain between delivery bed and door); human resources (i.e. No absence of staff for unjustified reasons during last 3 months); record keeping (i.e. Admission card correctly filled out) and fees (i.e. delivery fee). Process quality was further categorized into distinct services: ANC (i.e., weight, blood pressure, breast exam and check for edema completed during physical examination); labor and delivery (i.e. APGAR noted during the 1st, 5th and 10th minutes); PMTCT (i.e. Proper monitoring of infants born to HIV positive mothers) and postnatal care (i.e. Postnatal consultations compliance with quality standards). Coverage indicators included indicators such as rate of women having postnatal consultation and number of uncomplicated deliveries.

Number of incentivized indicators by category

Indicator type	Burundi	Lesotho	Senegal	Zambia	Zimbabwe
Structural quality					
Infection prevention items	8	15	3	0	0
Key equipment	34	18	5	12	7
Medications	5	8	0	2	1
Patient amenities	2	8	6	0	0
Human resources	3	3	5	0	0
Record keeping	4	13	4	1	0
Fees	1	1	0	0	4
Process quality					
ANC process	2	14	3	14	3
Delivery Process	6	6	7	3	5
PMTCT Process	5	1	6	2	1
Postnatal Process	0	0	2	0	1
Coverage	13	14	5	9	11

The programs varied in the number and type of indicators. Zimbabwe's program had 33 incentivized indicators relevant to maternity care while Lesotho's had 101. Programs with more incentivized indicators tended to have a larger proportion in structural quality, while Zimbabwe's were equally split between structural quality, process quality and coverage measures.

Implementation and control districts

	First implementation districts	Second implementation districts	Control districts	Unconditional financing districts
Burundi	Bubanza, Cankuzo and Gitega	Makamba and Bururi	Karuzi, Rutana, Ruyigi, Ngozi, Kirundo	N/A

Lesotho	Mokhotlong and Thaba-Tseka	Mafeteng and Mohale's Hoek	Botha-Bothe, Berea, Maseru and Qacha's-nek	N/A
Senegal	Kaffrine and Kolda	N/A	Tambacounda, Sédhiou, Kédougou and Ziguinchor	N/A
Zambia	Mumbwa, Lufwanyama, Lundazi, Mwense, Mporokoso, Isoka, Mufumbwe, Siavonga, Gwembe and Senanga	N/A	Chadiza, Chavuma, Chibombo, Chinsali, Kazungula, Mpongwe, Mazabuka, Milenge, Mpulungu, and Shangombo	Chilubi, Itezhi-Tezhi, Kalabo, Kapiri Mposhi, Kawambwa, Masaiti, Mwinilunga, Nakonde, Namwala and Nyimba
Zimbabwe	Binga, Centenary, Chegutu, Chikomba, Chipinge, Chiredzi, Gokwe South, Gwanda, Gweru Rural, Kariba Nyaminyami, Manangwe, Mazowe, Mutare, Mutoke, Mwenezi and Nkayi	N/A	Bikita, Bindura, Bubi, Chirumhanzu, Hurungwe, Kadoma/Sanyati, Makoni, Matobo, Mt. Darwin, Nyanga, Shurugwi, Umguza, UMP, Umzingwane, Hwedza and Zaka	N/A

PBF implementation in Burundi

Burundi’s PBF program was rolled out in three phases. Bubanza, Cankuzo and Gitega provinces implemented the scheme in December 2006, Makamba and Burui provinces implemented in October 2008, and the remaining provinces implemented in April 2010. This study only includes the first two phases of implementation in order to use the remaining provinces as control areas.

As of 2014, performance-based financing accounts for 40% of the total average health facility budget [55]. Facilities receive payments based on the quantity and quality of health services provided. Quantity is measured through twenty-three output indicators. Health care facilities report monthly to the Ministry of Health about quantities of health services delivered for each indicator. Reported quantities are verified and validated by a provincial committee through unannounced observation visits to facilities [55].

In addition to the quantity-based payments, facilities can receive a quality bonus of up to 25 percent [55]. Quality is assessed quarterly by local regulatory authorities on a randomly chosen day using a checklist containing 220 items grouped into the following topics: general infrastructure and communication, business plan, income and costs, hygiene and sterilization, outpatient consultations, family planning, laboratory services, inpatient care, management of essential drugs, availability of essential drugs, maternal care, surgery, tuberculosis screening, vaccination, and antenatal care. The total payment to a facility is calculated as a weighted sum of the number of provided services in the previous three months times their unit payment multiplied by the quality bonus, which ranges between 1 and 1.25 depending on the score obtained from evaluation of facilities based on results of the checklist assessment [55].

An equity bonus is further used to support facilities that face major problems that cannot be resolved by itself and are critical to its performance [132]. The calculation of the bonus accounts for geographic remoteness, the poverty of its clients, and the needs of the staff and facility.

Performance bonuses awarded to the staff through the incentive cannot exceed 30% of the health facility's overall income [132].

PBF implementation in Lesotho

The PBF program in Lesotho had a phased implementation approach. There were two pilot districts: Quthing in April 2014 and Leribe in January 2015. These two districts are excluded from this study. Mokhotlong and Thaba-Tseka began implementation in July 2016 and Mafeteng and Mohale's Hoek began implementation in October 2016, which are analyzed in this study.

PBF payments are determined based on the quantity of services provided, quality of services provided, and for health centers, the relative remoteness of the health center [133]. Quantity performance is assessed monthly, quality performance is measured quarterly, and payments are made quarterly. Facilities report fourteen quantity indicators, which are then verified by the Performance Purchasing Technical Assistance team. There are per-unit incentives for each service that are adapted based on facility performance so that, there are higher incentives for lower performing indicators and small incentives when the service is near capacity.

The quality score is based on a quality assessment checklist (80% weight) and a client satisfaction survey (20% weight) [133]. The assessment checklist and satisfaction tools are specific to the level of the health system (health center vs hospital), and the checklists are reviewed every year for changes to the indicators or revising the points assigned to indicators. The quality performance score translates to a graduated bonus on top of the quantity incentive payment. Health centers receive no quality bonus for quality scores below 50, and 65% bonus for quality scores between 90-100. Hospitals receive no bonus for quality scores below 50 and the full bonus for quality scores between 95-100.

In addition, health centers in remote areas receive additional bonuses to address inequities in retaining staff, higher transportation and communication costs and limited access to other services [133]. Health centers located outside urban areas but with access to public transport and network receive a 10% bonus over the quantity produced (not quality), while those in remote areas with infrequent public transportation and unreliable network receive a 20% bonus. Health centers in urban areas and hospitals receive no remoteness bonuses.

A minimum of 50% of the incentive payments must be used for improvement of service delivery, and a maximum of 50% may be used for motivation bonuses for health center or hospital staff. In addition, district health management teams may receive PBF incentives with an emphasis on quality of supportive supervision and essential support for the PBF project.

PBF implementation in Senegal

Senegal's PBF program was piloted in two regions, Kaffrine and Kolda, in April 2012, and was expanded to four additional regions in May 2016 [134]. Only the two pilot regions were evaluated in this study.

PBF payments are based on both coverage and quality targets. Coverage targets are set for each health facility based on the previous year's target. Payments are disbursed if quarterly and annual coverage targets are met and are deflated by a quality score. Quantity and quality performance is reported by the facility and verified by the Regional Management Committee, facility visits, and household surveys. 25% of PBF payments must be used to cover operational costs, and a maximum of 75% can be used to pay individual incentives to health workers [135].

In addition to the supply-side incentives, Senegal's program also incorporates demand-side incentives, with vouchers for four antenatal care visits and skilled deliveries [134].

PBF implementation in Zambia

Zambia's PBF evaluation contained three arms: an PBF intervention group that received PBF performance-based grants and Emergency Obstetric and Neonatal Care (EmONC) equipment; an unconditional financing arm that received the EmONC equipment and funding equivalent to the average of the PBF performance grants as input financing; and a pure control group that received nothing [57]. This study uses the pure control as the comparison districts in the main analysis and the unconditional financing group in a secondary analysis.

Ten districts were randomized to each evaluation arm. Districts selected for the evaluation approximated the median population health, socio-economic condition, and health governance capacity for the provinces in which they were located. Three districts in each rural province were selected as well as six districts in Northern and Southern provinces; within each province the selected districts were then randomized to either the PBF intervention or one of the two treatment arms.

The PBF program used a contracting-in strategy [57]. PBF payments were based on nine maternal and child health output indicators and ten dimensions of quality. Quantity unit incentives ranged from USD 0.20 for curative consultations to USD 6.40 for institutional deliveries by skilled birth attendants. The ten quality indicators were assigned different weights, then received bonuses on top of the quantity payments, with a quality score of 61%-69% corresponding to an additional 15%, scores of 70%-79% received an additional 25% and scores over 80% received an additional 50% on top of the quantity payments. A small number of more remote health facilities were randomly assigned to receive 25% higher prices for all output indicators. In addition, the district medical offices received performance bonuses for fulfilling a set of supervision and management functions. Quantity and quality scores were externally verified.

Health facilities were required to spend a minimum of 40% of their PBF payments on operational activities, while a maximum of 60% could be spent on staff motivational bonuses [57].

PBF implementation in Zimbabwe

After PBF pilots in two districts, Zimbabwe's PBF program was implemented in 16 districts. The pilot districts were excluded from this study [56]. 32 districts were purposively selected from the universe of 64 districts and pair-matched on the following characteristics: geographic accessibility, type and level of health facilities, average facility catchment population, proportion of staff in position, presence of key staff such as the district medical/health officer, and health service utilization rates for antenatal and postnatal care coverage, institutional delivery and immunization rates. One district in each pair was selected by the Ministry of Health and Child Care into the PBF treatment arm, while the others were used as control districts.

PBF payments were based on quantity, quality and a remoteness bonus [56]. Rural health centers received a unit price for 16 quantity indicators, while district hospitals quantity payments were based on five indicators mostly related to deliveries. Remoteness bonuses up to 30% of the quantity payments were based on population density, distance to the nearest referral facility and availability of roads, public transportation and communications. Quality of services was measured through a client satisfaction survey and a balanced score card covering structural quality, process quality, organization and management systems. Quality scores translated into bonuses on top of the quantity and remoteness payments of up to 25%.

Health facilities were required to spend a minimum of 75% of their PBF payments on improving working conditions at the facility, while up to 25% could be spent on staff bonuses [56].

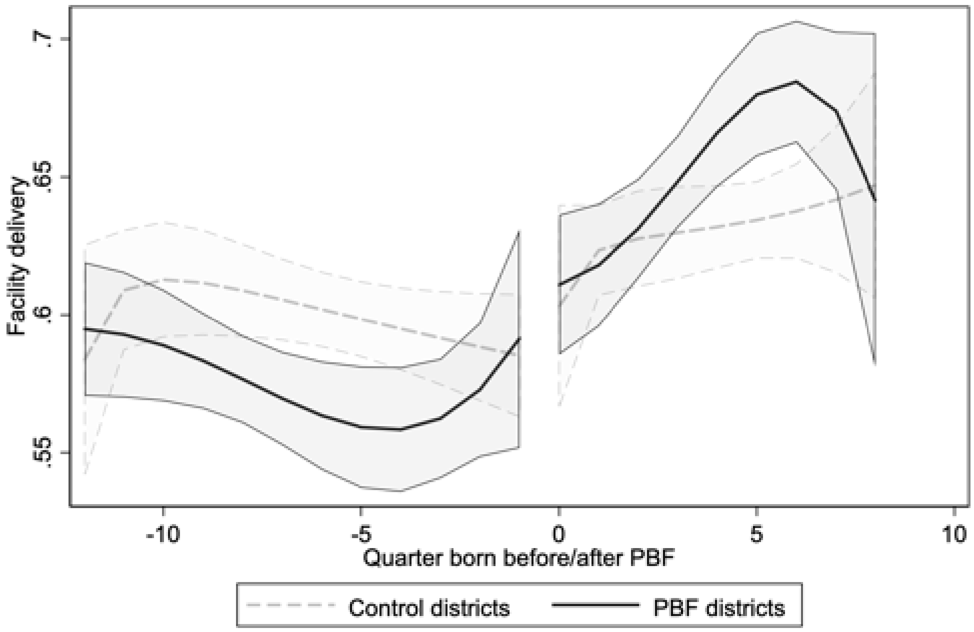
In addition to the PBF payments, intervention districts received training on effective PBF implementation and abolishment of formal user fees for services that PBF was targeting [56].

Text A.2. Notable concurrent socio-political events

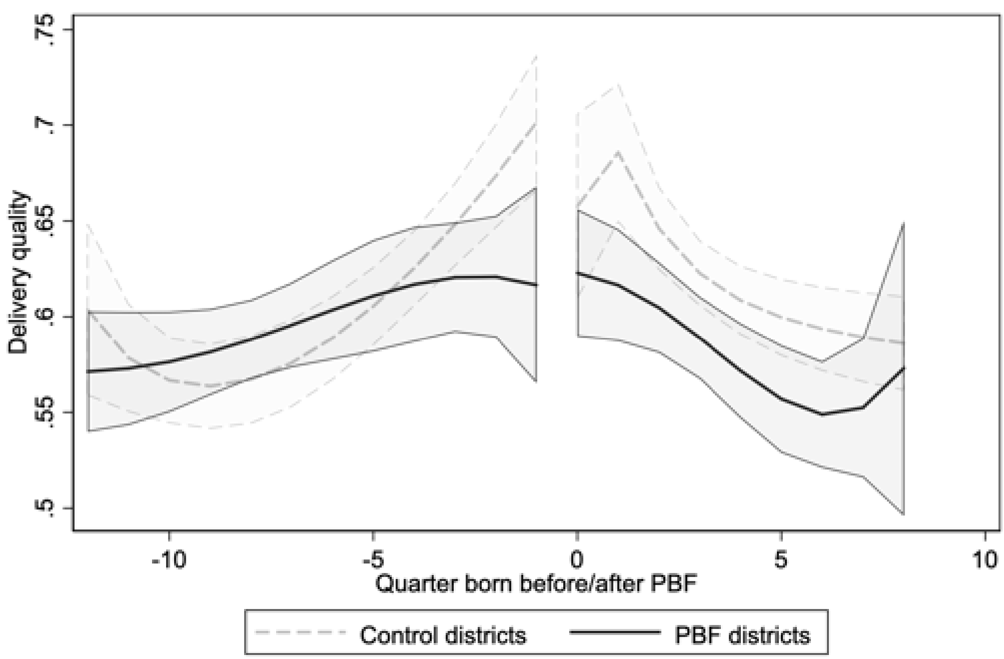
In addition to the concurrent health interventions taking place (Table 2.1), several socio-political events in the study countries are notable. In Burundi when PBF implementation began in 2006, reconstruction was occurring following conflict between rebel groups and the government, though skirmishes continued to erupt. In Senegal, a new president was elected shortly before PBF was implemented in 2012, and major flooding occurred later in the year, affecting some implementation and control districts. Finally, in Zimbabwe during 2012 implementation, there was relative economic growth between two crashes in 2008 and 2016 though there were nevertheless challenges in hiring and retaining health workers. None of these events are expected to differentially impact health outcomes in control versus implementation areas.

Figure A.1. Trends in secondary outcomes over study period

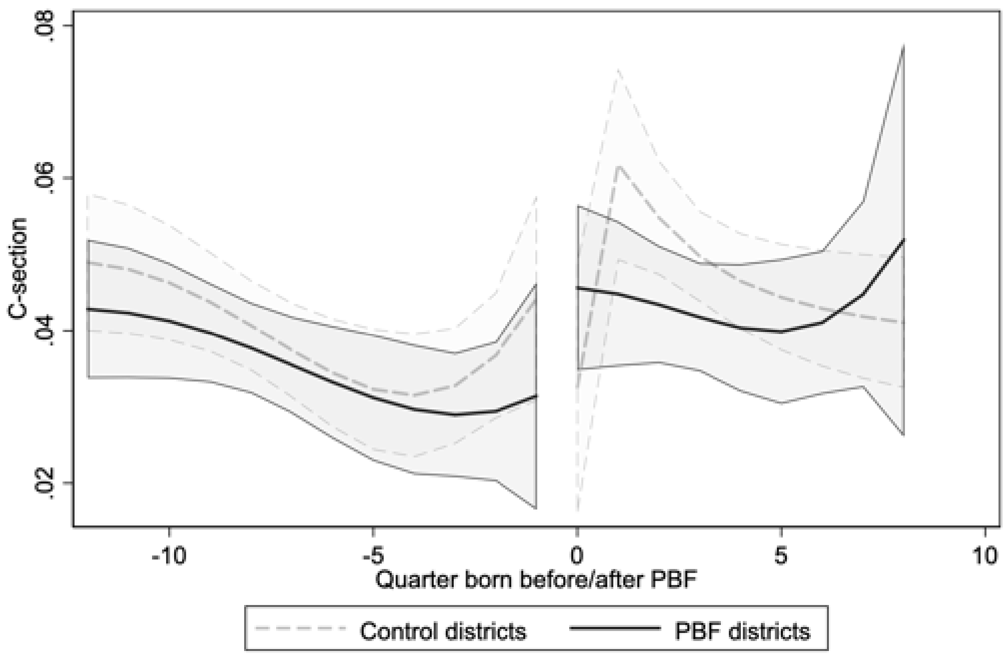
A. Facility delivery trends



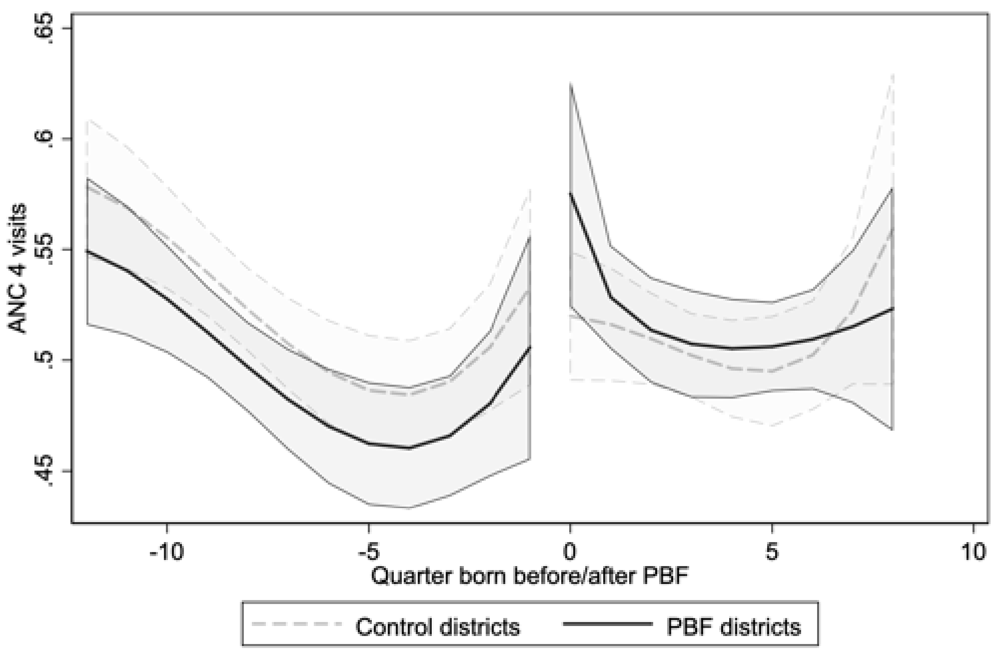
B. Delivery quality trends



C. Cesarean section trends



D. Antenatal care utilization trends



E. Antenatal care quality trends

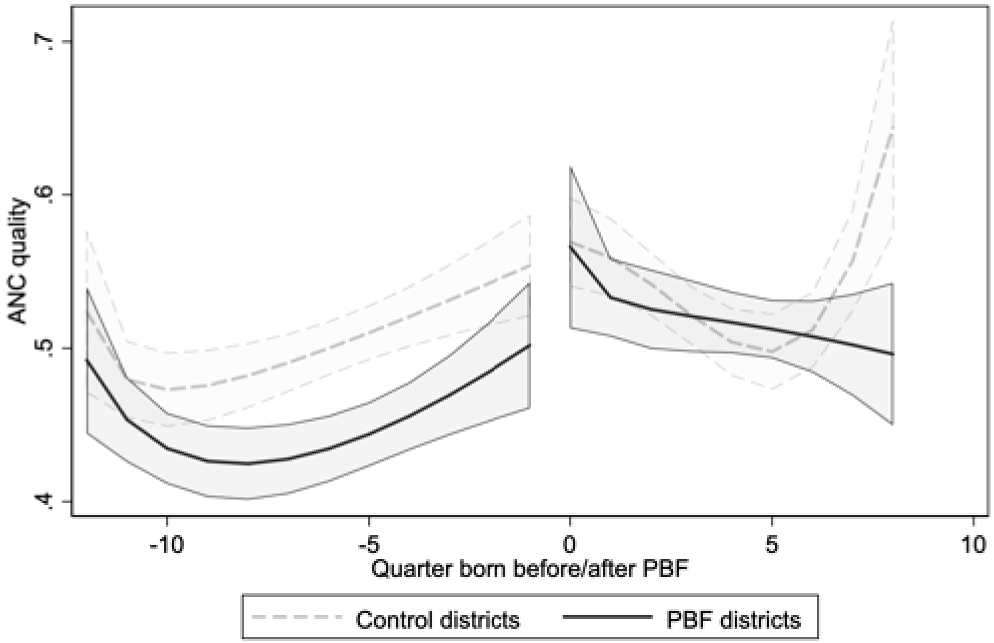
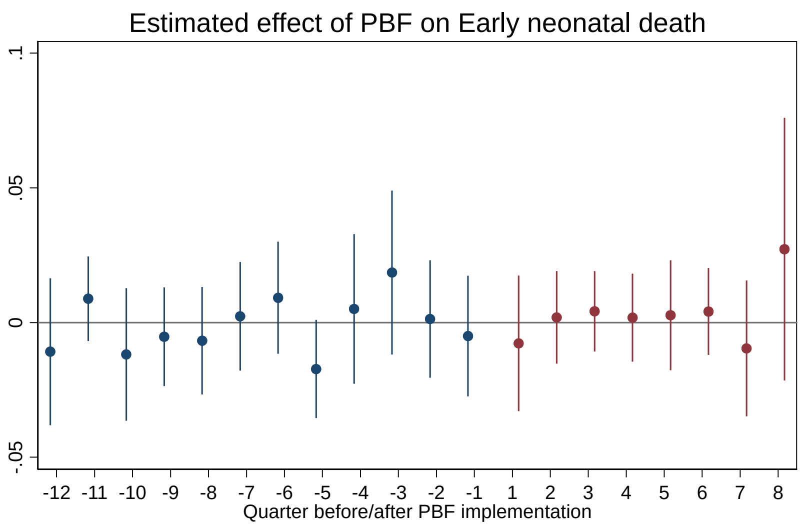
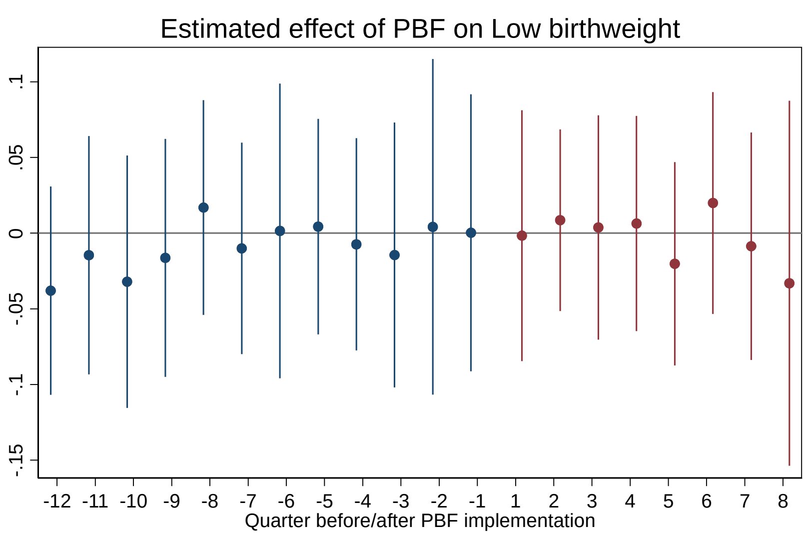
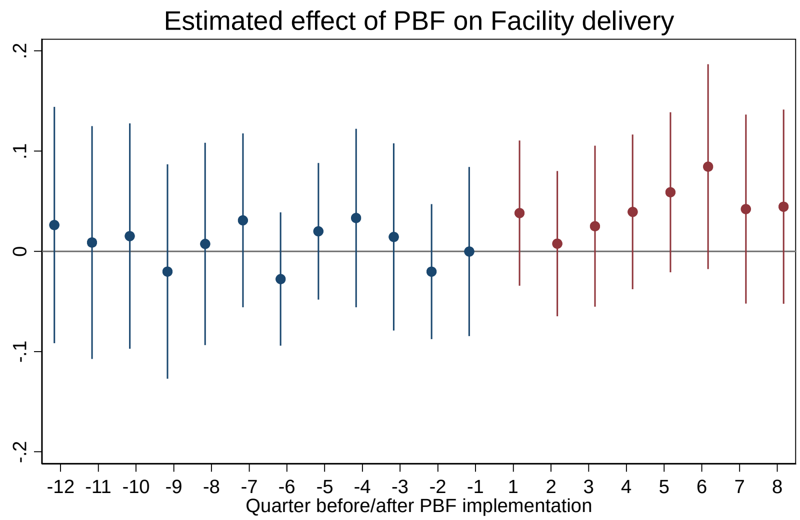


Figure A.2. Modeled effects in pre-trends and over time

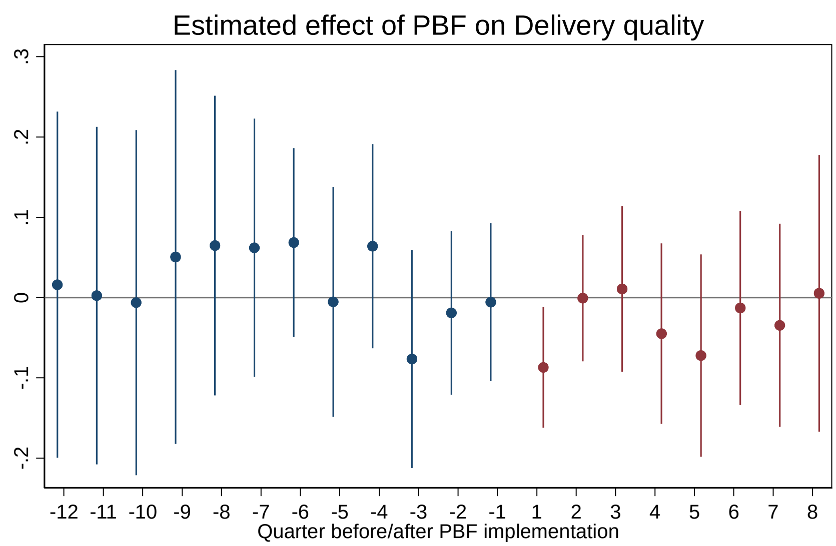
These graphs show the estimated effect of PBF relative to time zero (start of PBF implementation), using a multi-period difference and difference framework. Dummy indicators are included in the models for PBF implementation interacted with the quarter of interest, along with month and district fixed effects. These can be used to evaluate if there were differences in trends prior to implementation (blue points and confidence intervals), or if there were any delayed PBF effects (red).



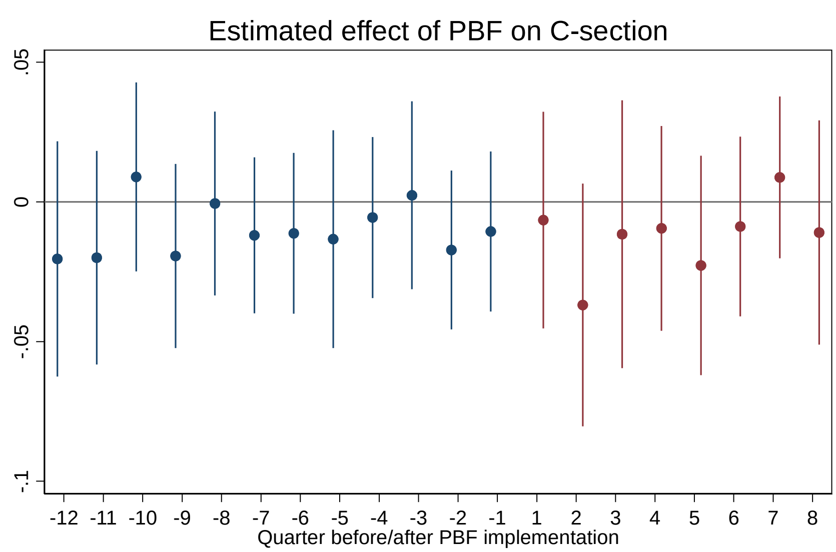




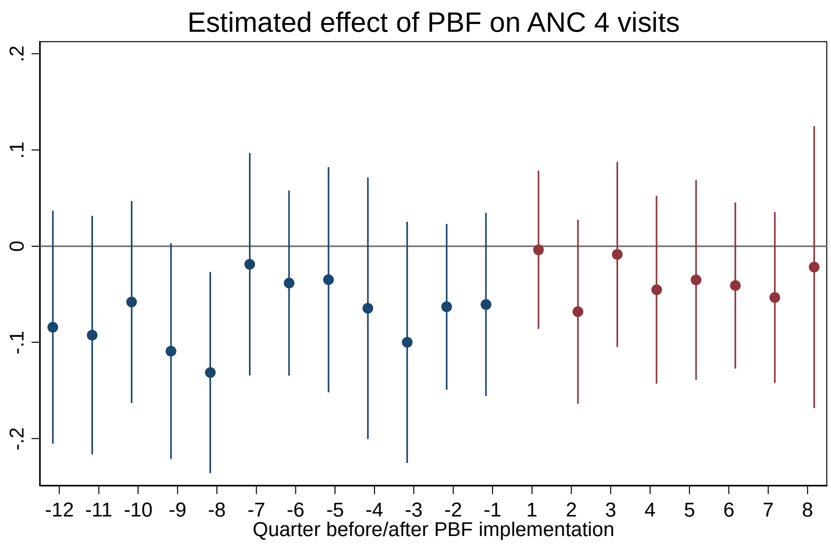
Estimated effect of PBF on Facility delivery
.2
.1
0
-.1
-.2
-12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8
Quarter before/after PBF implementation



Estimated effect of PBF on Delivery quality
.3
.2
.1
0
-.1
-.2
-12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8
Quarter before/after PBF implementation



Estimated effect of PBF on C-section
.05
0
-.05
-.1
-12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8
Quarter before/after PBF implementation


Estimated effect of PBF on ANC 4 visits
.2
.1
0
-.1
-.2
-12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8
Quarter before/after PBF implementation

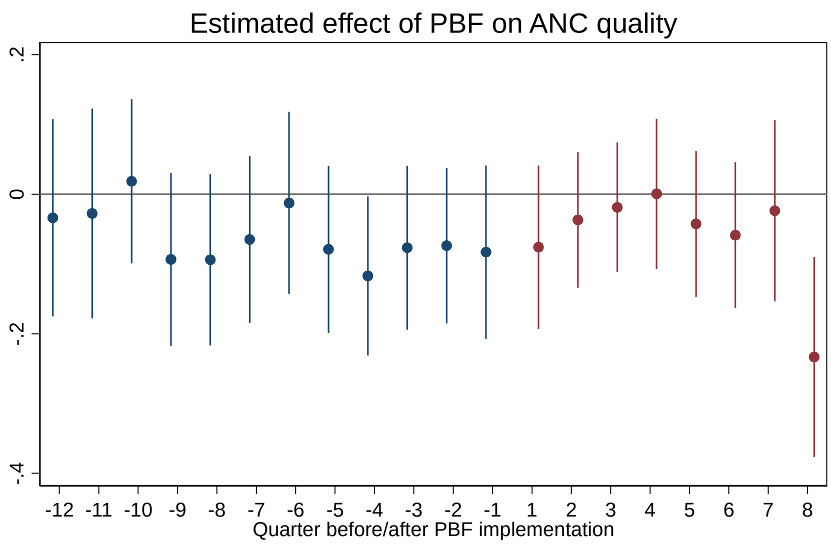

Estimated effect of PBF on ANC quality
.2
0
-.2
-.4
-12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8
Quarter before/after PBF implementation

Table A.1. Robustness checks

Pooled impact on alternative quality measures and birthweight measurement

Outcome	Percentage point change	95% CI	N
Low birthweight (recorded observations only)	0.012	(-0.02,0.04)	11226
Delivery quality (mean)[a]	-0.019	(-0.05,0.01)	13054
ANC quality (mean)[a]	0.015	(-0.01,0.04)	14510
Birthweight recorded[b]	0.011	(-0.1,0.12)	28619

[a]Defined as the percent of delivery or ANC items received
[b]Defined as whether birthweight copied from a record (as opposed to mother's report)

Pooled results in all study countries except for Burundi

Outcome	Percentage point change	95% CI	N
Early neonatal death	0.00	(-0.01,0.01)	20331
Low birthweight	0.01	(-0.01,0.03)	20331
Facility delivery	0.03	(-0.01,0.07)	19331
Delivery quality	-0.05	(-0.14,0.05)	11835
C-section	0.00	(-0.02,0.01)	19279
ANC 4 visits	0.04	(-0.02,0.10)	13590
ANC quality	0.02	(-0.05,0.09)	13714

Pooled results using all non-implementation districts in the country as a comparison

Outcome	Percentage point change	95% CI	N
Early neonatal death	0.00	(-0.00,0.01)	64194
Low birthweight	0.00	(-0.01,0.02)	64194
Facility delivery	0.04	(-0.00,0.07)	51730
Delivery quality	-0.03	(-0.10,0.04)	35446
C-section	0.00	(-0.01,0.01)	51478
ANC 4 visits	0.04	(-0.00,0.08)	34491
ANC quality	0.03	(-0.01,0.08)	34902

Pooled results using second implementation date in Burundi and Lesotho

Country	Outcome	Percentage point change	95% CI	N
Burundi	Early neonatal death	0.00	(-0.01,0.01)	5588
Burundi	Low birthweight	0.01	(-0.12,0.13)	5588
Burundi	Facility delivery	-0.03	(-0.1,0.04)	2154

Burundi	Delivery quality	0.02	(-0.16,0.19)	1212
Burundi	C-section	-0.01	(-0.02,0.01)	2160
Burundi	ANC 4 visits	0.10	(-0.01,0.21)	1263
Burundi	ANC quality	0.12	(0.07,0.18)	1267
Lesotho	Early neonatal death	0.00	(-0.04,0.04)	1899
Lesotho	Low birthweight	-0.10	(-0.19,-0.01)	1899
Lesotho	Facility delivery	0.06	(-0.01,0.13)	1213
Lesotho	Delivery quality	-0.16	(-0.27,-0.05)	1054
Lesotho	C-section	-0.03	(-0.21,0.15)	1213
Lesotho	ANC 4 visits	0.13	(-0.07,0.34)	1213
Lesotho	ANC quality	-0.08	(-0.39,0.22)	1230

Pooled results with country fixed effects instead of district fixed effects

Outcome	Percentage point change	95% CI	N
Early neonatal death	0.004	(-0.00, 0.01)	28,619
Low birthweight	0.007	(-0.04,0.05)	28,619
Facility delivery	0.048	(-0.02,0.11)	21,471
Delivery quality	-0.058	(-0.13,0.02)	13,054
C-section	-0.004	(-0.02,0.01)	21,424
ANC 4 visits	0.001	(-0.04,0.04)	14,383
ANC quality	0.008	(-0.08,0.09)	14,510

Pooled subset results among poor and rural women

Outcome	Percentage point change	95% CI	N
Early neonatal death	0.00	(-0.01,0.02)	9456
Low birthweight	0.00	(-0.04,0.03)	9456
Facility delivery	0.02	(-0.04,0.08)	7834
Delivery quality	-0.05	(-0.15,0.06)	3333
C-section	-0.01	(-0.02,0.01)	7817
ANC 4 visits	0.08	(-0.01,0.17)	4980
ANC quality	0.05	(-0.03,0.14)	5010

PBF vs unconditional financing in Zambia

Outcome	Percentage point change	95% CI	N
Early neonatal death	0.00	(-0.03,0.02)	3432
Low birthweight	-0.02	(-0.09,0.05)	3432
Facility delivery	0.05	(-0.05,0.14)	3418
Delivery quality	-0.06	(-0.15,0.03)	2282

C-section	-0.01	(-0.04,0.01)	3424
ANC 4 visits	0.01	(-0.09,0.11)	2267
ANC quality	0.06	(-0.02,0.14)	2307

Table A.2. Ex-post power calculation assumptions

We conducted ex-post power calculations to estimate the minimum detectable effect (MDE) possible with the available sample size. We calculated the MDE with the following assumptions:

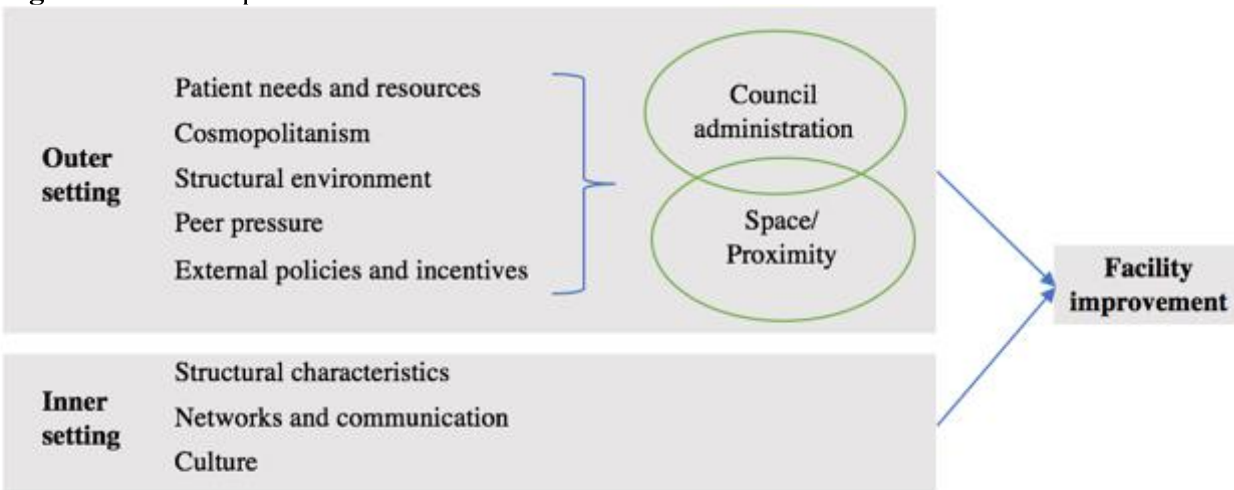
1. The observations are clustered based on the actual number of districts in the study: 40 treatment and 35 control
2. The sample size is based on the actual number of matched observations for each dependent variable, see table
3. Initial mean and rho are calculated from all matched observations at baseline, see table
4. Power=0.80, Alpha=0.05

Minimum detectable difference and assumptions

	N treat	N control	Mean	Rho	MDE
Early neonatal death	13164	18484	0.022	0.003	-0.007
Low birthweight	7853	11842	0.182	0.003	-0.020
Facility delivery	10074	12607	0.586	0.037	0.064
Delivery quality	6167	7663	0.601	0.046	0.073
C-section	10040	12595	0.036	0.009	0.012
ANC 4 visits	6948	8585	0.512	0.017	0.048
ANC quality	6999	8664	0.481	0.020	0.051

Appendix B: Chapter 3 Supplementary materials

Figure B.1. Conceptual model



Text B.1. Additional details on the MCMC hierarchical spatial autoregressive model

The hierarchical spatial autoregressive model from Dong and Harris[16] was used to jointly estimate the hierarchical and spatial components of the variation in facility improvement. The analysis used the HSAR package in R and the spatial weight matrices were created using the spdep package.

The HSAR package uses Bayesian Markov Chain Monte Carlo (MCMC) algorithms to fit the HSAR model. This method is preferred for spatial econometrics models that have complex variance-covariance structures or spatial panel data models with random effects. The model takes the form:

$$y_{ij} = \rho W_i y + \beta X_{ij} + \theta_j + e_{ij}$$

$$\theta_j = \lambda M_j \theta + u_j$$

Where y is the change in Star Rating in facility i and council j, ρ is a spatial lag term at the facility level, W_i is the facility inverse distance weight matrix for facilities within 50 km, X is the vector of covariates included in the full model, λ is a spatial lag term at the council level and M_j is a queen's contiguity spatial weight matrix for councils.

The first level (facility) spatial weight matrix was defined as an inverse distance weight matrix for facilities within 50 km. This indicates that closer facilities will have a greater weight than more distant facilities, and facilities beyond 50 km have no weights. Euclidean distances were used, so the weights may not accurately reflect the proximity of facilities based on road distances or travel times. The second level (council) weight matrix was defined as a queen's contiguity matrix; that is any council that shared any boundary or corner was considered a neighbor. Both weight matrices are row normalized to ensure the maximum value of the spatial autoregressive parameters are less than 1.

We ran 10,000 simulations of the model with a burn in of 5000 and thinning factor of 5. To implement the MCMC algorithms, we must specify the prior distributions for the following parameters: first level variance (σ_e^2), second level variance (σ_u^2), betas for each predictor, first level spatial lag (ρ), and second level spatial lag (λ). We used the variance estimates from the full model as the priors in the HSAR model (0.42 for σ_e^2 and 0.11 for σ_u^2). The beta priors were set to those from an ordinary least squares regression with the outcome and all predictors included. The spatial lag term priors were each set to 0.1. We conducted a robustness check in which perturbed these priors and estimated the value when setting the priors to every 0.05 between 0.0 and 0.5. The results were robust to these alternate priors: the range of mean ρ estimates were from 0.331 to 0.340 and the range of mean λ estimates were from 0.333 to 0.376.

As a sensitivity analysis, we also constrained λ to be zero, assuming that the spatial patterning only occurs at the facility level rather than the council level.

Table B.1 Bunching at Star Rating thresholds in the underlying domain scores

Assessment	Baseline						Reassessment					
Threshold	20		40		60		20		40		60	
	T	SE	T	SE	T	SE	T	SE	T	SE	T	SE
Minimum domain (determines Star Rating)	**0.22**	0.04	**0.19**	0.08	0.33	0.25	**0.52**	0.12	**0.24**	0.06	0.06	0.07
Mgmt. & staff performance domain	**0.35**	0.06	0.05	0.06	0.27	0.12	**0.35**	0.19	**0.30**	0.07	**0.23**	0.06
Service charters & accountability domain	**0.31**	0.07	**0.17**	0.05	0.04	0.10	**0.34**	0.19	**0.50**	0.08	**0.35**	0.06
Safe & conducive domain	**0.40**	0.06	-0.01	0.06	0.20	0.12	**0.50**	0.15	**0.22**	0.07	**0.35**	0.06
Quality of care domain	**0.38**	0.12	**0.24**	0.05	**0.27**	0.08	-0.21	0.35	**0.49**	0.07	**0.21**	0.05

Notes: We used the McCrary (2008) method to identifying bunching in the running variable for regression discontinuity designs. We tested each domain at the thresholds of 20, 40 and 60, as well as for the minimum domain in each facility, which determines the facility's Star Rating score. The null hypothesis is for a smooth distribution of scores above and below the thresholds; bunching just above the threshold signals the possibility of manipulation. The table above shows the T statistics and standard errors for each domain and threshold. Positive t statistics indicate bunching above the threshold, bolded numbers indicate statistical significance ($\alpha < 0.5$).

Table B.2. Sensitivity analysis using the change in the overall score (0-100) as the outcome rather than the number of stars (0-5)

	Null model		Full model		HSAR model	
	Coef.	95% CI	Coef.	95% CI	Coef.	95% CI
Sum of people within 5 km radius of facility (ln)			0.8	(0.4,1.1)	0.6	(0.3,1)
Institutional delivery percent in council			2.7	(-5.7,11.2)	3.2	(-0.2,6.6)
Female primary education percent in council			3.5	(-8.3,15.2)	4.1	(-0.6,8.9)
Healthcare decisions percent in council			-9.5	(-23.8,4.8)	-4.7	(-10.4,1.1)
Facilities in council per population			-0.1	(-0.3,0.1)	0.0	(-0.1,0.1)
Urban council			1.3	(-2.9,5.5)	1.6	(-0.2,3.5)
Distance to major road (ln)			-0.1	(-0.2,0.1)	-0.1	(-0.2,0.1)
Distance to large city (10 kms)			0.1	(-0.1,0.2)	-0.1	(-0.2,0)
Percentile rank at baseline			3.7	(0.6,6.9)	0.1	(-2.8,3)
RBF Participation			5.9	(4.2,7.6)	4.1	(2.6,5.7)
RBF ineligibility due to low baseline			11.0	(9.1,12.8)	8.8	(7.1,10.5)
Starter fund			0.4	(-1.5,2.4)	-0.5	(-2.4,1.4)
Ownership (Public ref.)						
Private			-4.2	(-3.4,-0.8)	-4.7	(-5.9,-3.4)
Non-profit			-2.1	(-1.2,1.2)	-2.4	(-3.6,-1.2)
Level (Dispensary ref.)						
Health center			3.0	(6.3,8.3)	2.8	(1.8,3.8)
Primary hospital			7.3	(-12,-8.1)	7.2	(5.3,9.2)
Baseline performance			-10.1	(-1.6,1.6)	-8.2	(-9.7,-6.8)
First level lag term					0.4	(0.3,0.5)
Second level lag term					0.2	(-0.1,0.5)
Constant	19.1	(17.6,20.6)	20.1	(9.3,31)	10.5	(5,16.1)
Council variance	63.9		41.78		16.6	10%
Individual variance	168.9		133.29		141.5	90%
Moran's I of residuals (p value)	0.07	0.00	0.05	0.00		

Table B.3. Sensitivity analysis that excludes facilities that scored 3 or 4 stars at baseline

	Null model		Full model	
	Coef.	95% CI	Coef.	95% CI
Sum of people within 5 km radius of facility (ln)			0.08	(0.05,0.1)
Institutional delivery percent in council			0.36	(-0.06,0.78)
Female primary education percent in council			0.23	(-0.38,0.84)
Healthcare decisions percent in council			-0.45	(-1.19,0.29)
Facilities in council per population			-0.01	(-0.02,0)
Distance to major road (ln)			-0.01	(-0.02,0)
Distance to large city (10 kms)			0.00	(0,0.01)
Percentile rank at baseline			0.06	(-0.14,0.26)
RBF Participation			0.36	(0.27,0.46)
RBF ineligibility due to low baseline			0.60	(0.5,0.7)
Ownership (Public ref.)				
Private			-0.32	(-0.39,-0.25)
Non-profit			-0.10	(-0.17,-0.04)
Level (Dispensary ref.)				
Health center			0.36	(0.31,0.42)
Primary hospital			0.79	(0.67,0.9)
Baseline performance			-0.72	(-0.83,-0.61)
Constant	1.03	(0.96,1.1)	0.71	(0.15,1.27)
Council variance	0.15		0.11	
Individual variance	0.62		0.42	
Moran's I of residuals (p value)	0.05	0.00	0.05	0.00

Notes: Facilities that scored 3 or 4 stars at baseline may encounter a ceiling effect in improvement. These are excluded in order to ensure the models are not unduly influenced by these high performers.

Table B.4. Sensitivity analysis where $\lambda = 0$ in the HSAR model

	HSAR model	
	Coef.	95% CI
Sum of people within 5 km radius of facility (ln)	0.07	(0.05,0.09)
Institutional delivery percent in council	0.30	(0.11,0.49)
Female primary education percent in council	0.23	(-0.06,0.51)
Healthcare decisions percent in council	-0.17	(-0.51,0.18)
Facilities in council per population	0.00	(-0.01,0.00)
Urban councils	0.06	(-0.04,0.16)
Distance to major road (ln)	-0.01	(-0.02,0.00)
Distance to large city (10 kms)	0.00	(-0.01,0.00)
Percentile rank at baseline	-0.18	(-0.34,-0.01)
RBF Participation	0.29	(0.20,0.38)
RBF ineligibility due to low baseline	0.50	(0.40,0.59)
Starter fund	0.08	(-0.03,0.19)
Ownership (Public ref.)		
Private	-0.32	(-0.39,-0.24)
Non-profit	-0.11	(-0.17,-0.04)
Level (Dispensary ref.)		
Health center	0.34	(0.28,0.40)
Primary hospital	0.74	(0.63,0.85)
Baseline performance	-0.58	(-0.66,-0.50)
First level lag term	0.34	(0.27,0.40)
Constant	0.31	(-0.01,0.62)
Council variance	0.05	
Individual variance	0.44	

Notes: Setting $\lambda = 0$ does not allow for spatial patterning at the council level. This constrains the spatial patterning in the HSAR model to the facility level.

Figure B.2 Council level random effects from null and full models

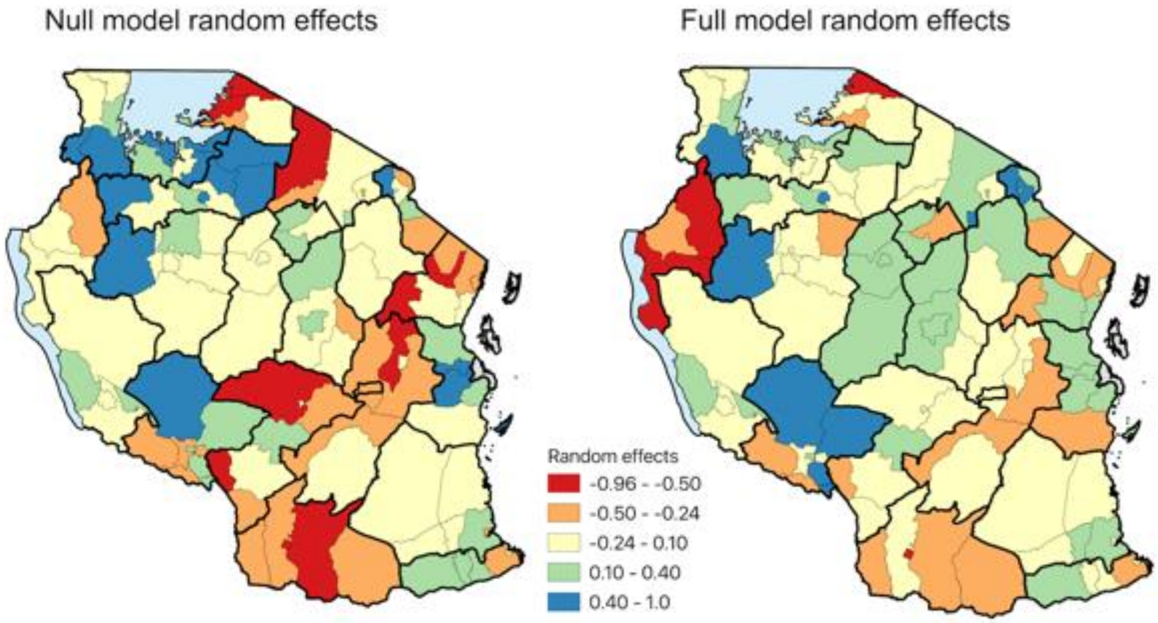


Table B.5. Comparison of analytic sample with facilities that were only assessed at baseline

	Analytic sample	No reassessment
Baseline star rating		
0	34%	47%
1	52%	42%
2	12%	10%
3	2%	1%
4	0%	0%
N facilities	5,595	137

Notes: The facilities only assessed at baseline may be considered lost to follow up.

Appendix C: Chapter 4 Supplementary materials

Figure C.1. Examples of provider arrangements during labor and delivery

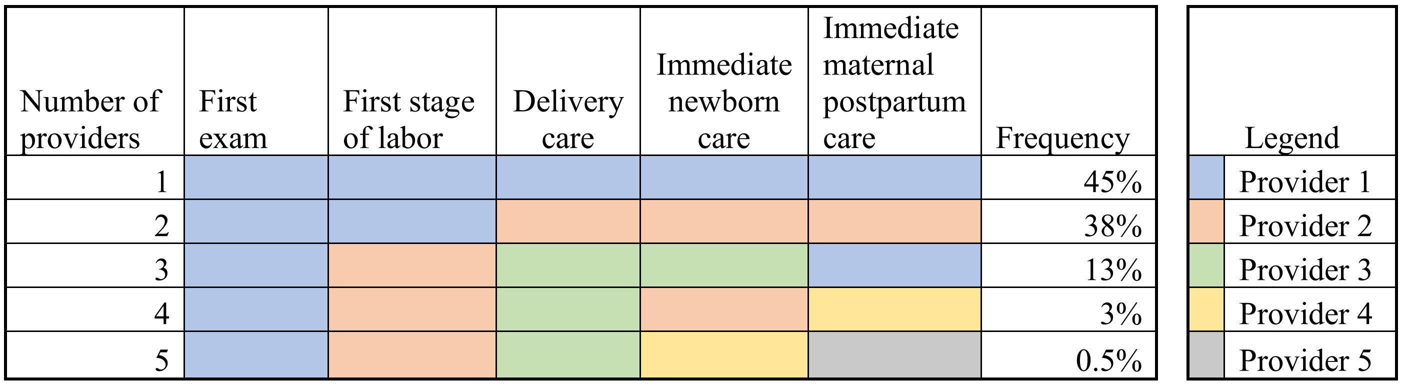


Table C.1. Quality of intrapartum and immediate postpartum care processes with no time constraints on actions

	Percent completed	N
Average of first exam actions	47%	823
Checks woman's HIV status	68%	809
Asks whether woman has experienced headaches or blurred vision	6%	823
Asks whether woman has experienced vaginal bleeding	7%	823
Takes blood pressure during initial client assessment	73%	822
Takes pulse during initial client assessment	67%	822
Washes hands before initial examination	13%	823
Wears gloves before vaginal examination	99%	781
Average of first stage of labor actions	48%	795
At least once, explains what will happen in labor	42%	824
Prepares uterotonic drug to use for AMTSL	81%	791
Uses partograph during labor	53%	793
Prepares bags and masks for neonatal resuscitation	14%	783
Average of third stage of labor actions	88%	828
Correctly administers uterotonic (no time requirement)	93%	820
Assesses completeness of placenta and membranes	74%	819
Assesses for perineal and vaginal membranes	90%	819
Ties or clamps cord	99%	828
Average of immediate newborn care	81%	694
Immediately dries baby with towel	98%	694
Places newborn on mother's abdomen skin-to-skin	64%	690
Average of immediate maternal postpartum stage actions	37%	824
Takes mother's vital signs after birth	3%	823
Palpates uterus after birth	72%	824
Assists mother to initiate breastfeeding	37%	695
Alternate Quality of Intrapartum and Immediate Postpartum Care Processes index	58%	828

Figure C.2. Decomposition of variance in quality of care

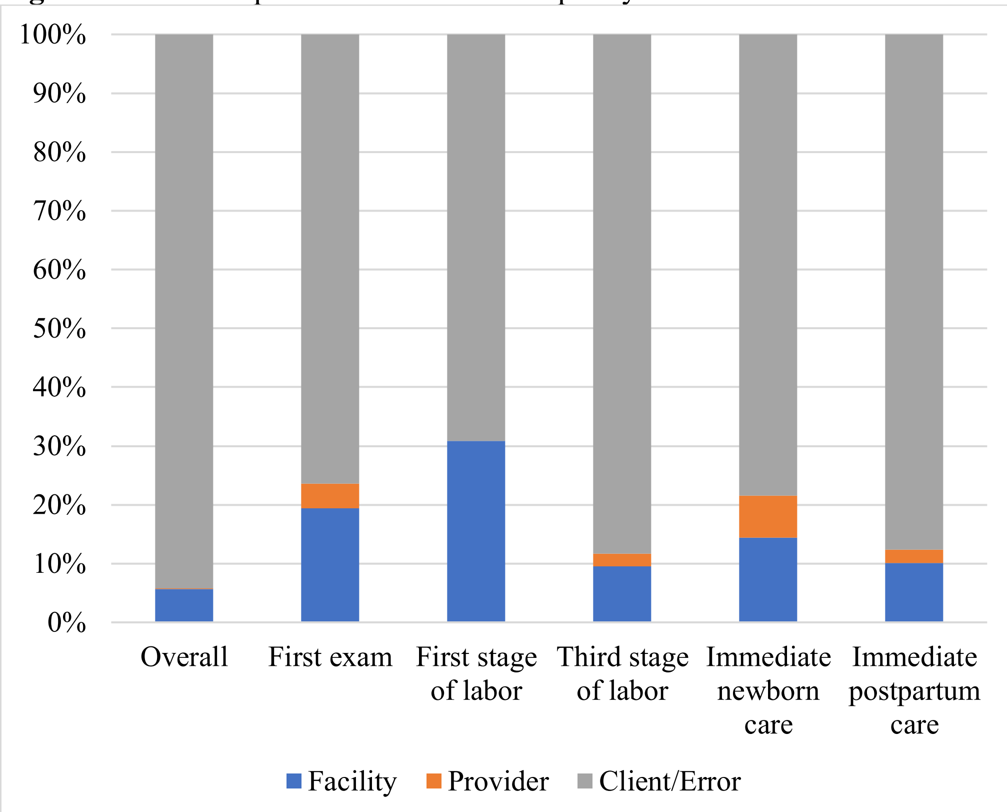
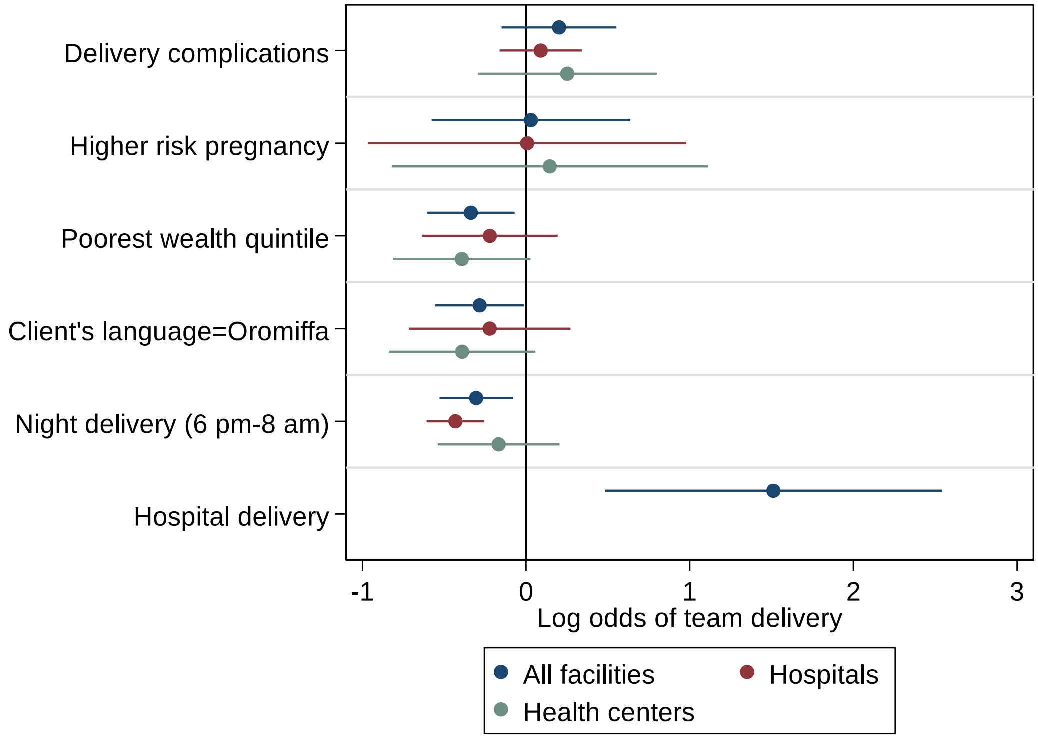


Table C.2. Quality of care by number of providers

Number of providers	Quality of care Mean	N observations	p-value
1	0.518	371	
2	0.506	317	0.228
3	0.473	109	0.002
4	0.500	27	0.490
5	0.557	4	0.553
Total	0.507	828	

Notes: p-value for difference relative to solo deliveries (1 provider)

Figure C.3. Predictors of having multiple providers during a delivery



Notes: Points and confidence intervals show the log odds of predictors association with having more than one provider attend the delivery, with observations at the delivery level. Delivery complication is neonatal resuscitation, newborn referred to the NICU, or mother treated for post-partum hemorrhage or eclampsia. Higher risk pregnancy is mother grand multiparous (5 or more births), younger than 18 or older than 35, or has multiple births.

Table C.3. Group characteristics association with overall quality of care

	Quality of care (full index)	
	Coef.	p value
Group characteristics		
Average team performance (excluding index observation)	0.047	0.00
At least one senior cadre in group	-0.01	0.04
At least one experienced provider	-0.001	0.92
Number of providers	0.002	0.91
Context and environment		
Delivery complication	0.01	0.50
Higher risk pregnancy	0.004	0.68
Night delivery (morning reference)	0.005	0.68
Client wealth index (poorest reference)		
Wealth 2	0.027	0.29
Wealth 3	0.062	0.00
Wealth 4	0.037	0.00
Wealth 5 (wealthiest)	0.036	0.02
Facility (Dil Chorra Hospital ref.)		
Sabien Primary Hospital	0.035	0.00
Biyowale Health Center	0.041	0.00
Legeharae Health Center	0.046	0.17
Melka Jebdu Health Center	0.072	0.02
Wahil Health Center	0.045	0.00
Gende Gerada Health Center	0.036	0.02
Goro Health Center	0.063	0.02
Jelobelina Health Center	0.083	0.02
Constant	0.441	0.00
N	457	

Figure C.4. Index provider and coworker provider performance among 2 provider deliveries (N=1326)

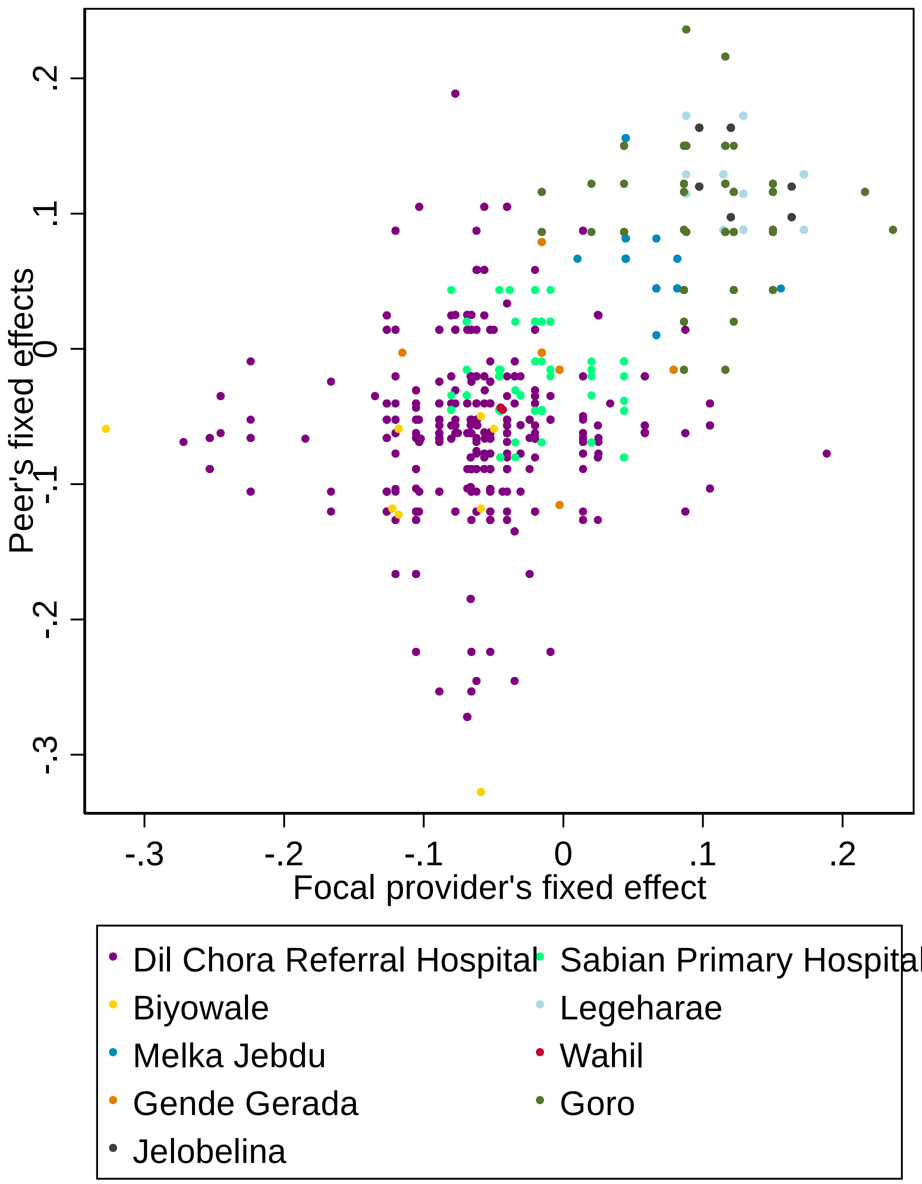


Figure C.5. Index provider and coworker provider performance among deliveries with >2 providers (N=674)

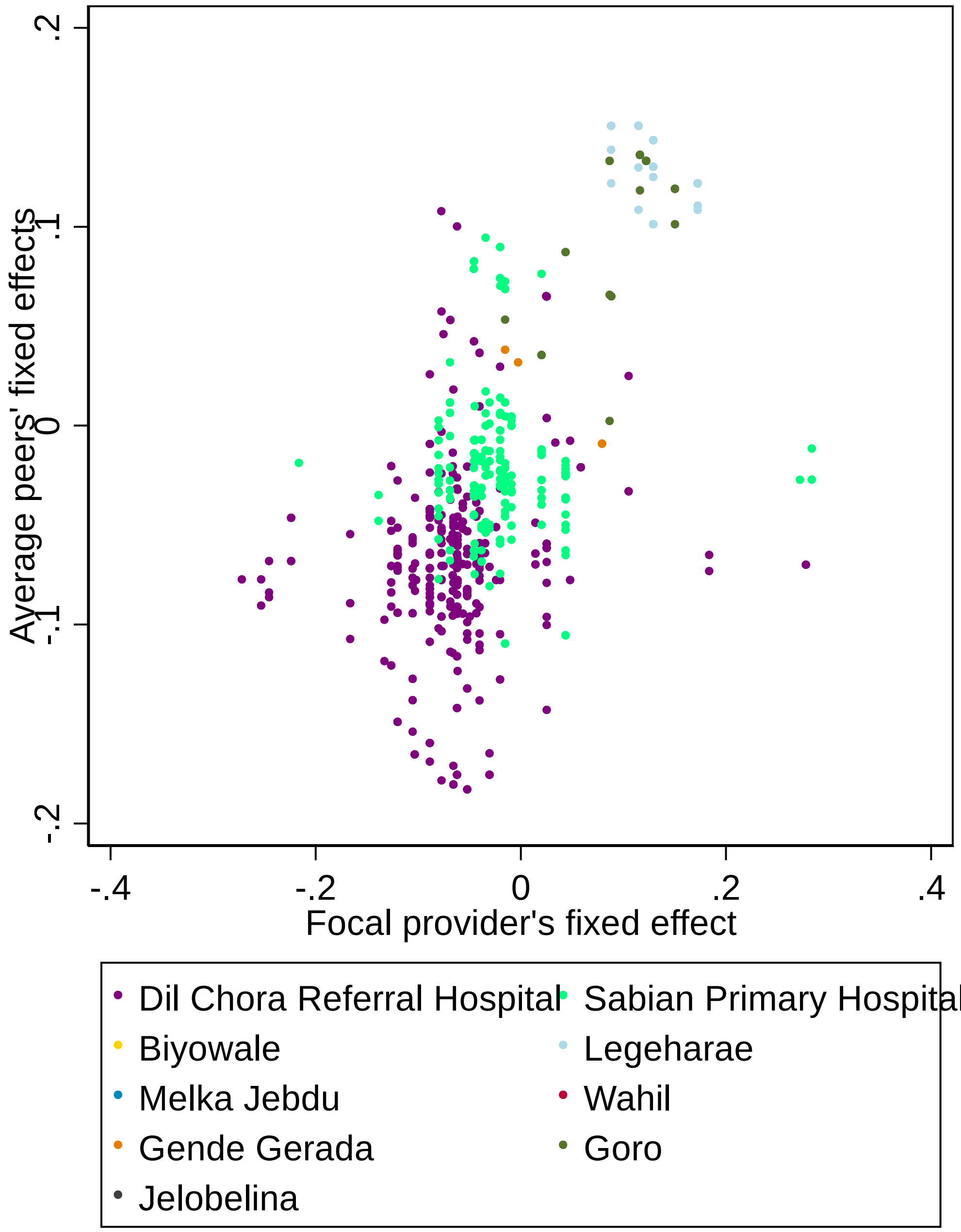


Table C.4. Bivariate relationships between independent variables and quality of care

	Coef.	p value	95% CI	N
Coworkers' performance across all deliveries	0.059	0.00	(0.04,0.07)	2189
Index provider is most senior in group	-0.056	0.01	(-0.1,-0.02)	2189
Index provider is most experienced in group	0.003	0.82	(-0.02,0.03)	2189
Number of providers	-0.006	0.53	(-0.03,0.01)	2189
Years of experience	0.003	0.44	(0,0.01)	2101
Delivery complication	-0.002	0.92	(-0.04,0.04)	2189
Higher risk pregnancy	-0.001	0.97	(-0.04,0.04)	2186
Night delivery	0.012	0.40	(-0.02,0.04)	2189
Provider cadre (Midwife or nurse diploma ref.)				
Midwife or nurse Bsc.	0.047	0.09	(-0.01,0.1)	2149
Health officer	-0.052	0.17	(-0.13,0.02)	2149
GP	-0.037	0.14	(-0.09,0.01)	2149
IESO	0.024	0.85	(-0.22,0.27)	2149
Client wealth index (poorest reference)				
Wealth 2	0.043	0.05	(0,0.09)	1990
Wealth 3	0.077	0.00	(0.03,0.12)	1990
Wealth 4	0.026	0.18	(-0.01,0.06)	1990
Wealth 5 (wealthiest)	0.039	0.05	(0,0.08)	1990
Delivery stage (first exam ref.)				
First stage of labor	-0.001	0.96	(-0.04,0.03)	2189
Third stage of labor	0.228	0.00	(0.2,0.26)	2189
Immediate newborn care	0.294	0.00	(0.25,0.33)	2189
Immediate maternal postpartum care	-0.323	0.00	(-0.36,-0.29)	2189
Facility (Dil Chorra Hospital ref.)				
Sabien Primary Hospital	0.028	0.06	(0,0.06)	2189
Biyowale Health Center	0.067	0.20	(-0.04,0.17)	2189
Legeharae Health Center	0.189	0.00	(0.15,0.23)	2189
Melka Jebdu Health Center	0.139	0.00	(0.09,0.19)	2189
Wahil Health Center	-0.020	0.10	(-0.04,0)	2189
Gende Gerada Health Center	0.034	0.31	(-0.03,0.1)	2189
Goro Health Center	0.197	0.00	(0.16,0.23)	2189
Jelobelina Health Center	0.189	0.00	(0.12,0.26)	2189

Table C.5. Multivariable associations between group characteristics and quality of care by delivery stage

	First exam		First stage of labor		Third stage of labor	
	Coef.	p value	Coef.	p value	Coef.	p value
Group characteristics						
Peers' performance across all deliveries	-0.005	0.77	0.042	0.01	0.020	0.10
Peers are more senior cadre than index	-0.016	0.65	-0.097	0.00	-0.041	0.13
Peers are more experienced than index	0.000	0.98	-0.002	0.96	-0.011	0.56
Number of providers	-0.011	0.42	0.039	0.05	-0.002	0.89
Index provider characteristics						
Years of experience	-0.001	0.75	0.000	0.92	0.000	0.90
Provider cadre (Midwife or nurse diploma ref.)						
Midwife or nurse Bsc.	-0.065	0.05	-0.140	0.00	-0.034	0.20
Health officer	-0.142	0.08	-0.016	0.82	-0.161	0.12
GP	-0.002	0.97	-0.135	0.00	-0.053	0.13
IESO	-0.206	0.05	-0.088	0.57	0.035	0.77
Context and environment						
Delivery complication	0.044	0.02	-0.039	0.24	0.012	0.66
Higher risk pregnancy	-0.022	0.52	0.038	0.28	0.013	0.65
Night delivery (morning reference)	0.006	0.75	0.002	0.93	-0.021	0.22
Client wealth index (poorest reference)						
Wealth 2	0.013	0.67	0.095	0.02	-0.004	0.88
Wealth 3	0.038	0.22	0.123	0.00	0.066	0.05
Wealth 4	0.014	0.63	0.066	0.11	0.030	0.37
Wealth 5 (wealthiest)	0.002	0.94	0.060	0.11	0.046	0.10
Facility (Dil Chorra Hospital ref.)						
Sabien Primary Hospital	0.156	0.00	0.013	0.62	-0.022	0.51
Biyowale Health Center	-0.150	0.00	0.122	0.36	0.132	0.01
Legeharae Health Center	0.114	0.01	0.188	0.00	0.116	0.00
Melka Jebdu Health Center	0.104	0.08	0.173	0.02	0.106	0.00
Wahil Health Center	0.078	0.09	-0.157	0.00	-0.002	0.94
Gende Gerada Health Center	0.085	0.16	0.122	0.00	0.076	0.24
Goro Health Center	0.152	0.00	0.344	0.00	0.050	0.15
Jelobelina Health Center	0.133	0.07	0.243	0.00	0.158	0.01
Constant	0.470	0.00	0.380	0.00	0.729	0.00
N obs	454		440		453	

Table C.5 (Continued)

	Immediate newborn		Immediate maternal postpartum	
	Coef.	p value	Coef.	p value
Group characteristics				
Peers' performance across all deliveries	0.036	0.09	0.023	0.15
Peers are more senior cadre than index	-0.059	0.21	0.004	0.90
Peers are more experienced than index	0.000	0.99	-0.030	0.12
Number of providers	-0.021	0.34	-0.007	0.61
Index provider characteristics				
Years of experience	0.000	0.98	-0.003	0.35
Provider cadre (Midwife or nurse diploma ref.)				
Midwife or nurse Bsc.	-0.073	0.16	0.065	0.03
Health officer	-0.506	0.06	0.043	0.72
GP	-0.176	0.00	-0.025	0.52
IESO	-0.273	0.09	-0.071	0.09
Context and environment				
Delivery complication	0.030	0.34	-0.005	0.87
Higher risk pregnancy	0.003	0.96	-0.006	0.86
Night delivery (morning reference)	0.021	0.36	0.027	0.22
Client wealth index (poorest reference)				
Wealth 2	0.040	0.31	0.015	0.60
Wealth 3	0.110	0.02	0.035	0.30
Wealth 4	0.039	0.37	0.043	0.20
Wealth 5 (wealthiest)	0.084	0.06	0.053	0.06
Facility (Dil Chorra Hospital ref.)				
Sabien Primary Hospital	-0.041	0.42	-0.034	0.29
Biyowale Health Center	0.335	0.00	0.171	0.23
Legeharae Health Center	0.129	0.06	0.122	0.01
Melka Jebdu Health Center	0.089	0.11	0.101	0.15
Wahil Health Center	-0.272	0.00	0.109	0.00
Gende Gerada Health Center	0.217	0.23	0.004	0.97
Goro Health Center	0.178	0.01	-0.003	0.96
Jelobelina Health Center	0.160	0.05	0.106	0.04
Constant	0.778	0.00	0.063	0.34
N obs	387		455	

Table C.6. Comparison of average vs best coworker performance for deliveries with 3 or more providers

	Coworkers' average performance		Best coworker's performance	
	Coef.	p value	Coef.	p value
Group characteristics				
Coworkers' average performance across all deliveries	0.045	0.01		
Best coworker's performance across all deliveries			0.023	0.08
Coworkers are more senior cadre than index	-0.054	0.04	-0.062	0.02
Coworkers are more experienced than index	-0.024	0.27	-0.023	0.29
Number of providers	0.008	0.68	0.003	0.88
Index provider characteristics				
Years of experience	-0.005	0.25	-0.004	0.29
Provider cadre (Midwife or nurse diploma ref.)				
Midwife or nurse Bsc.	-0.053	0.05	-0.06	0.04
Health officer	-0.173	0.34	-0.181	0.33
GP	-0.088	0.02	-0.093	0.02
IESO	0.001	0.99	-0.004	0.97
Context and environment				
Delivery complication	0.015	0.52	0.017	0.47
Higher risk pregnancy	0.018	0.57	0.018	0.58
Night delivery (morning reference)	0.052	0.01	0.052	0.01
Client wealth index (poorest reference)				
Wealth 2	-0.028	0.40	-0.03	0.37
Wealth 3	0.022	0.47	0.017	0.56
Wealth 4	-0.019	0.57	-0.023	0.50
Wealth 5 (wealthiest)	-0.015	0.63	-0.019	0.55
Delivery stage (first exam ref.)				
First stage of labor	-0.01	0.71	-0.011	0.68
Third stage of labor	0.208	0.00	0.208	0.00
Immediate newborn care	0.247	0.00	0.246	0.00
Immediate postpartum care	-0.335	0.00	-0.335	0.00
Facility (Dil Chorra Hospital ref.)				
Sabien Primary Hospital	0.034	0.23	0.044	0.12
Legeharae Health Center	0.052	0.45	0.102	0.11
Gende Gerada Health Center	0.139	0.16	0.162	0.10
Goro Health Center	0.1	0.09	0.141	0.01
Constant	0.453	0.00	0.46	0.00
N	615		615	

Figure C.6. Association between coworker performance and quality by provider cadre rank

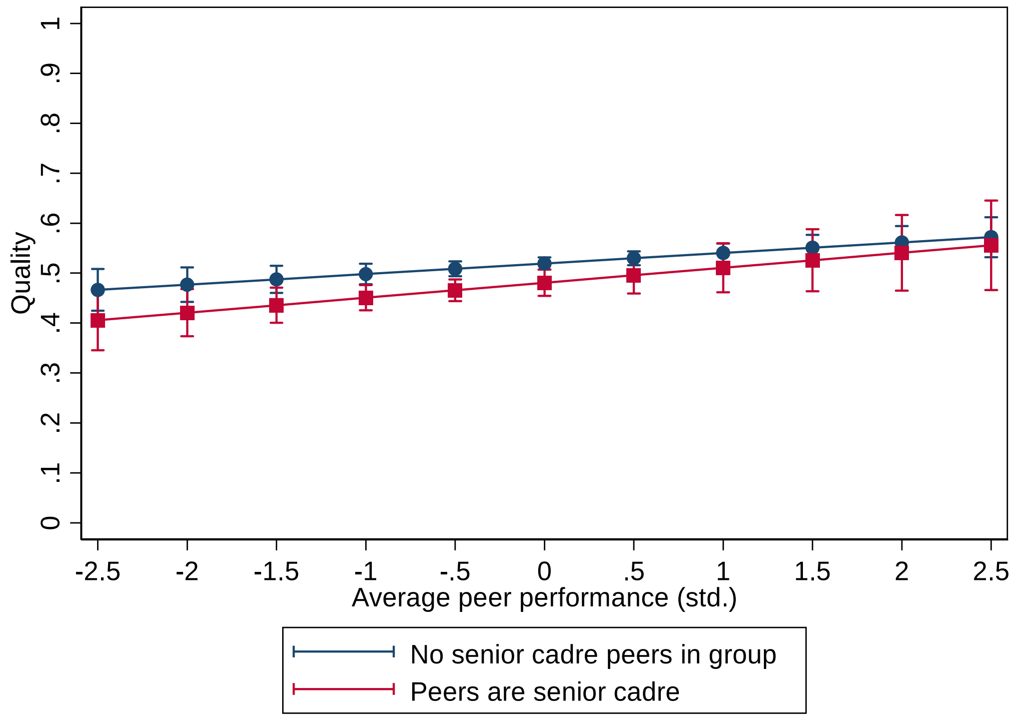


Figure C.7. Association between coworker performance and quality by provider seniority (years of experience)