Increasing the Use of Verification Results within Performance-Based Financing Programs

Trina Gorman

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading committee:

Gabrielle O'Malley, Chair

Bernardo Hernandez

Geoff Garnett

Program authorized to Offer Degree:

Global Health

University of Washington

Abstract

Increasing Data Use and Targeting of Verifications
within Performance-Based Financing Programs

Trina Gorman

Chair of the supervisory committee:
Gabrielle O'Malley
Department of Global Health

Background: Within Performance-Based Financing (PBF) programs, independent verification is a known cornerstone to ensure the reported results are accurate. While verifications typically include reviews of documentation (quantity verification) and patient follow up (community verification), the literature suggests many programs focus payment calculations and data analysis on the quantity verification results, which measure internal record alignment but do not assess the validity of records (e.g. whether records represent delivered services). Further, to reduce the high cost of verifications, there has been a call for risk-based sampling whereby estimated probabilities of overreporting determine both the frequency and intensity of verification efforts. However, there is very little research on the factors that are associated with overreporting, which program planners need to focus sampling in high-risk areas.

Objective: For a Voluntary Medical Male Circumcision (VMMC) PBF program in Zimbabwe that took place from 2016-2018, Aim 1 sought to understand the extent to which reported VMMCs could be verified in records and with patients, the paths and processes that led to the verification results, and if the two sources (records, patients) aligned at the facility-level. For the same program, Aims 2 and 3 both sought to understand the factors associated with overreporting.

Methods: This study used programmatic secondary data from verifications activities, additional data gathered from Zimbabwe's Ministry of Health and Child Care (MoHCC) for explanatory variables, as well as primary data collected through interviews with community verifiers. For Aim 1, descriptive statistics are presented as well as the reasons behind the results from the views of the verifiers'; the correlation between the quantity and community verification performance scores was also assessed. For Aim 2 and Aim 3, generalized mixed effects models were employed each with a different binary outcome to assess the relationship between health facility characteristics and overreporting. Aim 2 was at the facility-level and overreporting was defined based on the quantity verification results. Aim 3 was at the patient-level and overreporting was defined based on the community verification results.

Results: For Aim 1, 25% (36,877/146,924) of reported VMMCs were reviewed and 3,676 patients were interviewed across 41% (144/355) of VMMC locations. We found that quantity verification results were not a good proxy for community verification results and that multiple parties, not just health facility staff, were responsible for data discrepancies—realities that point to the need for our proposed results to action framework. In Aim 2, we found that VMMC locations that were further away from the fixed health facility were less likely to overreport and that each additional staff at the VMMC location was associated with a 73% increase in the odds of overreporting. For Aim 3, we found that patients in the target age range, which were renumerated at a higher price point, were less likely to be interviewed and over two times more likely to be classified as overreported.

Conclusions: Our findings are part of emerging evidence that suggest quantity verifications can portray a misleading and overly positive assessment of reported results tied to payment. Because tying payment to facility records alone risks overpaying for services and misreporting performance, programs should continue investing in community verifications and in-depth analysis of the results. The appropriate actions to take with community verification results often vary based on who was responsible for the issue and the underlying cause. To help navigate these complexities and increase the use of community verification findings, PBF programs should consider using and improving our proposed results to action framework.

Table of Contents

List of Tables

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

Background and significance

Performance-Based Financing (PBF) is payment for pre-determined, verified results (e.g., outputs, outcomes) rather than payment for inputs (medicine, salaries) (1). This approach has become increasingly popular over the last 15 years within funding streams to low- and middle-income countries (LMICs). The largest funding source for PBF is the Health Results Innovation Trust Fund (HRITF) administered by the World Bank, which was started in 2007 and has over $477 million in PBF funding from multiple donors (2). The World Health Organization has also endorsed PBF as a more strategic way to purchase health services compared to input-based approaches (3,4). The theory that underpins these schemes is that they increase the quantity and quality of services through a system of rewards, penalties and results-driven incentives(5,6). Tying payment to specific services and behaviors aligns priorities among funders, governments, and providers, and ensures payment is only disbursed for what was accomplished.

Verification of the reported results is critical to ensure that the rewards and penalties are fair. Without verification, reported results may not reflect reality for two reasons. First, the financial incentives within the payment structure which are intended to nudge actors towards improved performance are also perfectly aligned to entice fraudulent reporting. Second, reporting systems might have unreliable data due to insufficient resources (e.g. training, oversight, data entry time), particularly in LMICs. Verifications typically include reviews of documentation (quantity verification) to assess whether records are complete and patient follow up (community verification) to measure the validity of records (e.g. the extent to which records represent delivered services). While verifications are known to play a key role in PBF schemes, the literature base about the results of these efforts is limited, especially for community verification. This dissertation will advance PBF programming by helping to fill the three gaps outlined below, using results from a verification of Population Services International's (PSI) VMMC program in Zimbabwe.

First, more research is needed to understand the implementation of PBF schemes, particularly in terms of the extent to which patients actually receive the intended services. The evidence-base around the effectiveness of PBF has grown in the past ten years; however, most evaluations are quantitative and lack a thorough explanation of how the outcomes were achieved (through mixed method or process evaluation components). While some researchers have aimed to fill this gap to better understand the implementation (7–11), there has been little focus on the results of community verifications. These results play an important role because they measure the validity of records while reviewing documentation only identifies bookkeeping errors.

Second, more research is needed to inform guidance on risk-based sampling for verifications. Most practitioners agree that PBF should move towards risk-based sampling to reduce costs (8,12–15) given the required resources for verifications often represent a significant portion of program funds (8) (16) (17). At some point, these efforts hit diminishing returns and the allotted resources should be repurposed to providing additional services or for health system strengthening. Risk-based sampling, has been shown to reduce costs (19) however, there is very little research on the factors associated with risk (e.g., overreporting)—factors that program planners need in order to target sampling in the riskiest areas.

Third, thorough analysis comparing community verification results to quantity verification results is needed to help substantiate—or draw into question—a common practice, i.e. where programs focus data analysis and payment calculations on the results from quantity verification. This practice relies on the assumption that documentation represents real patients—an assumption that needs vetting.

Dissertation aims

Using data from verifications performed for a VMMC program implemented by Population Services International (PSI) Zimbabwe, in combination with primary interview data to fill targeted data gaps, the aims of this dissertation are:

- Aim 1 - To understand the extent to which reported VMMCs could be verified in records and with patients, the paths and processes that led to the results, and if the two sources (records, patients) aligned at the facility-level (Chapter 2).
- Aim 2 - To understand the factors associated with overreporting, where overreporting means the number of reported services exceeded the number of services that could be substantiated with facility records (Chapter 3).
- Aim 3 - To understand the factors associated with overreporting, where overreporting means a patient did not plausibly confirm receipt of the reported VMMC. We also assess factors associated with when a patient was not interviewed (Chapter 4).

CHAPTER 2. MOVING FROM RESULTS TO ACTION WITH COMMUNITY VERIFICATION: A CASE STUDY FROM A PERFORMANCE BASED FINANCING PROGRAM IN ZIMBABWE

Trina Gorman, Bernardo Hernandez Prado, Geoff Garnett, Loida Erhard, Taurai Kambeu, Brian Maponga, Jabulani Mavudze, Sinokuthemba Xaba, Getrude Ncube, Gabrielle O'Malley

Abstract

Background: Most Performance-Based Financing (PBF) programs perform community verifications to confirm reported outcomes. However, the literature suggests that the results from these efforts are often not used to hold actors accountable if overreporting is discovered. Instead, many programs tie payment to the results from facility record reviews, which measure internal record alignment but do not assess the validity of records (e.g. whether records represent delivered services).

Objective: We sought to understand the extent to which reported voluntary medical male circumcisions (VMMCs) could be verified in records and with patients, the paths and processes that led to the results, and if the two sources aligned at the facility-level.

Methods: For a VMMC program in Zimbabwe that took place from 2016-2018, we performed a mixed method verification including quantity verification, community verification, and verifier interviews to help understand the community verification findings. We also assessed the correlation between the quantity and community verification performance scores to see whether facilities that have strong record keeping tended to also have strong validation from patients and vice versa.

Results: Of the 36,877 reported VMMCs, 94% were verified with facility records. However, among records selected for community verification, only 55% (2,010/3,676) of patients were interviewed. Among those interviewed, 17% (342/2,010) provided answers that did not plausibly match facility records. Community verifiers reported that some patients admitted providing incorrect contact information to avoid follow-up and most verifiers suspected staff had fabricated data. We found no correlation between quantity and community verification performance scores at the facility level.

Conclusion: Overall, the results from the quantity verification were not a good proxy for the community verification. Programs that pay based on facility records alone risk overpaying for services and misreporting performance. The appropriate actions to take with community verification results often vary based on who was responsible for the issue and the underlying cause. To help navigate these complexities and increase the use of the community verification results, PBF programs should consider using and improving our proposed results to action framework.

Introduction

HIV/AIDS is the leading cause of death in Zimbabwe (20), totaling 20,000 deaths per year (21). Substantial resources have been dedicated to HIV/AIDS prevention, including the expansion of Voluntary Medical Male Circumcision (VMMC) to reduce transmission. The Ministry of Health and Child Care (MoHCC) has leveraged Performance Based Financing (PBF) as part of its VMMC strategy. PBF is payment for pre-determined, verified outcomes rather than payment for inputs (1)—a funding approach that has been increasingly implemented in low and middle income countries, totaling over $19 billion in funding since 2000 (22). The theory that underpins PBF is that aligning provider payment with provided services increases health care worker productivity and service quality (5,6).

Despite potential advantages of PBF schemes, they are continually challenged by weaknesses in the health management information systems (HMIS) that track reported outcomes as well as by the potential for perverse effects (e.g., fraud, gaming, etc.)(15,23,24). Independent verification is critical to ensure reported outcomes, payment, and the perceived health benefit reflect reality (8,25). Verification activities aim to identify and reduce overreporting, which can be either intentional (e.g. workers fabricating data to increase pay) or unintentional (e.g. bookkeeping errors due to insufficient time or training). Typically, PBF verifications include reviews of paper-based facility records (also called quantity verification) to ensure records are complete and interviews with patients (called community verification, client tracing) to ensure records represent delivered services since records could be fabricated (8,25).

While the explicit purpose of community verification is to determine the validity (26) of the paper records, the literature suggests the results from these efforts are often neglected both for programmatic decision making and in the literature. A review identified ten PBF verifications that include quantitative results, including eight from the grey literature (16,27–33) and two that were peer reviewed (8,34). Only seven of the ten articles include community verification (8,16,28,29,32–34) and of those that did, three explicitly report that the results were not used to hold actors accountable for overreporting or fraud (8,28,34). This implies that in six out of the ten verification studies, payments were tied to patient records alone, which is counter to PBF guidance that verifications should "signal to providers that there is a strong chance that one will be caught if one cheats (by claiming phantom patients)" (25). Tying payment to records alone would be less of a concern if there were empirical evidence that facilities having strong records tended to also have strong community verification results and vice versa. However, we found no studies in the published literature that assessed the correlation of these two performance scores.

In addition, the available literature provides little detail on the methods and findings of community verification. Most studies only contain high level quantitative results without details regarding indicator definitions or the processes that led to the results; as an example, only three (16,32,34) of the seven studies (8,16,28,29,32–34) reported on both the portion of selected patients that were interviewed and the portion of interviewed patients whose responses matched facility records.

Measuring the extent to which facility records represent delivered services is essential for the effective implementation of PBF. However, the literature is scant regarding community verification results and the paths and processes that led to them. This study helps to fill this gap by describing a quantity and community verification for a program designed to encourage scaling VMMC in Zimbabwe, including the reasons behind the results from the view of the verifiers.

Methods

Study Setting

HIV in Zimbabwe has been a significant public health concern for many years. The incidence of HIV/AIDS of Zimbabwe in 2019 was 214 new infections per 100,000 people per year - a significant decrease from the 669 new cases per 100,000 people in 2009 (35). The VMMC program in Zimbabwe launched in 2009, following randomized trials that showed circumcising men reduces the risk of female-to-male HIV transmission by up to 60% (36–38).

PBF Program Design

Starting in 2015, the Bill & Melinda Gates Foundation supported Population Services International (PSI) in partnership with the MoHCC to help scale VMMC using PBF. This study focused on VMMCs performed as part of a grant from May 2016 to April 2019, whereby PSI as well as facility staff were remunerated for each VMMC. Initial payments were made based on the reported VMMCs and then adjustments (reductions or increases) were applied as necessary following the ex-post verification activities. At the start of the program, staff in 12 health facilities in 12 districts performed VMMCs in their own facilities and on outreach at VMMC locations; the program was then decentralized to where 32 facilities performed VMMCs across 355 locations and 18 districts.

Verification Design

The verification was based on MEASURE Evaluation’s Routine Data Quality Assessment tool, adapted to the local context (39) and completed at four time points. At each time point verifiers were trained over three days; the verifiers included PSI staff, MoHCC staff, and contractors. For each time point, verification activities were performed between 1-12 months following when the VMMC was reportedly performed.

Quantity verification: Verifiers visited health facilities to quantify the alignment between the number of sufficiently complete facility records (e.g. 10 fields needed to be filled on each patient's form) and the number of reported VMMCs in the HMIS. They recorded the verification results on a paper form and then supervisors entered the results into Excel. A purposeful sample of facility records were selected to fulfill requirements from the funder. The sample included 25% of reported VMMCs and focused on facilities that reported the largest numbers of VMMCs; 41% of VMMC locations (144/355) were selected. Once selected, a census of the facility records that fell into the time period were reviewed.

Community verification: Among facility records that were sufficiently complete, 10% were randomly selected for community verification (2.5% of reported VMMCs). Patients or their guardians (in the case of minors) were interviewed on the phone if possible or at their homes; responses were captured on a mobile phone. They were asked if they were circumcised, and if they were, they were asked three additional questions about the procedure, i.e. the general time period, district, and VMMC method (surgery or prepex device). If their response from one or more of these questions did not plausibly align with data captured from the facility record into Excel, the VMMC was considered unverified.

Community verifier interviews: Structured phone interviews with verifiers who interviewed patients were conducted to learn about their views regarding what caused the verification results. Verifiers who performed less than 10 patient interviews and who only worked during the first verification were excluded given program changes and recall bias concerns. Interviews were performed over Zoom in English. Responses were transcribed in real time into an electronic survey and audio recorded; the recording was used to complete the transcription following the interview. Verifiers were asked to provide Likert scale rankings regarding how frequently they encountered certain scenarios. They also ranked their certainty level regarding whether data fabrication had occurred by patients or VMMC staff, where 0 meant they had no suspicion and 10 meant they were certain. Verifiers were also asked open-ended questions to explain their rankings.

The protocol was approved by the research council of Zimbabwe (approval #MRCZ/E/237) and the MoHCC at all appropriate levels. The University of Washington's human subject's department classified the research as exempt. Both patients and verifiers who participated in interviews provided oral consent.

Analysis

Descriptive results are presented for the quantity and community verification. To assess the correlation between the two performance scores, the percentage difference between the verified VMMCs (e.g., sufficiently complete facility records) and reported VMMCs was calculated for the quantity verification; the same was calculated using the community verification results. Pearson Correlation coefficients were then calculated along with a univariate linear regression model to assess the linear association between the two results. Each result was further defined using a binary measure of overreporting if the reported number of VMMCs exceeded the verified number of VMMCs by 10% or more, a standard in the field (40,41).

Close-ended questions were summarized using Stata (StataCorp. 2021. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC) and qualitative responses were coded using Dedoose software (Sociocultural Research Consultants LLC, Los Angeles, CA) through both deductive and inductive approaches. An initial codebook was developed using a parent code for two general areas of inquiry: reasons why patients were not found and reasons why patient responses did not match the facility data for the four required questions. Nested within these codes, each survey question was assigned a child code and additional grandchild codes were developed inductively through consensus between two coders.

Results

Figure 2.1 summarizes the results from the quantity and community verification. Ninety-four percent (34,713/36,877) of selected records were sufficiently complete (C). Among records selected for community verification (Table 2.1), 55% (2,010/3676) of patients were interviewed (E). Table 2.2 summarizes the reasons for patients that were not interviewed, including that 18% were marked as both unknown and had either insufficient or invalid contact information. Among interviewed patients, 83% (1,668/2010) met the funders requirements to verify the VMMC (G).

In terms of the correlation of the performance scores at the facility-level, we found no statistically significant correlation based on both the Pearson Correlation coefficient (-.034, p=.68) and linear regression (-.018, p=.69). When instead using a binary measure of overreporting with a 10% cutoff, 43% (58/136) of results were aligned (either both overreporting or both not overreporting).

For the verifier interviews, thirty-four community verifiers were interviewed among 37 invited. Those interviewed had completed 76% (1536/2010) of the community verifications. In the rest of this section, we summarize why patients were not interviewed (Step F) and why patient responses did not plausibly verify the VMMC (Step H) from the views of the verifiers.

Figure 2.1. Cascade of sample and results

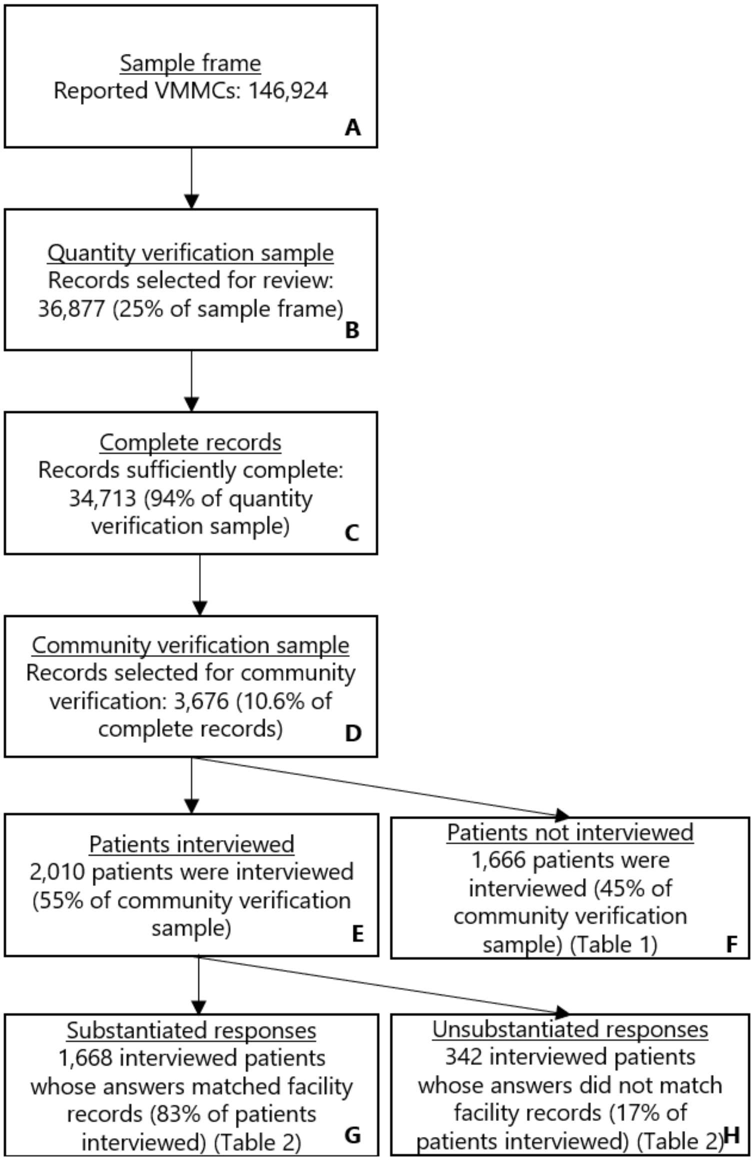


Table 2.1. Description of patients selected for community verification

Description	**Mean (SD) or Proportion**	**N**
Characteristics of patients		
Age of patient in years	17 (0.55)	3676
Method of circumcision		
prepex	8%	3623
surgical	92%	3623

Characteristics of patient's VMMC location		
Type of facility		
district	25%	3676
mission	19%	3676
rural clinic	46%	3676
rural hospital	10%	3676
Type of base funding		
council	27%	3659
government	48%	3659
mission	24%	3659
private	1%	3659

Table 2.2. Summary of reasons patients could not be interviewed

Reason Patient Was Not Interviewed		**n**	**Percent of Selected Patients**
Known in community	... but not available (working, relocated, passed away)	250	7%
Unknown in community	... and insufficient contact information	503	14%
	... and sufficient contact information	495	13%
	... invalid address	143	4%
	Without further explanation	160	4%
Other	Unspecified	28	1%
	Lost interview data	87	2%
Total		1666	45%

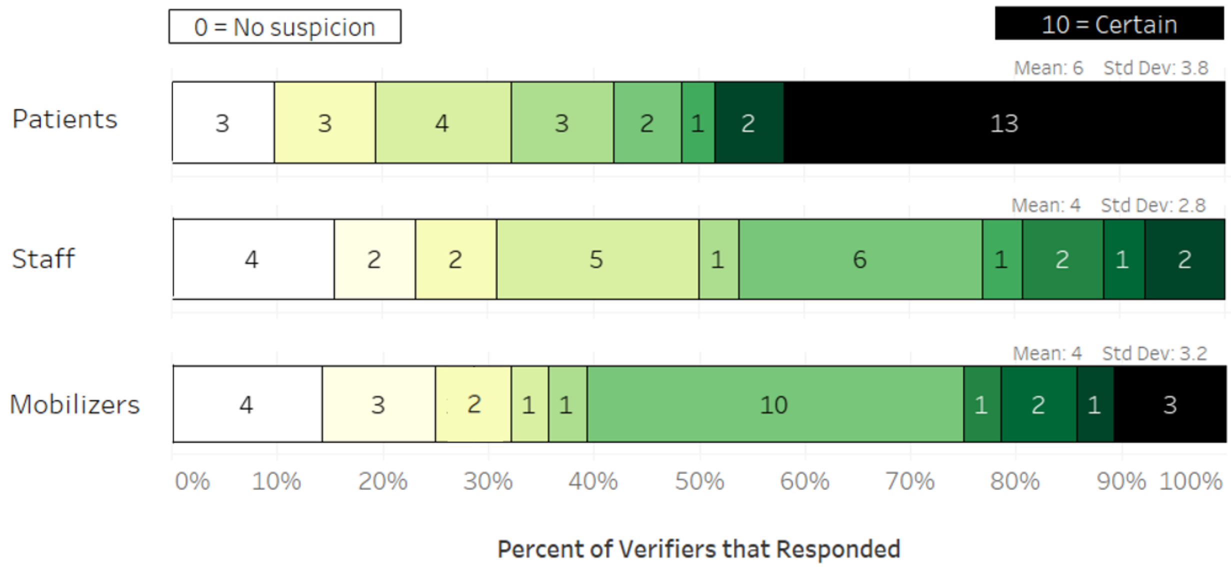
Reasons patients were not interviewed

Verifiers highlighted three main reasons why they thought such a large portion (35%) of patients were reported as unknown in their community (Table 2.2): suspected falsification of data, unintentional bookkeeping errors, and verification design decisions.

Suspected falsification of data

Nearly all verifiers suspected that at least some data on the facility records had been made up by either staff or patients. When asked to rank their certainty level regarding *who* had provided false data, Figure 2.2 summarizes their responses.

Figure 2.2. Verifier's certainty level of falsified data, per group



Nearly half of verifiers reported that patients had admitted providing false/misleading contact information (e.g. nicknames, uncle's name) because they didn't want to be followed-up; their reasons included social stigma, desire for privacy, or because a minor had been circumcised without his parent's consent.

> *I met quite a number of beneficiaries and having a chat with them, they would explain "when I was provided the VMMC service, I actually used the wrong name." The first name was right, but the surname was wrong or vice versa. (ID 109)*

> *In the scenario of students, they were mobilized in the school for circumcision. Then they would go home with the consent form and parents would disagree for them to participate. The kids would sign their own forms with false names, so their parents don't know. (ID138)*

Most enumerators reported having some suspicion that clinicians or mobilizers fabricated data. Verifiers reported their suspicion was primarily due to experiences they had during the field work. For example, some verifiers described seeing questionable records while extracting patient information at facilities (e.g., fields that appeared to have been duplicated, repeated patient signatures, or different ink types).

> *You would find that on some CIRs [clinical intake form] you would see that even the consent forms have been signed by one person, sort of forged the signature. You would see the information is false [because] the same signature was on multiple CIRs. (ID 132)*

The information was similar to another CIR that you looked at. Especially when doing the CIR review - people would call each other and say "it appears that the information was copied from another CIR into another CIR". (ID 121)

Another reason for suspicion was invalid or conflicting contact information from the facility record, such as when the address was an area without dwellings (bus stop, demolished houses) or when the plot number went beyond the possible range.

"This area was demolished (old houses in the colliery) so everyone knows that no one stays in those addresses anymore. But you would find these addresses being recorded as successful male circumcisions. Why would you record this as an address when [you know it] is no longer in use?" (ID 128)

"Like the address 21 Harare Street. It's impossible to have a street in such a rural area." (ID 126)

Staff behavior was also mentioned as an underlying cause of suspicion, such as when mobilizers were not able to help find patients they had reportedly previously mobilized. Four verifiers said a mobilizer intentionally brought them to the wrong patient to try to manipulate the verification results; three verifiers said they saw nurses filling in facility records for VMMCs when they arrived for the quantity verification though there were no VMMC patients at the time.

If you are moving around a village, at least you should be able to locate 70-80% of the people. Once you start going around and no one knows the name or surnames, you start wondering where the mobilizer got the information from to get them for circumcision as they are so busy and the walking distances are so far. (ID 113)

Verifiers said they expected to not find some patients, but many said that the large proportion of unknown patients made them suspicious data had been fabricated. On average, verifiers expected that 15% of the time village leaders and other contacts would not know a person in their village, in contrast to the 35% of patients that were reported as not known in this study.

I was suspicious when we would talk to all the people who should know in a village set up - they have homes close together and should know each other. We would visit the shops etc. and the people are sure the person isn't there. They would refer to someone who knows everyone in the village, and they would say there is no one here by that name. Then we would follow up and ask the mobilizer - he would say "I think he relocated" but it wasn't convincing. (ID 106)

For other verifiers, it wasn't only their experiences from the fieldwork that made them suspicious but also their belief that the financial incentive coupled with economic hardships and unrealistic targets created too great of a temptation.

> *From the home visits and checking their attitude, I would say they were really chasing targets because they are paid per client. So to me, looking at the economic situation they were out there to make money and those clients are not easy to get…. (ID 119)*

Despite their suspicions, most verifiers admitted that they didn't have hard evidence about fabrication by staff.

> *We didn't have tangible evidence - we didn't have mechanisms to follow-up to investigate more on these instances. But we had very strong suspicions... we felt, if something was done - like an investigation - there was a high chance the conclusion or finding would point to instances where the teams had been fixing some figures (ID 121)*

However, three verifiers said they had no suspicion of fabrication by either clinicians or mobilizers.

> *I don't think they provided wrong data because whenever we find them, they would take you straight to the client's home. ... These guys they really knew exactly where the client lives or they would point you in the right direction. (ID 104)*

Unintentional bookkeeping errors

Several verifiers believed that some missing patients were due to unintentional bookkeeping errors, exacerbated by high workloads. A few verifiers recalled asking staff why so many patients could not be found, which helped reduce suspicion for some verifiers but not others.

> *I had to confront the nurses that were responsible because we were failing to locate the clients. She provided insight that maybe at times they were under too much pressure during the day with the VMMC procedures that they wouldn't fill in the documentation…. So, the [reason for the] funny CIR is that the client did exist, but they just failed to fill in the details that would be required. But I still do believe there was some funny stuff going on – they aren't busy all the time with outreach. (ID 114)*

Verification design decisions

Most verifiers said a main cause of not locating patients was a lack of detailed contact information since the program did not require specific addresses; at times a large village or school was all that was listed. Some verifiers mentioned more patients could have been found with more transportation funds and time for the field work.

Reasons interview responses did not plausibly match the facility record

7.5% of interviewed patients (151/2010) were self-reported as not circumcised. Verifiers said these 151 patients commonly reported they had registered for the VMMC but then did not complete the procedure. When verifiers were asked how they thought a complete record was found at the facility for these patients, they explained that it was common for staff to pre-fill facility records with basic information about patients, leaving just a few fields for afterward. Verifiers were generally split regarding whether they thought the remaining fields on these uncircumcised patients' records were intentionally fabricated or filled in by mistake.

Table 2.3 shows the outcome for the remaining 1859 of interviewed patients (excluding those that reported not being circumcised), including 9.5% (191/2010) of patients who reported being circumcised but provided an implausible answer regarding the time period, district, or method (grey cells). 114 patients reported the facility and method plausibly, but not the date; 58 patients reported the facility and date plausibly but not the method. To learn about these 58 patients, verifiers were asked about the survey question that asks patients about the method used in their procedure; nearly one quarter of verifiers thought that patients at times may not have understood the difference between the two methods even after verifiers provided additional explanations. Finally, it is also possible that some verifiers interviewed the wrong patients: when asked whether they completed a survey even when they were not certain that the respondent was the right person, most said never, while four verifiers said rarely or sometimes.

Table 2.3. Reported date, method, and facility from circumcised patients

		Time period not plausible?		Time period plausible?		
		Method plausible?		Method plausible?		
		No	Yes	No	Yes	Total
Clinic plausible?	No	1	6	1	6	14
	Yes	5	114	58	1668	1845
	Total	6	120	59	1674	1859

Discussion and program recommendations

This is the first mixed method study examining a quantity and community verification with detailed findings including the paths and processes that led to them. Overall, we found that quantity verification results were not a good proxy for community verification results for two reasons. First, while the quantity verification results were high (94%), only 45% of participants were reachable for community verification interviews, and among those interviewed, 17% could not be validated. Second, we found no correlation between the two performance scores at the facility-level. The purpose of verification is to detect overreporting by staff and most verifiers' did suspect some data fabrication in this regard; however, verifiers also believed discrepancies were caused by others (PSI, patients, verifiers themselves) and they only had hard evidence about patients fabricating data. These findings support two points of discussion in the literature.

First, verifiers' suspicion that staff fabricated data echoes evidence from a series of studies on the unintended consequences of PBF. Turcotte-Tremblay et al. performed in-depth case studies in Burkina Faso using data gathering techniques inspired by anthropology, such as living in the facilities for two weeks, and found that in two (34,42) out of three studies (34,42,43), there was widespread data fabrication by staff paid per output in order to increase payments. In one study, staff in all six study facilities routinely spent considerable time creating documentation for unperformed services (42). Another study found that when the verifiers tasked with verifying the patient data were paid per patient themselves, they also falsified data in the majority of cases using deliberate and organized strategies (34). In the author's third study, staff still manipulated the program but instead to gain non-financial benefits (43).

In terms of understanding why community verification results are often not analyzed (16,44) or used to inform payment (8,28,34), our findings reinforce a primary reason reported by this same study (34). The authors explain that multiple parties were responsible for data discrepancies, as the verifiers in our study also reported, which made it challenging to interpret the results and agree on appropriate actions (34). For example, a discrepancy between a patient's response and facility data does not necessarily mean overreporting occurred (34). While it could have been caused by staff who recycled patient data (overreporting), it could have instead been caused by a patient's memory lapse or a verifier who interviewed the wrong person (not overreporting, but measurement error).

Given these complexities, a framework may be useful to guide action based on community verification results. Mapping results to actionable recommendations has been identified as an important step in increasing the use of data for program improvement (45,46). Table 2.4 presents a "Results to Action" framework, which organizes the type of result, responsible party, and potential underlying causes as found in our study and other studies (8,34,42,44,47). Below we elaborate on possible follow-up actions, drawing on our findings, the available literature, and the PBF program theory.

Table 2.4. The Results to Action Framework for Community Verifications

Type of Result	Responsible Party	Potential Underlying Cause	Possible Follow-up Actions	
			How to Avoid This Underlying Cause from Occurring?	**Apply Sanctions if Underlying Cause is Unknown?**
Patients not reachable for an interview	Health facility staff	Intentional data fabrication or fraud (e.g., ghost patients)	- Perform verifications and apply sanctions using the results - Ensure sufficient staffing; pay and train staff well	Only when programs can estimate the portion of missing patients that are likely to be due to the patient and verification team, or when programs are confident that this portion is low.
		Unintentional bookkeeping errors		
		Insufficient contact information collected from patients (e.g., typos, nicknames, vague addresses, patients with the same name, too few contact options)	- Collect detailed contact information & references	
	Patient	Insufficient contact information provided either intentionally (e.g., patients avoiding being found) or unintentionally (e.g., wrong phone number, nicknames, name changes over time)	- Avoid tying payment to services associated with stigma or services that are very personal	
		Unavailability either unintentionally (e.g., transient populations/visitors, working class, moved) or intentionally (e.g., patients avoiding survey attempts)	- Avoid tying payment to services used by transient populations - Schedule verification interviews when patients are likely to be home (e.g. not during working hours)	
	Verification team	Lack of effort to find patients	- Ensure sufficient staffing; pay and train verifiers well	
		Insufficient resources (e.g., limited transportation funds, time restrictions)	- Invest sufficient resources	
Patients whose responses do not plausibly match health facility data	Health facility staff	Intentional data fabrication or fraud (e.g., recycling of data from real patients)	- Perform verifications and apply sanctions using the results - Ensure sufficient staffing; pay and train staff well	Yes, this is imperative to uphold the accountability structure
		Unintentional bookkeeping errors		
	Patients	Lack of truth telling (e.g., stigma, concern for privacy, social desirability)	- Avoid tying payment to services associated with stigma or services that are very personal	
		Memory lapse	- Avoid tying PBF payment to services people are likely to forget - Avoid selecting patients for verification that are not likely to remember the services (e.g the elderly) - Ensure verifiers gain rapport	
		Misunderstanding the interview questions	- Avoid selecting patients for verification that are not likely to understand the questions (e.g the elderly); hire experienced verifiers; train/pilot thoroughly	
	Verification team	Data falsification (e.g. a patient's responses are fabricated)	- Ensure sufficient staffing; pay and train verifiers well	
		Data entry errors when extracted from the facility		
		The wrong patients are interviewed		
		Questions are asked incorrectly		
		Data entry errors when patient responses are transcribed		
		Data from patients are incorrectly matched to data from facilities		

Three primary recommendations are proposed in this framework for future research and program implementation. First, it is important that PBF programs reduce payment when a patient's response does not plausibly match the facility record. The theoretical foundation for PBF is that it helps solve the principal-agent problem through paying the agent contingent on verified outcomes—so that the principal pays only for what has been achieved (22,48,49). The entire purpose of the verification is therefore to measure *achievements* (25). If community verification results are not used, it weakens the power of the incentive (8); there is virtually no way to identify cheating based on records alone (43). When it is not possible to verify whether patients received services, PBF may not be optimal in that setting (25).

Second and in contrast, programs should be cautious about reducing payment for patients who are not found for interviews unless there are good estimates regarding the portion of errors for which patients and verification teams are likely to be responsible. While overreporting could be disguised as patients who were not found for interviews, our findings align with past studies showing it's common for verifiers to have insufficient resources to find all patients (8,34,44), and for patients to avoid follow-up (29) or be away when verifiers visit (32–34). Reducing pay without sound estimates regarding these occurrences risks unfairly penalizing the program.

Third, programs need to make sure that verifiers produce data that is reliable by ensuring teams are well trained, paid sufficiently and on time, and have reasonable oversight (44). Evidence from various fields including academia—from low and middle income countries (50–52) to high income countries (53,54)—has shown that data quality suffers when accountability and incentive structures are weak or perverse. Counter to this, the same Burkina Faso study mentioned earlier paid verifiers on a per patient basis only when patient responses matched the health facility data (34); verifiers were also not paid if a patient was unavailable but had not permanently moved. Unsurprisingly, the verifiers fabricated data to increase their pay. Lack of adequate pay

(8,34), training(44), and time for verifiers (34,44) has been reported by many community verifications. While there is increasing focus on avoiding perverse incentives for PBF staff (55–59), programs cannot lose sight of the importance of verifier incentives as well.

This study has several limitations. First, the selection of facilities for quantity verification was purposeful, and as a result, so were the patients selected for follow-up. Most notably, facilities that performed fewer VMMCs were not included so the findings might not generalize to those areas. The low portion of patients that were interviewed is important to understand challenges with finding patients, and yet reduces the internal validity of findings from interviewed patients. Our interviews with verifiers only included their views; additional interviews with VMMC staff would have strengthened our understanding of the context. Further, VMMC is a sensitive, stigmatized topic so findings might not translate to other programs where patients are more open to follow-up. Finally, interviews with the verifiers covered sensitive topics about data fabrication, so are subject to a set of biases. Of most concern was social desirability bias and recall bias, which we mitigated with a thorough informed consent, sequencing of topics, and repeatedly removing any pressure to answer.

Conclusions

We found that when verifying reported VMMCs in a PBF program in Zimbabwe, the results based on the completeness of facility records were not a good proxy for accuracy when compared to findings from patient interviews regarding whether the services were actually delivered. While verifications are complex in that inaccurate results can be caused by many parties, reducing payment in cases when patient responses do not match facility records helps ensure programs are paying for *achieved* results and that there is a real threat of sanctioning for intentional overreporting—two core theoretical requirements for PBF.

CHAPTER 3. FACTORS OF OVERREPORTING BASED QUANTITY VERIFICATION RESULTS IN A VMMC PBF PROGRAM IN ZIMBABWE

Trina Gorman, Pia Arce, Gabrielle O'Malley, Taurai Kambeu, Brian Maponga, Jabulani Mavudze, Sinokuthemba Xaba, Getrude Ncube, Bernardo Hernandez Prado

Abstract

Background: Performance-based financing (PBF) in low- and middle-income countries has generated considerable funding and interest in the last 15 years. Verification of the reported results is essential given the incentives intended to increase productivity could also entice intentional overreporting. To verify reported results, programs typically include reviews of facility records (quantity verification) and interviews with patients (community verification), both of which produce different but complimentary measures of overreporting. Unfortunately, these activities are often expensive given the large number of records and patients commonly verified. A less expensive approach is to verify a smaller sample in areas expected to have a higher risk of overreporting, called risk-based sampling. To do so, program planners need to know which factors are associated with overreporting to ensure they are sampling the right areas, but there is very little research on such factors.

Objective: The purpose of this study is to understand the factors associated with overreporting in a PBF setting, where overreporting means the number of reported services exceeded the number of services that could be substantiated with facility records.

Methods: Using quantity verification results for a Voluntary Medical Male Circumcision (VMMC) program in Zimbabwe that took place from 2016-2018, a generalized mixed effects model was employed. Missing data were addressed through bootstrapping-based expectation-maximization multiple imputation and results were weighted using inverse probability of treatment to account for differences between the selected VMMC locations and all VMMC locations in the program.

Results: 144 VMMC locations were included in the model. After adjusting for various facility-level and district-level variables, VMMC locations that were further away from the fixed health facility were less likely to overreport; each 10 kilometers in distance was associated with a 29% decrease in the odds of overreporting (OR: 0.71, 95% CI: 0.56-0.90). Each additional staff at the VMMC location was associated with a 73% increase in the odds of overreporting (OR:1.73 95% CI: 1.22-2.45).

Conclusion: To help reduce the cost and better target quantity verifications, we identified factors associated with overreporting for reported VMMCs in Zimbabwe including that distance from the fixed health facility and additional staff were significant determinants. While this study was exploratory given it was the first of its kind in this context, programs should consider employing similar methods to inform risk-based sampling decisions. We also recommend qualitative research to understand the role of higher staff workloads on overreporting, as opposed to catchment and staff size separately.

Introduction

Performance-based financing (PBF) programs aim to align payment with provided services in order to increase both the productivity of health care workers and service quality. Independent verification—including reviews of paper-based facility records (also called quantity verification) and interviews with patients (called community verification, counter verification, or client tracing)(60)—is a cornerstone of PBF programs to ensure the reported outcomes are accurate. Inaccuracies can be either intentional (e.g. data fabrication) or unintentional (e.g. bookkeeping errors) but in either case, implementers use verifications to identify and reduce overreporting, which is when the number of reported services exceeds the number of services delivered to patients.

Programs commonly store the aggregated reported outcomes in a health management information system (HMIS). When assessing overreporting within the HMIS, the source data used to compare to the aggregated totals is important since different sources will vary in terms of the amount and type of potential measurement error. For quantity verifications, overreporting is estimated *based on facility records* (e.g., when the total number of services in the HMIS exceed the number of services substantiated in records). For community verifications, overreporting is estimated *based on patient responses* (e.g., when a patient's report of a service does not plausibly match the facility record). Assessing both types of overreporting is important; quantity verifications assess whether records are complete and community verifications measure the validity of records (e.g. the extent to which records represent delivered services).

As PBF programs have grown in the last 15 years, verifications have received criticism for their high cost, which have been reported as high as 16% of all program costs in Burundi (16) and 23% in Afghanistan (17). Such high costs are in stark contrast to general guidelines that M&E funding should range between 5-10% of program costs (61). Further, from a utilization focused evaluation perspective (62), PBF funders and implementers need to be able to justify the opportunity costs of funds spent on verification rather than implementation – a challenge demonstrated by a program in Benin, where for each 1 USD paid to health providers, around 0.50 cents was used for verification efforts (8).

High costs are often due to the large number of records and patients verified. Traditionally within quantity verification, programs sample a large portion—if not all—facility records for review and among them, patients are selected for community verification. Sampling methods vary, but most aim for a representative sample within resource constraints (60). An alternative approach is to verify a reduced number of records and patients, targeted in areas expected to have a higher risk of overreporting (risk-based sampling), an approach that has been increasingly recommended (8,12–15). As an example of the potential for cost savings, a maternal and child health (MCH) program implemented by Cordaid in Zimbabwe reduced verifications costs by 47% by implementing this approach (19). To implement risk-based sampling, program planners need to know which factors are associated with overreporting to ensure they are sampling the right areas, but there is very little research on such factors. In terms of factors associated with poor data quality based on facility records, while there are studies that speculate about potential factors in PBF settings (8,14,16,28–30,63,64), there is only one study that has assessed such factors quantitively (65) but instead focused on misreporting, which includes both overreporting and under—the latter being of less interest to funders since it does not result in overpayment (41). Additional papers studied the best machine learning methods, not factors (41); we also studied factors associated with overreporting for this same program (Chapter 4), but instead using community verification results as the data source.

Our research contributes to this neglected topic by examining factors that are associated with overreporting, where overreporting is based on quantity verification results. We use data from a PBF program that implemented Voluntary Medical Male Circumcision (VMMC) from 2016-2018 in Zimbabwe, a program that aimed to reduce the spread of HIV, one of Zimbabwe's primary health concerns over the last decades.

Methods

Study Setting

In Zimbabwe, the incidence of HIV/AIDS in 2019 was 214 new infections per 100,000 people per year - a significant decrease from the 669 new cases per 100,000 people in 2009(35) . This progress was in part due to the Ministry of Health and Child Care's (MoHCC) adoption of VMMC as a priority prevention strategy, following randomized trials that showed circumcising men reduces the risk of female-to-male transmission by up to 60% (36–38).

PBF Program

Program Design

Starting in 2015, the Bill & Melinda Gates Foundation supported Zimbabwe's VMMC program through a series of grants to Population Services International (PSI). This study focuses on a grant from May 2016 to April 2019. VMMC staff performed the procedures at VMMC locations, which were either fixed health facilities or locations staff traveled to during outreach. During busy times such as campaigns, staff would travel outside of their typical catchment to help facilities expected to be short staffed. Mobilizers were assigned to districts to sensitize communities, create demand, and enroll interested patients. Mobilizers were categorized as gold or silver based on their performance. PSI as well as VMMC staff and mobilizers were remunerated for each VMMC performed, including a higher amount for patients between 15 to 29 years, the target age range. To document the VMMCs, clinicians filled in a facility record, consisting of a single paper form for each patient plus a separate informed consent. Documented VMMCs from each location were aggregated monthly and then entered into the HMIS. Initial payments were made based on the number of reported VMMCs and then adjustments were applied as necessary following the ex-post verification activities outlined below.

Verification Design and Sample

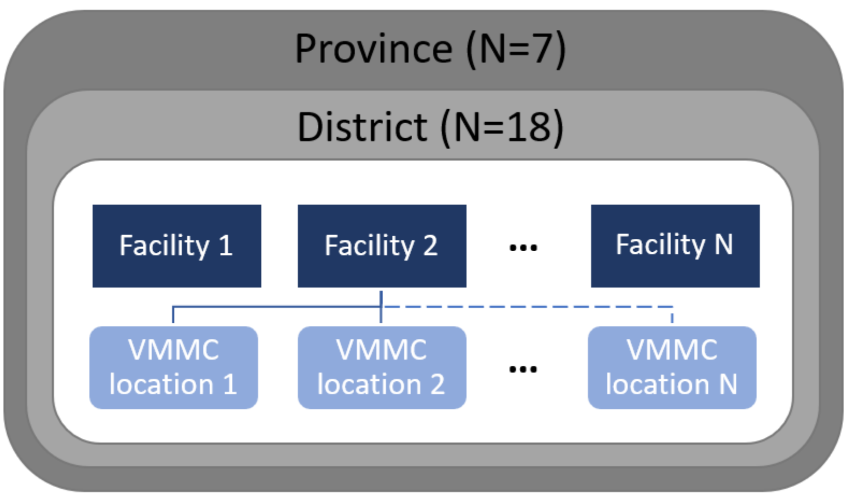
The verification activities were based on MEASURE Evaluation's Routine Data Quality Assessment tool(39), adapted to the local context and completed at four time points, totaling two rounds of verification for each district in the program. At each time point, verifiers (PSI staff, MoHCC staff, and contractors) were trained over two days to perform two activities. First, the teams performed a quantity verification, where they visited health facilities to quantify the alignment between the number of sufficiently complete facility records and the number of reported VMMCs in the HMIS. To be sufficiently complete, the funder required that 10 specific fields were filled out on the form. Verifiers recorded the results on paper and then supervisors entered the results into Excel. Second, the teams performed a community verification, where they visited patients to quantify the alignment between the number of VMMCs that could be confirmed as received and the number of sufficiently complete facility records from the quantity verification. The results from the community verification are outside of the scope for this analysis, and have been summarized in Chapter 2 and Chapter 4.

Among reported VMMCs in the HMIS, PSI purposefully sampled facility records in order to fulfil the funder's requirement, which was to review the records of 25% of reported VMMCs. To select the sample, the team divided the districts in half; VMMC locations in the first half could be selected for the first and third verifications, and VMMC locations in the second half could be selected in the second and fourth verifications. Before each verification, between 3 and 9 months of VMMCs prior to the verification start date were selected in order to be on track to reach the 25% requirement. Within the selected months, VMMC locations that reported above 50 VMMCs and were from the relevant districts were selected for review.

Dataset for this Study

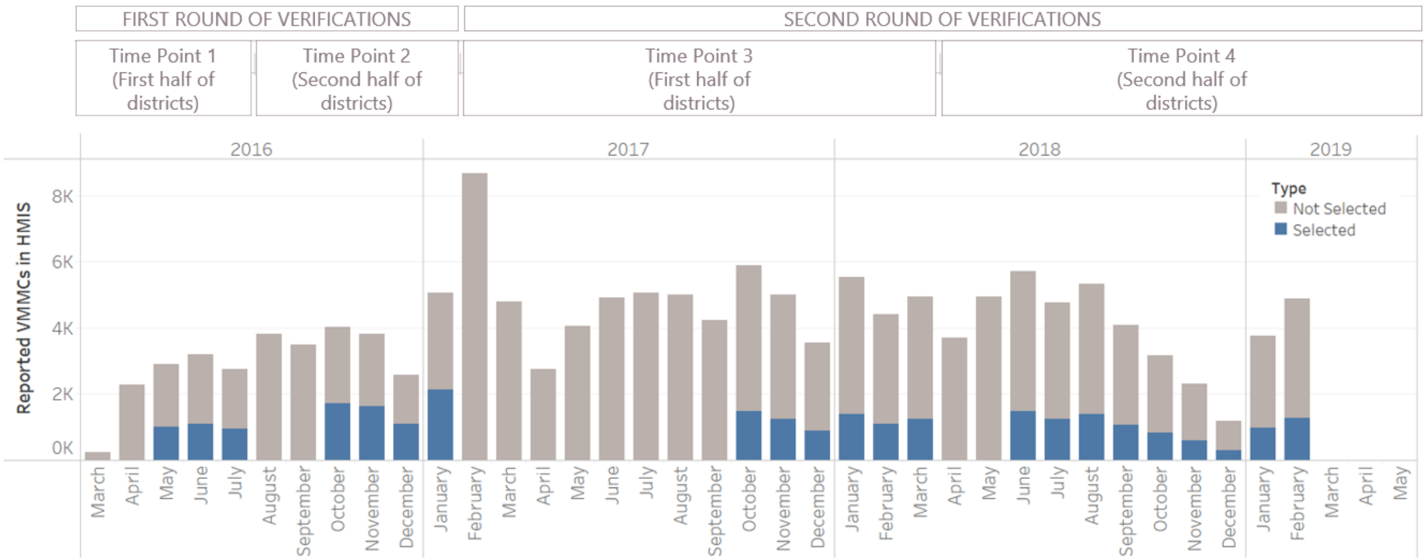
This study uses secondary data from the program activities described above and additional data gathered from the MoHCC. The program provided a dataset where each row represented the quantity verification results from one VMMC location at one time point. VMMC locations were nested within the fixed health facilities whose staff performed the procedures, districts, and provinces (Figure 3.1).

Figure 3.1: Data structure



Some VMMC locations in the original dataset were selected at two time points but these were dropped for this analysis to eliminate bias caused by the influence of the VMMC location being verified a second time, though the dataset contains some repeated measures at the health facility level. Figure 3.2 shows the resulting portion of records that were reviewed in the dataset used in this analysis, which included outcomes from 31/31 fixed health facilities and 144/355 VMMC locations.

Figure 3.2: Summary of data collection and sample for this analysis



For each VMMC location, there are two sources of data. First, the results of the quantity verification from the program, including the number of VMMCs that were reported and the number of VMMCs that were sufficiently complete. Using these data, a binary outcome of overreporting was calculated as follows. First the percentage difference between the number of records that were sufficiently complete and the number of VMMCs that were reported in the HMIS was calculated. Results were then classified as overreporting if the reported number of VMMCs exceeded the verified records by 10% or more, a standard in the field. The second source of data includes health facility and location-level characteristics that were used as explanatory variables (Table 3.2). A list of variables of interest was generated based on the literature, and then PSI and the MoHCC collected those that were available from district-level VMMC officials. The MOHCC also provided GPS locations, which were used to calculate the linear distance between the VMMC location and the health facility. Based on the literature, we hypothesized that the number of staff and supervisory visits would be negatively associated with overreporting and that the number of reported VMMCs and catchment populations would be positively associated (29,66).

Table 3.1: Summary of explanatory variables

Level	Variable
District	Largest number of gold mobilizers that were active at one time during the grant period
	Largest number of silver mobilizers that were active at one time during the grant period
	Largest number of VMMC vehicles operating at one time during the grant period
Health Facility	Number of VMMCs reported in the HMIS for the health facility at all VMMC locations throughout the grant period
	Largest number of trained clinical VMMC staff (nurses & doctors) that were employed at one time during grant period
VMMC Location	Number of supervisory visits provided by the MOHCC
	Size of eligible population in catchment area
	Distance in kilometers from the VMMC location to the health facility based on GPS coordinates.
	Type of facility (e.g., primary, secondary, tertiary, central care)
	Type of base funding (government, local council, mission, private) of the location as a whole

Analysis

To identify the factors associated with overreporting, a generalized linear mixed effects model was used including a random intercept for the health facility and the district to account for the nested data structure. The model used the binomial family and logit link. To account for differences between the selected VMMC locations and all VMMC locations in the program, the results were weighted with Inverse Probability of Treatment by Logistic Regression (IPTW) using the only two variables that were available for all locations: the district and the total number of VMMCs reported by the location. The weights were stabilized to reduce the variance of the effect estimate and trimmed at 1% and 99%. To address missingness in two variables (catchment population was missing for three locations, number of supervisory visits was missing for ten locations) bootstrapping-based expectation-maximization multiple imputation was used. After imputing five datasets, regressions were run on each imputed dataset. The coefficients and standard error estimates were then pooled using Rubin's rules (67). More details about the weighting and imputation can be found in Annex I and Annex II. All models were also estimated for the subset of locations that had complete data (complete case analysis) to test the sensitivity of the findings to missing data.

All analyses were performed in R version 2022.12.0 (R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.) or STATA version 17.0 (StataCorp. 2023. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC.). The Amelia II package version 1.8.1 was used for imputation(68).

Ethical Approval

The protocol was approved by the research council of Zimbabwe (approval # MRCZ/E/237) and the MoHCC at all appropriate levels. The University of Washington's human subjects department classified the research as exempt.

Results

Characteristics of VMMC locations selected for quantity verification

The sample of VMMC locations is summarized in Table 3.2; Annex III contains the same restricted to the 134 locations that had complete data. The fixed health facilities whose staff performed the procedures reported an average of 5,714 VMMCs reported across all locations; facilities had 5 VMMC staff and 8 mobilizers (4 of each type) on average. These facilities shared on average 1.7 vehicles across the district. The catchment population of each VMMC location ranged from 590 to 156,842 with an average of 8,669. VMMC locations were often a rural clinic (69%) funded by the government (49%). Among all location-level results, 31% (41/144) were classified as overreported with an average percentage difference of -6.41 and median of -3.74 (IQR: -12.2 to -.75).

Factors associated with overreporting

Table 3.3 presents findings for factors associated with overreporting. The three facilities with private funding had no variation in the outcome so were not included in the models. After controlling for various facility and district-level variables, the number of VMMC staff was positively associated with overreporting;

each additional staff member was associated with a 73% increase in the odds of overreporting (OR:1.73 95% CI: 1.22-2.45). Distance between the VMMC location and the fixed facility was negatively associated; each 10 kilometer increase in the distance was associated with a 29% reduction in the odds of overreporting (OR: 0.71, 95% CI: 0.56-0.90). The analysis of VMMC locations with complete data showed that the size of the catchment population was negatively associated with overreporting, but this was not significant when imputed data were included. Other explanatory variables were not significantly associated.

Table 3.2: Description of VMMC locations

	Mean or Proportion	SD	Min	Max	Observations
Reported VMMCs of the facility	5714.70	3844.40	22	14829	144
Number of staff	5.69	2.25	2	10	144
Number of gold mobilizers	4.22	0.96	1	6	144
Number of silver mobilizers	4.14	4.27	0	15	144
Number of vehicles	1.70	0.75	1	3	144
Population of location	8658.60	18309.70	590	156842	141
Distance in kilometers between the VMMC location and health facility	27.13	24.83	0	136.9	144
Supervisory visits	2.40	2.03	0	10	134
Type of facility*					
District	0.15	0.36	0	1	144
Mission	0.08	0.28	0	1	144
Rural clinic	0.69	0.46	0	1	144
Rural hospital	0.07	0.26	0	1	144
Type of base funding*					
Council	0.35	0.48	0	1	144
Government	0.49	0.50	0	1	144
Mission	0.13	0.34	0	1	144
Private	0.02	0.14	0	1	144

* Presents the proportion as each classification is represented with a binary (0/1) variable.

Table 3.3: Logit regression – Adjusted odds ratio for overreporting defined by facility records

	Results with Imputed Values (N=141)					Complete Case Analysis (N=131)				
	Adj.Odds ratio*	Robust std. err.	z	[95% conf. interval]		Adj. Odds ratio*	Robust std. err.	z	[95% conf. interval]	
Reported VMMCs of the facility (Log)	1.28	0.70	0.45	0.44	3.75	1.45	1.10	0.49	0.33	6.38
Number of staff	1.73	0.31	3.09	1.22	2.45	1.66	0.29	2.95	1.19	2.33
Number of gold mobilizers	1.20	0.50	0.45	0.53	2.72	1.31	0.49	0.72	0.63	2.72
Number of silver mobilizers	1.01	0.09	0.17	0.85	1.21	1.08	0.08	0.95	0.92	1.26
Number of vehicles	1.79	0.86	1.21	0.70	4.61	1.37	0.55	0.79	0.63	2.99
Population of location (Log)	0.62	0.19	-1.53	0.34	1.14	0.46	0.17	-2.05	0.22	0.96
Distance of 10 km between VMMC location and fixed health facility	0.71	0.09	-2.82	0.56	0.90	0.72	0.09	-2.65	0.57	0.92
Supervisory visits	1.29	0.19	1.72	0.96	1.72	1.30	0.22	1.57	0.94	1.82
<u>Type of facility</u>										
District	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Mission	1.15	1.26	0.13	0.13	9.88	1.04	1.16	0.04	0.12	9.16
Rural clinic	0.55	0.55	-0.59	0.08	3.94	0.32	0.33	-1.11	0.04	2.40
Rural hospital	0.50	0.45	-0.77	0.08	2.95	0.39	0.44	-0.83	0.04	3.61
<u>Type of base funding</u>										
Council	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Government	0.49	0.38	-0.93	0.11	2.20	0.49	0.48	00.73	0.07	3.29
Mission	0.36	0.24	11.54	0.10	1.31	0.31	0.22	11.64	0.07	1.26
Private	1.00	(omitted)				1.00	(omitted)			

Note: All odds ratios are adjusted by all the variables in the table.

Discussion

To our knowledge, this study is the first to assess factors associated with overreporting in a PBF setting, where overreporting means the number of reported VMMCs in the HMIS exceeded the number that could be substantiated with facility records. We found that locations that were further away from the fixed health facility were less likely to overreport and locations with more staff were more likely to overreport. Considering the complete case analysis, larger catchment areas were associated with a decrease in overreporting.

In general, the latter two findings are counter to our hypotheses based on the literature (65). We expected more staff to be associated with less overreporting, not more, because more staff might translate into more time to ensure facility records were complete. Further, we expected larger catchment areas to be associated with more overreporting, not less, because a larger populace would require more staff time to complete procedures and fill out records, resulting in more mistakes. One possible explanation is that the relationship between the two characteristics and overreporting is instead based on the interaction between the two, or, a measure of how busy the clinic is. Presumably, a larger catchment population would only translate into less time for record keeping if the location did not also have more staff to account for the increased workload. This theory aligns with the views of verifiers interviewed as part of another study from this same program who believed that unintentional bookkeeping errors were exacerbated by high workloads (Chapter 2). We were not able to further assess this in our analysis since staff provided support in different facilities and districts during busy times such as campaigns. However, future studies should assess the influence of varying workloads on data quality within the context of PBF.

We also found that overreporting was lower when VMMCs were performed farther from the fixed health facility. It's possible that this factor is also confounded by the omission of a measure for workload, i.e., if locations that were farther away tended to be less busy.

To our knowledge, Kuunibe et al performed the only other published analyses that was similar to ours (65). They studied factors associated with misreporting in a PBF program in Burkina Faso across 20 variables (excluding VMMC) from various health sectors. In their results, they also found the same negative association between larger catchment areas and misreporting in six out of the twenty variables; the number of staff and distance were also associated for a few variables but without consistent directions of association. It's expected to some degree that our findings would differ given differences in the outcome and variables but the inconsistency in their results could also support the importance of facility workload.

This study has several limitations. The selection of reported VMMCs was a purposeful sample of facilities and months, so may not generalize to the entire program. Most notably, lower producing facilities were not included, which could explain why we did not find a significant association with the number of reported VMMCs and our outcomes. Further, due to insufficient sample size in our unit of analysis, we could not control for dimensions of time, so seasonal trends and the influence of repeated quantity verifications on overreporting could be confounding factors.

Conclusion

To help reduce the cost and better target quantity verifications, we identified factors associated with overreporting for reported VMMCs in Zimbabwe including that distance from the fixed health facility and additional staff were significant determinants. While this study was exploratory given it was the first of its kind in this context, programs should consider employing similar methods to inform risk-based sampling decisions. We also recommend qualitative research to understand the role of higher staff workloads on overreporting, as opposed to catchment and staff size separately.

Annex

Annex I: Weighting Summary

In this study, VMMC locations were selected purposefully, including the exclusion of VMMC locations that reported low numbers of VMMCs. The purpose of weighting is therefore to account for differences between the selected VMMC locations and all VMMC locations in the program. Inverse probability of treatment weighting (IPTW) was used including four steps. First, a logistic regression was run using the outcome of whether the VMMC location was selected for the verification or not (1/0) and the only two independent variables that were available for all VMMC locations: the district and the total number of VMMCs reported by the location. Predictions for the probability of being selected (prob) for each VMMC location were then obtained based on the logistic regression results. Second, raw weights were calculated using the inverse of the probability of being selected (1/prob); this increases the weight of VMMC locations that had low probabilities of being selected, since they were originally misrepresented in the sample. Applying these weights to the study population creates a pseudo population in which confounders are equally distributed across selected and unselected groups (69). Third, to reduce the effect estimate's variance (69,70), the raw weights were stabilized by multiplying each weight by the proportion of VMMC locations that were selected among all VMMC locations. In the fourth step, the weights were trimmed at 1st and 99th percentiles to avoid the presence of extreme observations. Figure 3 summarizes the distribution of weights after each step and Figure 4 shows how the weights corrected the data, by comparing the unweighted and weighted distributions for an example variable (total number of VMMCs reported).

Figure 3.3: Distribution of weights

Distribution of raw weights | Distribution of stabilized weights | Distribution stabilized and trimmed weights

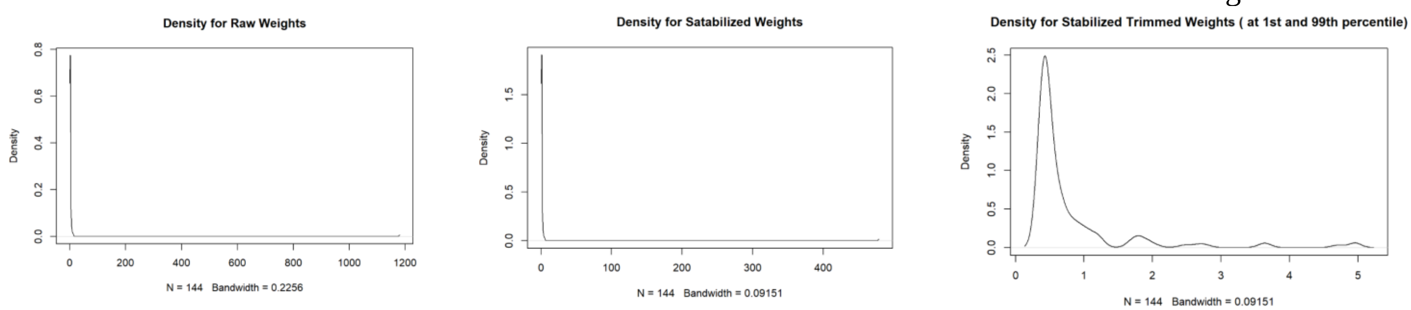
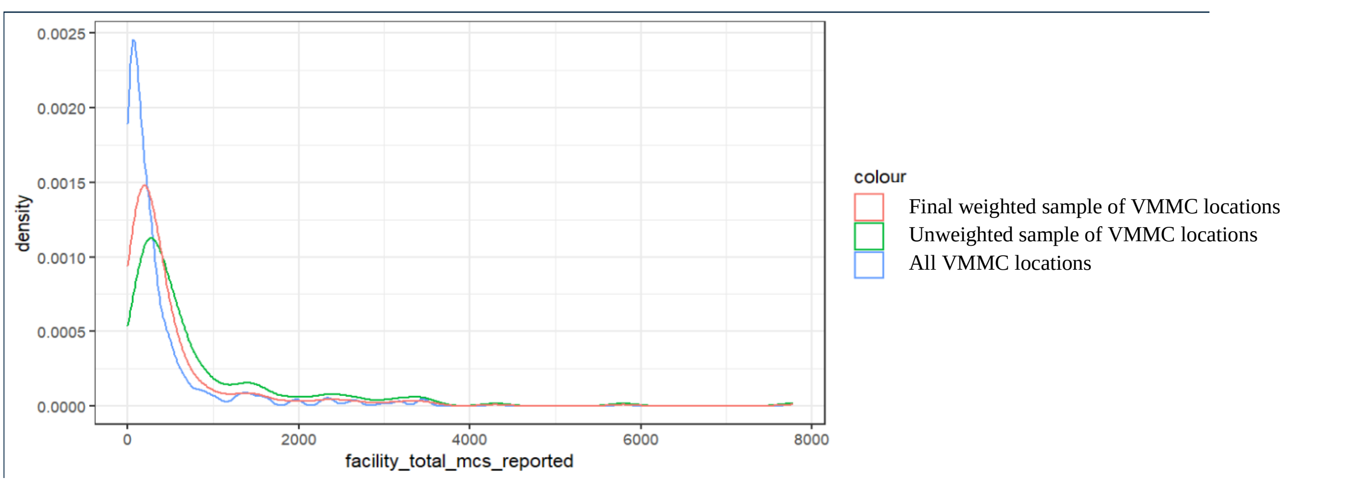


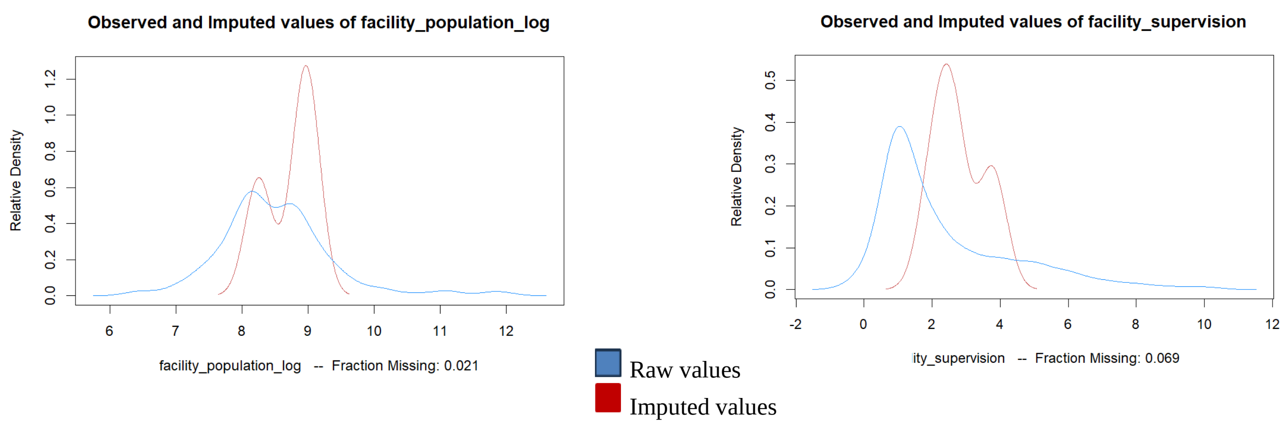
Figure 3.4: Distribution of VMMCs reported, before and after weighting



Annex II: Imputation Summary

The multiple imputation tool called Amelia II was used to impute missing observations. Amelia II uses a bootstrap-based algorithm where *m* imputations create *m* complete datasets, each with different imputed results. Amelia II considers the distribution of all other input variables and identifiers—aiming to not change any relationships in the data as it fills in missing values. In this study to impute estimates for the catchment population (n = 3) and the number of supervisory visits (n = 10), we employed 5 iterations using the same variables that are in the study model. The Stata mi command was used to pool the coefficients and standard error estimates from the 5 datasets using Rubin's rules. Figure 4 shows the distribution of the imputed values relative to the raw dataset.

Figure 3.5: Distribution of two variables, comparing the raw to imputed values



Annex III: Description facilities in the complete case analysis

Table 3.4: Description facilities in the complete case analysis

	Mean or Proportion	SD	Min	Max	Observations
Reported VMMCs of the facility	5771.9	3794.0	22.00	14829.0	131
Number of staff	5.756	2.317	2	10	131
Number of gold mobilizers	4.237	1.006	1	6	131
Number of silver mobilizers	4.252	4.352	0	15	131
Number of vehicles	1.664	0.761	1	3	131
Population of location	8286.4	18221.5	590.0	156842.1	131
Distance in kilometers between the VMMC location and health facility	27.72	25.33	0	136.9	131
Supervisory visits	2.351	2.011	0	10	131
Type of facility*					
District	0.15	0.36	0	1	131
Mission	0.092	0.29	0	1	131

Rural clinic	0.69	0.46	0	1	131
Rural hospital	0.061	0.24	0	1	131
Type of base funding*					
Council	0.36	0.48	0	1	131
Government	0.50	0.50	0	1	131
Mission	0.15	0.35	0	1	131
Private	0	0	0	0	131

* Presents the proportion as each classification is represented with a binary (0/1) variable.

CHAPTER 4. FACTORS OF OVERREPORTING BASED ON COMMUNITY VERIFICATION RESULTS IN A VMMC PBF PROGRAM IN ZIMBABWE

Trina Gorman, Pia Arce, Gabrielle O'Malley, Taurai Kambeu, Brian Maponga, Jabulani Mavudze, Sinokuthemba Xaba, Getrude Ncube, Bernardo Hernandez Prado

Abstract

Background: Verification is an established cornerstone of Performance Based Financing (PBF) to ensure that reported outcomes, corresponding payments, and estimated public health benefits are accurate. While most programs perform community verifications to confirm that reported services were received by patients, many focus payment calculations and program-wide analysis on reviews of facility records (quantity verification), which measure internal record alignment but do not assess the validity of records. This is partly due to the high costs of community verifications per service verified given the effort required to find and interview patients. There is general agreement that risk-based sampling can help reduce costs by targeting verification efforts in areas with the highest risk of overreporting—but to do so, implementers need to know the characteristics of reported services with higher risk, a neglected topic.

Objective: To help guide decisions for risk-based sampling, this study explores factors associated with overreporting within a PBF setting, where overreporting means a patient did not plausibly confirm receipt of the reported Voluntary Medical Male Circumcision (VMMC). We also assess factors associated with when a patient was not interviewed.

Methods: Using community verification data from a VMMC program in Zimbabwe from 2016-2018, two generalized mixed effects models were employed each with a different outcome. The first outcome indicated when patients were not interviewed among patients selected for community verification. The second outcome indicated when patients did not meet the funders four requirements to confirm receipt of the service among patients that were interviewed. Bootstrapping-based expectation-maximization multiple imputation was used to address missing data and results were weighted using inverse probability of treatment to account for differences between the interviewed patients and all patients in the program.

Results: 2565 records were selected for community verification. Among them, 45% (1153/2565) of patients were not interviewed. Among interviewed patients, 19% (263/1412) were classified as overreported. After controlling for various facility and district-level variables, we found that patients in the target age range, which were renumerated at a higher price point, were less likely to be interviewed and over two times more likely to be classified as overreported (OR:2.92, 95% CI: 2.38-3.59). Patients reported from outside the fixed health facility were more likely to be interviewed and less likely to be classified as overreported. There was also evidence that interviews performed in person rather than on the phone were worth the additional investment in resources (OR: 1.61, 95% CI: 1.20-2.16).

Conclusion: Analyzing community verification data can inform future risk-based sampling to reduce costs and ultimately ensure these activities are more sustainable. Programs should continue investing in community verifications and in-depth analysis of the results given emerging evidence that quantity verifications can portray a misleading and overly positive assessment of the data quality of reported VMMCs.

Introduction

Performance- based financing (PBF) has generated substantial interest among governments and funders in the last 15 years (22). The theory that underpins PBF is that aligning provider payment with provided services increases health care worker productivity and service quality. Independent verification is an established cornerstone of PBF programs to ensure that reported outcomes, corresponding payments, and estimated public health benefits are accurate.

Typically, programs store aggregated reported outcome totals in a health management information system (HMIS), adding up the number of services reflected in records (often paper-based) at health facilities. To verify the reported totals, PBF programs typically include reviews of facility records (also called quantity verification) and interviews with patients (called community verification, counter verification, or client tracing) (60). Ultimately, these activities aim to identify and reduce overreporting, which can be either intentional (e.g. workers fabricating data to increase their pay) or unintentional (e.g. bookkeeping errors due to insufficient time or training). Because overreporting can be caused by issues in the aggregation process or by inaccuracies about the service itself, it’s important for verifications to be explicit about the source data used to compare the reported totals against. For quantity verifications, overreporting is estimated *based on facility records* (e.g., when the number of services in the HMIS exceed the number substantiated in records). For community verifications, overreporting is estimated *based on patient responses* (e.g., when a patient’s report about a service does not match the record). Measuring both types of overreporting is important to ensure stakeholders can trust the outcomes; the former helps improve recordkeeping and the latter ensures records represent real delivered services (e.g. the validity of records)—a critical component of PBF given the same incentives that are intended to promote productivity can also entice intentional overreporting.

Even though PBF programs acknowledge the importance of confirming services were received and invest significant resources in community verification (22,60,63,64), the literature suggests that results from these activities are commonly not analyzed program-wide or used to adjust payment. Among eight case studies that included community verifications (8,16,28,29,32–34,44), five explicitly mention the lack of data use (8,16,28,34,44). This is concerning because recent studies suggest that the results from quantity verification might not be a good proxy for whether reported results were actually received by patients. As previously reported in this same study, we found that while 94% of reported VMMCs were substantiated with facility records, only 55% of patients were interviewed, and among those interviewed, 17% did not plausibly confirm receipt of the reported service (Chapter 2). We also found no correlation between the facility-level results from quantity verification and community verification, suggesting both measures are needed in order for payments to be correct. Further, Turcotte-Tremblay et. al. performed a series of in-depth case studies for a Burkina Faso PBF program and found that in two (34,57) out of their three cases studies (34,43,57), there was widespread data fabrication by staff whose payment was conditioned on results in order to increase their pay.

Despite the known limitations of quantity verification, they are often relied on because community verification is more costly and time-intensive per service verified (8). Costs may be reduced by targeting areas empirically most prone to overreporting, also called risk-based sampling (8,12–15). To implement risk-based sampling, program planners needs to know which services to oversample but there are no studies that assess factors associated with overreporting based on community verification results (or other ways to measure the validity of records). The only two studies on the topic both use facility records as the source data (65) (Chapter 3).

This study contributes to closing this gap by assessing facility-level and patient-level factors associated with overreporting, where overreporting means the patients was unable to plausibly confirm receipt of the service, using data from a PBF program designed to scale VMMC in Zimbabwe. Findings from these analyses can help implementers better target community verifications, and ultimately make them more sustainable.

Methods

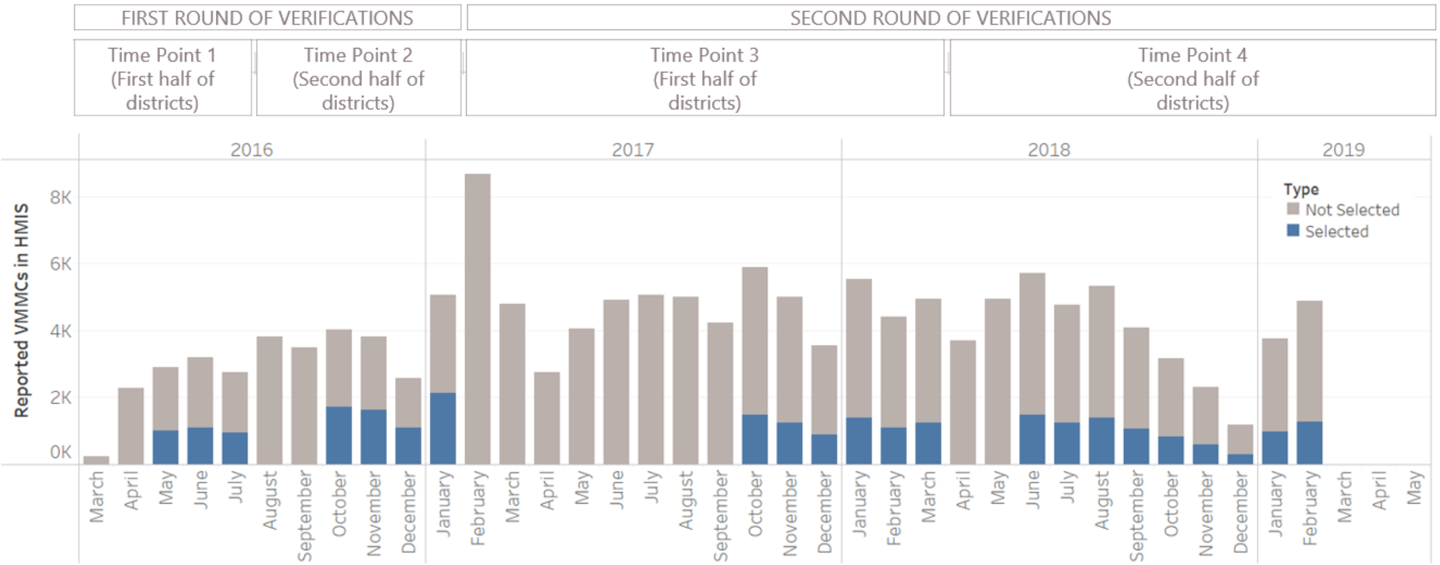
Study Setting and PBF Program

Program & Verification Design

The incidence of HIV/AIDs in Zimbabwe in 2009 was 669 new cases per 100,000 people – an estimate that has dramatically improved to 214 new infections per 100,000 people per year in 2019 (35). Part of this progress was due to the VMMC program, which the Ministry of Health and Child Care (MoHCC) launched in 2009. In 2015, the Bill & Melinda Gates Foundation began supporting Population Services International (PSI) to help scale the program using PBF. This study focuses on VMMCs reported between May 2016 and April 2019 as part of a grant. The program and verification have been described previously (Chapter 2, Chapter 3) but in brief, VMMCs were performed at VMMC locations, which were either fixed health facilities or locations staff traveled to on outreach. Mobilizers, categorized as gold or silver based on performance, sensitized communities and enroll interested patients across their assigned district. PSI and VMMC staff were paid on a per-unit basis for each VMMC performed. Patients between 15-29 years old were in the target age range, so PSI as an organization and mobilizers were renumerated at a higher price point for this group. The VMMC outcomes were aggregated into the HMIS each month, which is the data source used for verification as outlined below.

Verification teams performed quantity verifications and community verifications at four time points, with two rounds of verification for each district in the program (Figure 4.1). For the quantity verification, the alignment between the number of sufficiently complete facility records at VMMC locations and the number of reported VMMCs in the HMIS was assessed. These results have been explored previously (Chapter 2, Chapter 3). For community verification, patients were interviewed to assess the alignment between patients' reports of receiving the service and sufficiently complete facility records from the quantity verification. Patients selected for community verification were interviewed on the phone, if available, or traced to their home. To confirm receipt of the VMMC, patient responses (or in the case of minor's, their parent's responses) had to plausibly match four fields on the facility record: patients were asked if they were circumcised, and if they were, they were asked the general time period, district, and VMMC method (surgery of prepex device). This helped ensure that the patient's VMMC uniquely matched the selected record. Patient responses were compared after the interviews with data extracted from the facility record. If one or more of these four requirements did not align, the VMMC outcome was considered overreported.

Figure 4.1: Sample of facility records included in this analysis, from which patients were randomly drawn



PSI selected a purposeful sample of records and from the sample of records, they selected patients. The funder required that among reported VMMCs, 25% of records needed to be reviewed and 2.5% of patients needed to be interviewed. To select the sample records, the team divided the districts in half so that all districts were visited once per verification round; VMMC locations in the first half could be selected for the first and third verifications and VMMC locations in the second half could be selected in the second and fourth verifications. To select the portion of months to verify, PSI selected between 3 and 9 months of VMMCs prior to each time point in order to be on track to reach the 25% requirement. Within these months, VMMC locations that reported above 50 VMMCs and were from the relevant districts were selected for quantity verification. To select the sample of patient for community verification, 10% of sufficiently complete facility records were randomly selected.

Dataset for this Study

This study conducted secondary analysis using data from the program activities described above as well as data for the explanatory variables separately provided by the MoHCC. PSI provided a dataset at the patient-level; patients were nested within the VMMC locations where the procedures took place, the fixed health facility (whose staff performed the procedures), districts and provinces (Figure 4.2). For each patient, there were three sources of data. First, data that the verification team had extracted from facility records including the requirements to match with patient responses. Second, data from tracking sheets documenting whether the patient was found for the interview and if not, the reason. Third, the results from the community verification survey for the subset that were found and interviewed. For the explanatory variables, a list was created of the variables of interest based on existing literature, after which PSI and the MoHCC collected the variables that were available from district-level VMMC officials (Table 4.2). Patient-level characteristics were also included as explanatory variables. Our hypothesis was that a later verification round would be negatively associated with overreporting and that distance would be negatively associated with being interviewed, and positively associated with overreporting.

Figure 4.2: Data structure

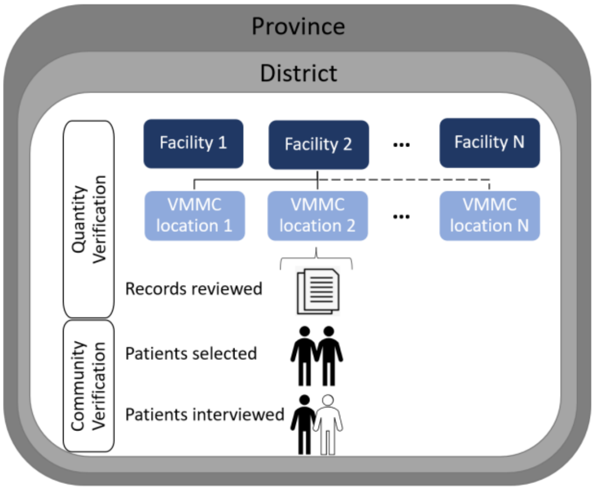


Table 4.1: Summary of explanatory variables

Level	Variables
District	Largest number of gold mobilizers that were active at one time during the grant period
	Largest number of silver mobilizers that were active at one time during the grant period
	Largest number of VMMC vehicles operating at one time during the grant period
	Whether it was the first or second round of verifications for the district
Health Facility	Number of VMMCs reported in the HMIS for the health facility at all VMMC locations throughout the grant period
	Largest number of trained VMMC staff (nurses & doctors) that were employed at one time during grant period
VMMC Location	Number of supervisory visits provided by the MOHCC
	Size of eligible population in catchment area
	Distance in kilometers from the VMMC location to the health facility based on GPS coordinates
	Type of facility where the VMMC took place (e.g., primary, secondary, tertiary, central care)
	Type of base funding where the VMMC took place (government, local council, mission, private)
Patient	Age of patient from the facility record
	VMMC method (prepex or surgical) from the facility record
	Whether the patient was interviewed on the phone or at home

Some VMMC locations in the program data were not selected for two verifications. To eliminate bias caused by the influence of the location being verified a second time, the dataset was restricted to only include the first verification from each location, though the dataset contains some repeated measures at the health facility level, which is controlled for using the verification round. Figure 4.2 summarizes the purposeful sample of facility records, from which patients were randomly sampled for this analysis.

Analysis

The purpose of this secondary data analysis was to examine the factors associated with overreporting, where overreporting means a service in facility records was not plausibly received by the patient. Two binary outcomes were assessed using two models. The first outcome indicated when patients were not interviewed (yes for not interviewed; no for interviewed) among patients selected for community verification (Model 1). The second outcome indicated when patients did not meet the funders four requirements to confirm receipt (yes for did not meet requirements; no for did meet) among patients that were interviewed (Model 2). For both outcomes, a generalized linear mixed effects model was used, including a random intercept for the health facility and the district to account for the nested structure of the data. Both models used the binomial family and logit link. The models were also stratified to assess whether findings differed between high/low performing facilities or between facilities with large/small catchment populations. For the former, facilities were classified as high performing if they reported more than the median number of VMMCs over the course of the grant period, and low otherwise. Facilities were similarly classified as having large catchment populations if they had more than the median population across sampled facilities, and low otherwise.

To account for patients that were selected but not interviewed (Model 2), inverse probability of treatment weighting (IPTW) was employed, using the explanatory variables that appeared to differentiate the interviewed from the not interviewed based on a t-test. Among the variables in Table 4.2, those that were not used included the number of VMMC staff and the number of supervisory visits. To reduce the variance of the effect estimate, the weights were stabilized and then trimmed at 1st and 99th percentiles. To address missingness in three variables (catchment population, number of supervisory visits, VMMC method) bootstrapping-based expectation-maximization multiple imputation was used. Five datasets were imputed separately for location level variables and patient level, each including all relevant independent variables as predictors. After combining the location-level and patient-level datasets, regressions were run separately on each of the imputed datasets and then the coefficients and standard error estimates were pooled using Rubin's rules (Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons; 2004). To test the sensitivity of the findings to missing data, all models were also estimated using only observations that had no missing data (referred to as complete case analysis). More details about the weighting and imputation can be found in Annex I and Annex II.

All analyses were performed in STATA version 17.0 (StataCorp. 2023. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC.) or R version 2022.12.0 (R Foundation for Statistical Computing, Vienna, Austria). Multiple imputation was performed using the Amelia II package version 1.8.1(68).

Ethical Approval

The protocol was approved by the research council of Zimbabwe (approval # MRCZ/E/237) and the MoHCC at all appropriate levels. The University of Washington's human subject's department classified the research as exempt. Patients who participated in interviews provided oral consent.

Results

2565 records were selected for community verification and included in the sample. Among them, 45% (1153/2565) of patients were not interviewed. Among interviewed patients, 19% (263/1412) were classified as overreporting. The exclusion of individuals from repeated verification rounds is the reason why this sample is smaller than in our previous analysis (Chapter 2).

Characteristics of patients selected for community verification

Table 4.2 describes the sample of patients. Among all selected patients, the majority were circumcised surgically (89%), and in the target age range 15-29 years old (58%); 62% were selected as part of the second round of verifications. Patients were from health facilities that on average reported 5658 VMMCs and had 5 clinical staff. 52% of patients were circumcised outside of the fixed health facility but within 50 kilometers. The average catchment population of the VMMC location was 12,425. Comparing selected patients with those that were interviewed, the largest differences included that interviewed patients were from facilities with lower reported VMMCs (5,187), smaller catchment populations (11,080), and had a greater proportion of patients below 15 years old (47%). Annex III describes the sample of patients that were included in the complete case analysis.

Factors associated with a patient that was not interviewed (Model 1)

Table 4.3 presents the findings for factors associated with when a patient was not interviewed after controlling for the other variables in the table. Relative to VMMCs performed at a fixed health facility, patients whose procedures took place any distance away were more likely to be interviewed; considering the stratified analysis, this appears to have been driven by facilities with large catchment populations (Annex IV). Older patients and patients who had a surgical procedure (OR: 2.21, 95% CI: 1.57-3.12) were both more likely to not be interviewed. In particular, patients that were in the target age range of 15-29 years had the highest ORs, resulting in nearly three times the likelihood of not being interviewed (OR:2.92, 95% CI: 2.38-3.59); this group was positively associated with overreporting in each of the stratified models, suggesting the size of the catchment or number of reported did not influence this outcome. The complete case analysis additionally showed that locations with larger catchment populations were negatively associated with not being interviewed (OR:0.81, 95% CI: 0.66-0.99) (i.e. are more likely to be interviewed in a large catchment), and patients from the second round of verifications were positively associated (OR:1.88, 95% CI: 1.27-2.80).

Factors associated with overreporting (Model 2)

Table 4.4 presents findings for factors associated with overreporting. Relative to VMMCs performed at a fixed health facility, procedures that took place away were less likely to be classified as overreporting as were procedures that took place as part of the second round of verifications (OR: 0.39, 95% CI: 0.25-0.61). Overreporting was more likely for fixed health facilities that reported greater numbers of VMMCs (OR: 2.00, 95% CI: 1.16-3.44) and VMMC locations that were rural hospitals (OR: 3.28, 95% CI: 1.28-8.40). In terms of patient level characteristics, overreporting was more likely among patients in the target age range (OR: 2.21, 95% CI: 1.49-3.26) and when the patient was interviewed in person (OR: 1.61, 95% CI: 1.20-2.16). Considering the stratified analyses, these results appeared to be largely driven by facilities that reported more VMMCs given each significant association held when the regression was run among patients from these facilities (Annex 5). The stratified analyses also showed that patients from the target age range were not associated with overreporting among facilities with small catchment areas, but the positive association held in facilities with large populations as well as facilities with both small/large reported VMMCs. And finally, the complete case analysis additionally showed that the number of gold mobilizers and catchment population were negatively associated with overreporting and surgical procedures were positively associated; location distance and rural hospitals were also not significant in the model with only complete data.

Table 4.2. Description of patients

	Patients selected for community verification					Patients interviewed for community verification				
	Mean or Proportion	SD	Min	Max	N	Mean or Proportion	SD	Min	Max	N
Number of reported VMMCs (Log)	5658.2	3650.9	22	14829	2565	5187.2	3278.5	22	14829	1412
Number of staff	5.436	2.088	2	10	2565	5.475	2.044	2	10	1412
Number of gold mobilizers	4.239	0.991	1	6	2565	4.277	0.967	1	6	1412
Number of silver mobilizers	4.050	4.154	0	15	2565	3.772	3.882	0	15	1412
Number of vehicles	1.795	0.727	1	3	2565	1.697	0.738	1	3	1412
Population of catchment	12425.2	20568.1	590	156842	2523	11080.3	21206.7	50	156842	1389
Location distance from facility										
[0]	0.412	0.492	0	1	2565	0.395	0.489	0	1	1412
[>0 to 50]	0.526	0.499	0	1	2565	0.539	0.499	0	1	1412
[>50 to 150]	0.062	0.240	0	1	2565	0.066	0.248	0	1	1412
Number of supervisory visits	2.532	2.048	0	10	2268	2.524	2.135	0	10	1350
Type of VMMC location										
District	0.194	0.396	0	1	2565	0.233	0.423	0	1	1412
Mission	0.142	0.349	0	1	2565	0.161	0.367	0	1	1412
Rural clinic	0.562	0.496	0	1	2565	0.541	0.498	0	1	1412
Rural hospital	0.102	0.303	0	1	2565	0.065	0.247	0	1	1412
Type of base funding										
Council	0.313	0.464	0	1	2565	0.288	0.453	0	1	1412
Government	0.465	0.499	0	1	2565	0.482	0.500	0	1	1412
Mission	0.207	0.405	0	1	2565	0.210	0.407	0	1	1412
Private	0.014	0.118	0	1	2565	0.021	0.144	0	1	1412
Verification round for district										
First	0.388	0.487	0	1	2565	0.409	0.492	0	1	1412
Second	0.612	0.487	0	1	2565	0.591	0.492	0	1	1412
Age of Patient										
Less than 15	0.374	0.484	0	1	2565	0.470	0.499	0	1	1412
15-29 years	0.583	0.493	0	1	2565	0.489	0.500	0	1	1412
30 or more	0.043	0.202	0	1	2565	0.041	0.199	0	1	1412
Method										
Prepex	0.111	0.314	0	1	2512	0.130	0.337	0	1	1396
Surgical	0.889	0.314	0	1	2512	0.870	0.337	0	1	1396
Interview Location										
On phone						0.461	0.499	0	1	1412
In person						0.539	0.499	0	1	1412

* Every category is shown as a dummy where the variable equals 1 and 0 otherwise.

Table 4.3: Logit regression results for not being interviewed (Model 1)

	Results with Imputed Values (N= 2,565)					Complete Case Analysis (N= 2,214)				
	Adj. Odds ratio*	Std. err.	t	[95% conf. interval]		Adj. Odds ratio*	Std. err.	t	[95% conf. interval]	
Number of reported VMMCs (Log)	1.38	0.23	1.89	0.99	1.92	1.34	0.24	1.62	0.94	1.90
Number of staff	1.13	0.09	1.55	0.97	1.33	1.10	0.07	1.37	0.96	1.25
Number of gold mobilizers	0.83	0.15	-1.00	0.58	1.19	0.88	0.12	-0.99	0.68	1.14
Number of silver mobilizers	1.00	0.05	-0.04	0.90	1.10	1.05	0.04	1.37	0.98	1.13
Number of vehicles	1.56	0.49	1.42	0.84	2.88	1.31	0.31	1.16	0.83	2.08
Population of catchment (Log)	0.98	0.09	-0.26	0.82	1.16	0.81	0.08	-2.07	0.66	0.99
Location distance (km)										
[0]	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
[>0 to 50]	0.62	0.15	-2.03	0.39	0.98	0.83	0.22	-0.70	0.50	1.39
[>50 to 140]	0.35	0.11	-3.39	0.19	0.64	0.48	0.16	-2.18	0.25	0.93
Number of supervisory visits	0.90	0.05	-1.76	0.80	1.02	0.99	0.05	-0.18	0.90	1.09
Type of facility										
District	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Mission	0.78	0.25	-0.76	0.42	1.46	0.90	0.29	-0.31	0.48	1.71
Rural clinic	1.05	0.29	0.20	0.62	1.80	0.68	0.19	-1.41	0.40	1.16
Rural hospital	1.45	0.42	1.29	0.82	2.56	0.71	0.23	-1.05	0.37	1.35
Type of base funding										
Council	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)

Government	1.10	0.16	0.63	0.82	1.46	1.31	0.20	1.73	0.96	1.77
Mission	1.09	0.28	0.34	0.66	1.81	0.95	0.25	-0.19	0.57	1.60
Private	0.46	0.25	-1.44	0.16	1.33	0.70	0.37	-0.67	0.25	1.97
Verification round										
First	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Second	1.14	0.20	0.74	0.81	1.61	1.88	0.38	3.13	1.27	2.80
Age of Patient										
Less than 15	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
15-29 years	2.92	0.31	10.25	2.38	3.59	2.99	0.33	9.87	2.41	3.72
30 or more	2.09	0.51	3.04	1.30	3.37	2.31	0.65	2.97	1.33	4.01
Method										
Prepex	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Surgical	2.21	0.39	4.56	1.57	3.12	2.29	0.44	4.32	1.57	3.34

Note: All odds ratios are adjusted by all the variables in the table.

Table 4.4: Logit regression results for overreporting defined by when a patient did not plausibly confirm receipt (Model 2)

	Results with Imputed Values (N=1,412)						Complete Case Analysis (N=1,334)					
	Adj. Odds ratio*	Std. err.	t	P>|t|	[95% conf. interval]		Adj. Odds ratio*	Std. err.	t	P>|t|	[95% conf. interval]	
Number of reported VMMCs (Log)	2.00	0.55	2.51	0.012	1.16	3.44	1.59	0.44	1.68	0.094	0.92	2.72
Number of staff	1.02	0.07	0.30	0.761	0.90	1.16	1.04	0.07	0.58	0.560	0.91	1.19
Number of gold mobilizers	0.83	0.08	-1.88	0.061	0.68	1.01	0.84	0.07	-2.14	0.032	0.72	0.99
Number of silver mobilizers	0.99	0.04	-0.26	0.791	0.91	1.07	1.02	0.03	0.53	0.595	0.96	1.08
Number of vehicles	0.81	0.22	-0.77	0.440	0.47	1.39	0.81	0.14	-1.19	0.235	0.57	1.15
Population of catchment (Log)	0.91	0.07	-1.22	0.221	0.77	1.06	0.87	0.06	-2.10	0.036	0.77	0.99
Location distance (km)												
[0]	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
[>0 to 50]	0.41	0.11	-3.28	0.001	0.24	0.70	0.80	0.28	-0.64	0.525	0.40	1.59
[>50 to 140]	0.36	0.15	-2.42	0.016	0.16	0.82	0.74	0.23	-0.95	0.340	0.40	1.38
Number of supervisory visits	0.91	0.06	-1.47	0.141	0.79	1.03	0.96	0.05	-0.81	0.416	0.87	1.06
Type of VMMC location												
District	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Mission	1.28	0.62	0.51	0.612	0.49	3.32	0.84	0.46	-0.33	0.742	0.29	2.44
Rural clinic	1.63	0.57	1.40	0.162	0.82	3.24	0.93	0.38	-0.17	0.869	0.42	2.08
Rural hospital	3.28	1.57	2.48	0.013	1.28	8.40	1.83	0.85	1.30	0.195	0.73	4.54
Type of base funding												
Council	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Government	0.91	0.21	-0.43	0.667	0.58	1.42	1.11	0.27	0.41	0.680	0.69	1.78
Mission	0.62	0.16	-1.85	0.064	0.38	1.03	0.78	0.17	-1.13	0.257	0.51	1.20
Private	0.27	0.21	-1.66	0.097	0.06	1.27	0.37	0.26	-1.39	0.163	0.09	1.50
Verification round												
First	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Second	0.39	0.09	-4.17	0.000	0.25	0.61	0.34	0.11	-3.25	0.001	0.18	0.65
Age of Patient												
Less than 15	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
15-29 years	2.21	0.44	3.96	0.000	1.49	3.26	2.21	0.47	3.74	0.000	1.46	3.34
30 or more	1.15	0.51	0.31	0.755	0.48	2.74	1.34	0.65	0.61	0.541	0.52	3.45
Method												
Prepex	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
Surgical	1.81	0.58	1.84	0.066	0.96	3.39	2.21	0.66	2.67	0.008	1.23	3.96
Interview Location												
On phone	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)	(base)
In person	1.61	0.24	3.17	0.002	1.20	2.16	1.62	0.29	2.66	0.008	1.13	2.30

Note: All odds ratios are adjusted by all the variables in the table.

Discussion

To our knowledge, this is the first study to quantitatively estimate the factors associated with being interviewed and overreporting of services verified by patient responses in a PBF setting. While both outcomes indicate that a facility record was not confirmed, the latter is of most interest to PBF funders since the errors leading to discrepant interviews are more often in the facility's control (Chapter 2).

We found that patients in the target age range, which were renumerated at a higher price point, were more likely to both not be interviewed and be classified as overreported. Two potential reasons for this were articulated by the verifiers who performed the fieldwork as presented in a previous study for this same project (Chapter 2). It was suspected by some verifiers that staff could have fabricated information on the facility records to increase their pay. Verifiers explained that a common cause of an unconfirmed VMMC by a patient was when a patient had signed up with a mobilizer but did not go through with the procedure; since mobilizers were paid more for this age group, there could have been opportunistic completion of records following the initial signup. Another reason for this outcome could be related to patients themselves providing false/misleading contact information to avoid follow-up, which many verifiers had heard happened directly from patients. If older patients (15-29 years as opposed to younger boys) were more likely to provide inaccurate information, this could have contributed to the negative association we found between age and being interviewed, but would not explain why the target age group was more likely to be overreported based on interviews.

It was surprising that patients that were circumcised away from the fixed health facility were more likely to be interviewed. We expected the opposite, given that to find such patients requires more resources (8), and some verifiers reported that they had insufficient time and resources to exhaustively look for all patients (Chapter 2). The fixed health facilities were often in district centers, so one possible explanation is that outside villages were smaller, making it easier for villagers to assist with locating patients.

The second round of verifications at the district level were associated with a decreased likelihood of overreporting, which is in line with various other studies that have shown that audits improve data quality (71,72). While the VMMC locations and patients had no repeated measures in our dataset, the VMMC staff from health facilities could have either been involved in an earlier verification and/or knew the verifications had occurred and as a result, improved record keeping and/or reduced any data fabrication. Finally, interviews performed in person were more likely to identify overreporting, showing tracing patients to their homes when they were not reachable by phone was worth the additional resources. This could be due to the verifier's improved ability to gain rapport and probe in person (66).

This study is part of an effort to better target and reduce the cost of PBF verifications. Reducing the cost of community verifications is of particular importance since an eyewitness account from a patient is a more valid data source than documentation created by staff who benefit when more services are reported. Without cost effective solutions, community verifications could be deprioritized, as shown by the suspension of these efforts in Benin (8). There is also emerging evidence that quantity verifications can portray a misleading and overly positive assessment of the data quality of reported VMMCs (34,43,57). Community verifications shed light on the extent to which reported VMMCs and the resulting health impact were actually realized. To reduce their cost, we identified factors associated with unconfirmed services, which programs could consider during risk-based sampling decisions.

This study has several limitations. Patients were a random sample from a purposeful sample of VMMC locations and months, so may not be representative of all reported VMMCs in the grant period. In particular, facilities that reported fewer VMMCs were not included in the sample, though we were still able to detect an association between the number of reported VMMCs and overreporting. Although our analysis was able to control for various facility and geographical characteristics, other unmeasured variables may confound the relationship between these factors and the two outcomes. In particular, we were unable to look at whether facilities with higher workloads influenced either outcome given the staff in this program would at times work in different catchment areas to alleviate high workloads. Finally, VMMC is a sensitive, stigmatized topic so findings might not translate to other programs if patients are more open to follow-up and disclosing personal health data during an interview.

Conclusion

Analyzing community verification data can inform future risk-based sampling as a step toward reducing the cost and increasing the sustainability of these efforts – efforts which crucially measure the validity of facility records. Further studies are needed to explore additional factors and to see whether our findings hold in different VMMC and PBF programs.

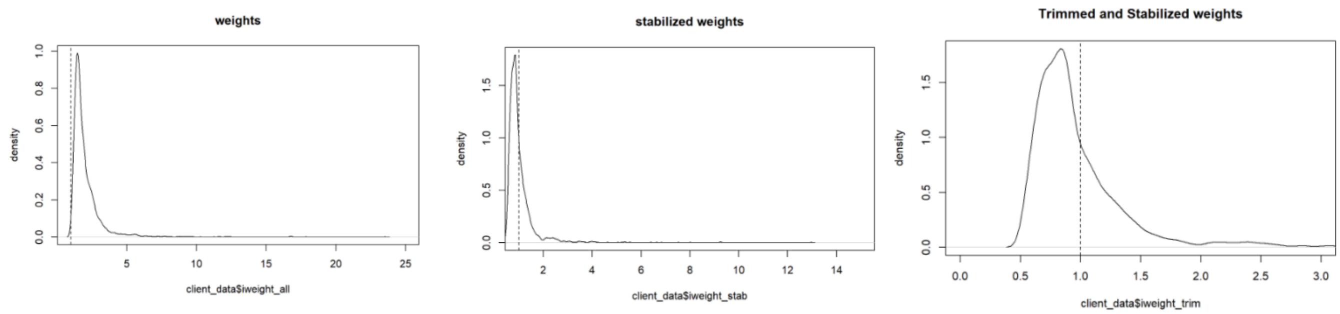
Annex

Annex I: Weighting summary

In this study, a random sample of patients from the facility records were selected for community verification, but not all patients were interviewed. Weighting the results therefore accounts for differences between the interviewed patients and all patients that were selected. Through six steps, inverse probability of treatment weighting (IPTW) was used. First, to determine which factors differentiated patients that were interviewed from patients that were not interviewed, t-tests comparing the two groups were conducted on all independent variables from Model 1. All variables were significantly different between the groups except for two: the number of trained clinical VMMC staff and the number of supervisory visits. Second, a logistic regression was run using the outcome of whether the patient was interviewed or not (1/0) and all independent variables that differentiated the two groups based on the first step. Third, predictions for the probability of being interviewed (prob) for each patient were then obtained based on the regression results. For the fourth step to increase the weight of patients that had low probabilities of selection, raw weights were calculated based on the inverse of the probability of being interviewed (1/prob). Fifth, the raw weights were stabilized by multiplying each weight by the proportion of patients that were interviewed among all selected patients. And finally, the weights were trimmed at 1th and 99^{th} percentiles to eliminate extreme outliers. Figure 3 shows the distribution of the weights at each step.

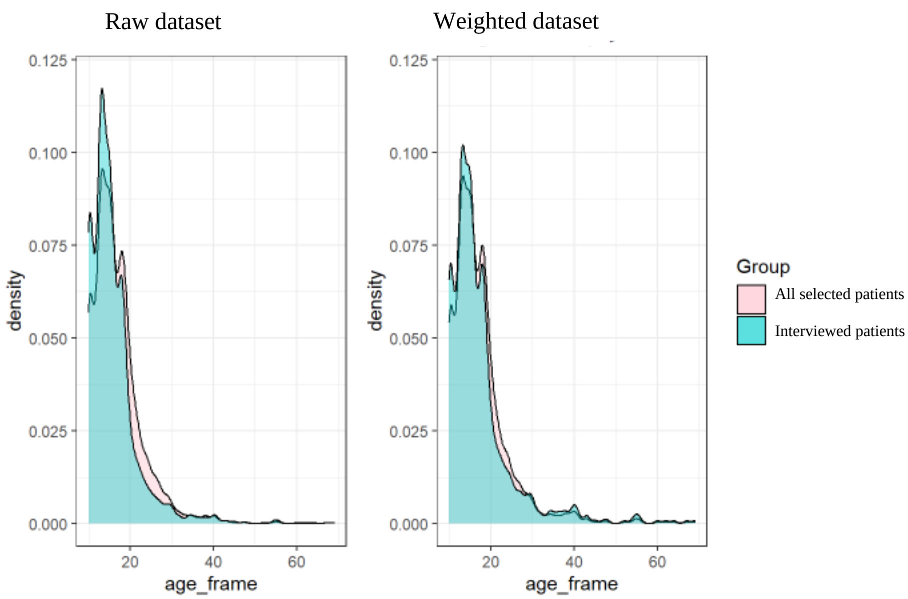
Figure 4.3: Distribution of weights

Distribution of raw weights	Distribution of stabilized weights	Distribution stabilized and trimmed weights



To check how well the weights corrected for bias, t-tests were again run but instead including the weights; about half of the independent variables were no longer significant between the patients that were interviewed and those that were not. To investigate further, the difference in means between the two groups were compared in the original data and the weighted data; comparing the two differences showed that the difference reduced for all variables. The results of these checks suggest the weights made the interviewed group more similar on average to the non-interviewed group. To further assess whether the interviewed group was adjusted appropriately relative to the entire unweighted sample, each variable was visualized such as with age in Figure 4.

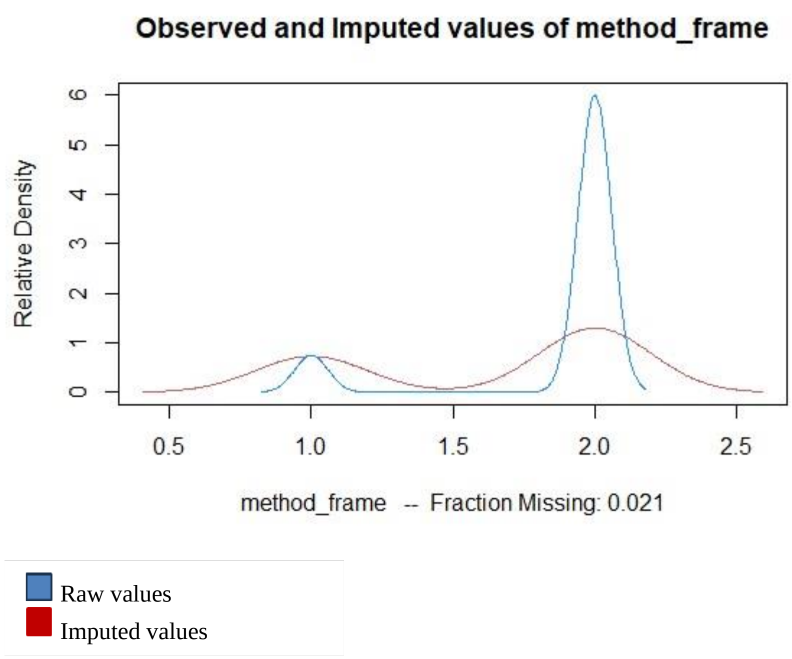
Figure 4.4: Distribution of patients age, comparing raw and weighted data



Annex II: Imputation summary

Amelia II in R software version 2022.12.0 (R Foundation for Statistical Computing, Vienna, Austria) was used for imputation. Amelia II uses a bootstrap-based algorithm to create *m* complete datasets from m imputations. Amelia works by leveraging all distributions in the dataset so as to not change any relationships. The patient-level dataset in this study had missing data for two facility-level variables (number of supervisory visits (n = 297) and the size of eligible population in catchment area (n = 42)) and one patient-level variable (VMMC method (n = 53)). First to impute the VMMC method, 5 iterations were employed (*m* = 5) on the patient-level dataset using the same variables as are in Model 1. Next for the two facility-level variables with missingness, the imputed values for each of the 5 imputed datasets from Aim 2 were merged in. Finally, the coefficients were combined in Stata using the mi command, which relies on Rubin's rules. Figure 5 presents the distribution of the raw data compared to the imputed values for the method.

Figure 4.5: Distribution of VMMC method, comparing raw to imputed values



Annex III: Description of patients in the complete case analysis

	Results with Imputed Values (N=2565)				Complete Case Analysis (N= 2,214)			
	Mean or Proportion	SD	Min	Max	Mean or Proportion	SD	Min	Max
Number of reported VMMCs	5658.2	3650.9	22.00	14829.0	5793.2	3776.4	22.00	14829.0
Number of staff	5.436	2.088	2	10	5.510	2.199	2	10
Number of gold mobilizers	4.239	0.991	1	6	4.290	1.038	1	6
Number of silver mobilizers	4.050	4.154	0	15	4.356	4.331	0	15
Number of vehicles	1.795	0.727	1	3	1.766	0.766	1	3
Population of catchment (Log)	8.782	1.017	6.380	11.96	8.589	0.835	6.380	11.96
Number of supervisory visits	2.532	2.048	0	10	2.492	2.024	0	10
<u>Location distance (kilometers)</u>								
[0]	0.412	0.492	0	1	0.343	0.475	0	1
[>0 to 50]	0.526	0.499	0	1	0.586	0.493	0	1
[>50 to 140]	0.0616	0.240	0	1	0.0714	0.257	0	1
<u>Type of facility</u>								
District	0.194	0.396	0	1	0.194	0.396	0	1
Mission	0.142	0.349	0	1	0.160	0.367	0	1
Rural clinic	0.562	0.496	0	1	0.579	0.494	0	1
Rural hospital	0.102	0.303	0	1	0.0664	0.249	0	1
<u>Type of base funding</u>								
Council	0.313	0.464	0	1	0.276	0.447	0	1
Government	0.465	0.499	0	1	0.472	0.499	0	1
Mission	0.207	0.405	0	1	0.235	0.424	0	1
Private	0.0140	0.118	0	1	0.0163	0.127	0	1
<u>Method</u>								
Prepex	0.111	0.314	0	1	0.110	0.313	0	1
Surgical	0.889	0.314	0	1	0.890	0.313	0	1
<u>Verification round</u>								
First	0.388	0.487	0	1	0.366	0.482	0	1
Second	0.612	0.487	0	1	0.634	0.482	0	1
<u>Age of Patient</u>								
Less than 15	0.374	0.484	0	1	0.384	0.487	0	1
15-29 years	0.583	0.493	0	1	0.582	0.493	0	1
30 or more	0.0421	0.201	0	1	0.0334	0.180	0	1

Annex IV: Logit regression results for not being interviewed, stratified by catchment population and number of reported VMMCs (Model 1)

Variables	Full Model (N=2565)	Stratified Analysis			
	Adjusted Odds Ratio*	Facilities with large catchment populations (N=1515)	Facilities with small catchment populations (N=1040)	Facilities that reported large numbers of VMMCs (N=1955)	Facilities that reported small numbers of VMMCs (N=610)
Number of reported VMMCs (Log)	1.38	1.60*	2.71	2.51*	1.70
Number of staff	1.13	0.85	1.27*	1.10	0.91
Number of gold mobilizers	0.83	1.00	0.88	0.85	1.19
Number of silver mobilizers	1.00	1.03	1.05	1.01	0.85
Number of vehicles	1.57	1.10	1.49	1.37	1.45
Population of catchment (Log)	0.99	0.65	0.50**	1.04	0.88
<u>Location distance (kilometers)</u>					
[0]	(base)	(base)	(base)	(base)	(base)
[>0 to 50]	0.62*	0.44*	0.95	0.59	0.65
[>50 to 140]	0.35***	0.13***	1.36	0.30**	0.28*
Number of supervisory visits	0.90	0.73**	1.06	0.96	0.93
<u>Type of facility</u>					
District	(base)	(base)	(base)	(base)	(base)
Mission	0.80	1.23	1.69	1.31	0.26
Rural clinic	1.07	0.99	0.25*	1.18	0.78
Rural hospital	1.46	5.93***	0.24	0.68	10.00***
<u>Type of base funding</u>					
Council	(base)	(base)	(base)	(base)	(base)
Government	1.10	1.64*	1.38	1.30	0.70
Mission	1.09	1.01	0.37	0.89	(omitted)
Private	0.46	3.46	0.35	0.21	0.42
<u>Verification round</u>					
First	(base)	(base)	(base)	(base)	(base)
Second	1.13	1.71*	2.93*	1.76*	1.11
<u>Age of Patient</u>					
Less than 15	(base)	(base)	(base)	(base)	(base)
15-29 years	2.92***	2.60***	3.31***	3.23***	2.05**
30 or more	2.10**	1.91*	2.29	2.48**	1.51
<u>Method</u>					
Prepex	(base)	(base)	(base)	(base)	(base)
Surgical	2.21***	1.95**	3.78**	2.57***	1.21
Constant	0.01*	0.84	0.00	0.00***	0.02
var(_cons[district])	1.59*	2.17	1.34	1.00	1.29
var(_cons[dis~t>par~e])	1.11	1.03	1.22	1.18	1.00
N	2565	1515	1040	1955	610

Note: All odds ratios are adjusted by all the variables in the table.

Annex V: Logit regression results for overreporting defined by when a patient did not plausibly confirm receipt, stratified by catchment population and number of reported VMMCs (Model 2)

Variable	Full Model	Stratified Analyses			
	Adjusted Odds Ratio* (N = 1412)	Facilities with large catchment populations (N = 763)	Facilities with small catchment populations (N = 634)	Facilities that reported large numbers of VMMCs (N = 1066)	Facilities that reported small numbers of VMMCs (N = 328)
Number of reported VMMCs (Log)	2.01*	1.34	2.66*	3.51*	1.92
Number of staff	1.02	1.21	1.12	0.97	0.95
Number of gold mobilizers	0.83	0.61	0.82**	0.81	1.19
Number of silver mobilizers	0.99	1.00	0.99	1.02	1.00
Number of vehicles	0.80	0.96	0.92	0.69	0.74
Population of catchment (Log)	0.92	1.36	1.25	0.82	1.06
<u>Location distance (kilometers)</u>					
[0]	(base)	(base)	(base)	(base)	(base)
[>0 to 50]	0.41**	0.34**	0.36*	0.16**	0.39*
[>50 to 140]	0.36*	0.37	0.56	0.09**	0.49
Number of supervisory visits	0.92	1.06	0.99	0.87	1.03
<u>Type of facility</u>					
District	(base)	(base)	(base)	(base)	(base)
Mission	1.30	1.63	0.67	1.66	0.12**
Rural clinic	1.68	3.17*	0.60	3.73	1.57
Rural hospital	3.34*	2.23	1.38	6.68**	4.57
<u>Type of base funding</u>					
Council	(base)	(base)	(base)	(base)	(base)
Government	0.90	0.47*	2.65	1.30	0.36
Mission	0.61	0.66	0.48	0.77	(omitted)
Private	0.27	(omitted)	1.35	0.68	(omitted)
<u>Verification round</u>					
First	(base)	(base)	(base)	(base)	(base)
Second	0.40***	0.25***	0.76	0.38*	0.43
<u>Age of Patient</u>					
Less than 15	(base)	(base)	(base)	(base)	(base)
15-29 years	2.21***	2.51***	1.62	1.98**	3.20***
30 or more	1.16	0.80	1.52	0.99	1.89
<u>Method</u>					
Prepex	(base)	(base)	(base)	(base)	(base)
Surgical	1.81	1.94	1.14	1.58	2.25
<u>Interview Location</u>					
On phone	(base)	(base)	(base)	(base)	(base)
In person	1.61**	1.47	2.07*	1.54*	1.89
Constant	0.00*	0.00*	0.00**	0.00	0.00
var(_cons[district])	1.48*	1.00	1.00	1.50	1.00
var(_cons[dis~t>par~e])	1.00	1.20	1.00	1.05	1.00

Note: All odds ratios are adjusted by all the variables in the table.

Chapter 5. Conclusion

The results of this dissertation highlight the importance of analyzing PBF verification data and performing research to understand the paths and processes that led to the results—all to inform appropriate action. While exploratory, Aim 2 (Chapter 3) presented methods that future programs can use to understand the factors associated with overreporting within quantity verifications to ultimately reduce the cost and better target these efforts. We also recommend qualitative research to understand the role of higher staff workloads on overreporting, as opposed to catchment and staff size separately.

Our findings from Aims 2 and 3 (Chapters 2 and 4) are part of emerging evidence that suggest quantity verifications can portray a misleading and overly positive assessment of reported results tied payment. Because tying payment to facility records alone risks overpaying for services and misreporting performance, programs should continue investing in community verifications. To increase the use of community verification findings, PBF programs should consider using and improving our proposed results to action framework. In particular, reducing payment in cases when patient responses do not match facility records helps ensure programs are paying for achieved results and that there is a real threat of sanctioning for overreporting—two core theoretical requirements for PBF.

REFERENCES

1. Musgrove P. Rewards for Good Performance or Results: A Short Glossary [Internet]. 2011 [cited 2017 Mar 29]. Available from: https://www.rbfhealth.org/resource/financial-and-other-rewards-good-performance-or-results-guided-tour-concepts-and-terms-and
2. Health Results Innovation Trust Fund. Mission | RBF Health [Internet]. [cited 2016 Oct 16]. Available from: https://www.rbfhealth.org/mission
3. McIsaac M, Kutzin J, Dale E, Soucat A. Results-based financing in health: from evidence to implementation. Bull World Health Organ. 2018 Nov 1;96(11):730-730A.
4. Soucat A, Dale E, Mathauer I, Kutzin J. Pay-for-Performance Debate: Not Seeing the Forest for the Trees. Health Syst Reform. 2017 Apr 3;3(2):74–9.
5. Witter S, Toonen J, Meessen B, Kagubare J, Fritsche G, Vaughan K. Performance-based financing as a health system reform: mapping the key dimensions for monitoring and evaluation. BMC Health Services Research. 2013;13:367.
6. Meessen B, Soucat A, Sekabaraga C. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? Bull World Health Organ. 2011 Feb 1;89(2):153–6.
7. Nimpagaritse M, Korachais C, Meessen B. Effects in spite of tough constraints - A theory of change based investigation of contextual and implementation factors affecting the results of a performance based financing scheme extended to malnutrition in Burundi. PLoS One. 2020;15(1):e0226376.
8. Antony M, Bertone MP, Barthes O. Exploring implementation practices in results-based financing: the case of the verification in Benin. BMC Health Services Research. 2017 Mar 14;17(1):204.
9. Ssengooba F, McPake B, Palmer N. Why performance-based contracting failed in Uganda--an “open-box” evaluation of a complex health system intervention. Soc Sci Med. 2012 Jul;75(2):377–83.
10. Ogundeji Y, Jackson C, Sheldon T, Olubajo O, Ihebuzor N. Pay for performance in Nigeria: The influence of context and implementation on results. Health Policy and Planning. 2016 Apr 1;31:czw016.
11. Performance-based financing in the context of the complex remuneration of health workers: findings from a mixed-method study in rural Sierra Leone | BMC Health Services Research | Full Text [Internet]. [cited 2021 Feb 9]. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-016-1546-8
12. Towards constructive rethinking of PBF: perspectives of implementers in sub-Saharan Africa | BMJ Global Health [Internet]. [cited 2021 Jan 11]. Available from: https://gh.bmj.com/content/3/5/e001036

13. Vergeer P, Heard AC, Josephson E, Fleisher L. Verification in results-based financing for health [Internet]. The World Bank; 2016 Dec [cited 2019 Mar 16] p. 1–22. Report No.: 112342. Available from: http://documents.worldbank.org/curated/en/504661485897909610/Verification-in-results-based-financing-for-health

14. Joseph Naimoli, Vergeer P. Verification at a Glance: A Series of Snapshots of Experiences in Verifying Performance Linked to Financial Incentives for Results-Based Financing (RBF) Programs from Selected Countries [Internet]. 2010. Available from: https://www.rbfhealth.org/sites/rbf/files/RBFVerification_at_a_glance.pdf

15. 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 Jan 1;3(1):e000664.

16. Verification of Performance in Results-Based Financing: The Case of Burundi | RBF Health [Internet]. [cited 2016 Oct 27]. Available from: https://www.rbfhealth.org/publication/verification-performance-results-based-financing-case-burundi

17. Salehi AS, Borghi J, Blanchet K, Vassall A. The cost-effectiveness of using performance-based financing to deliver the basic package of health services in Afghanistan. BMJ Global Health. 2020 Sep 1;5(9):e002381.

18. Cordaid International [Internet]. 2018 [cited 2021 Jan 7]. Blog: Towards a balanced rethinking of Performance-Based Financing. Available from: https://www.cordaid.org/en/news/towards-balanced-rethinking-performance-based-financing/

19. BMJ. BMJ Global Health blog. 2018 [cited 2021 Feb 7]. Towards a more balanced rethinking of performance-based financing. Available from: https://blogs.bmj.com/bmjgh/2018/02/09/towards-a-more-balanced-rethinking-of-performance-based-financing/

20. IHME. Institute for Health Metrics and Evaluation. 2019 [cited 2023 Jun 10]. Zimbabwe Snapshot. Available from: https://www.healthdata.org/zimbabwe

21. Zimbabwe [Internet]. [cited 2021 Apr 24]. Available from: https://www.unaids.org/en/regionscountries/countries/zimbabwe

22. GPOBA, Instiglio, World Bank Group. A Guide for Effective Results Based Financing Strategies [Internet]. 2018 [cited 2021 Oct 22]. Available from: https://documents1.worldbank.org/curated/en/265691542095967793/pdf/A-Guide-For-Effective-Results-Based-Financing-Strategies.pdf

23. Eijkenaar F. Key issues in the design of pay for performance programs. Eur J Health Econ. 2013 Feb 1;14(1):117–31.

24. Miller G, Singer Babiarz K. Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs [Internet]. National Bureau of Economic Research; 2013 [cited 2024 Feb 3]. (Working Paper Series). Available from: https://www.nber.org/papers/w18932

25. Fritsche GB, Soeters R, Meessen B. Performance-Based Financing Toolkit [Internet]. Washington, DC: World Bank; 2014 [cited 2023 Jun 6]. Available from: http://hdl.handle.net/10986/17194

26. The Road to Results [Internet]. [cited 2021 Oct 26]. Available from: https://elibrary.worldbank.org/doi/abs/10.1596/978-0-8213-7891-5

27. Verification in Results-Based Financing: The Case of the United Kingdom | RBF Health [Internet]. [cited 2021 Feb 1]. Available from: https://www.rbfhealth.org/resource/verification-results-based-financing-case-united-kingdom

28. Verification of Performance in Results-Based Financing : The Case of Community and Demand-Side RBF in Rwanda [Internet]. World Bank; Available from: http://hdl.handle.net/10986/20791

29. Verification of Performance in Results-Based Financing : The Case of Afghanistan [Internet]. World Bank; Available from: http://hdl.handle.net/10986/24077

30. Verification of Performance in Results-Based Financing Programs : The Case of Plan Nacer in Argentina [Internet]. World Bank; Available from: http://hdl.handle.net/10986/21712

31. Verification of Performance in Results-Based Financing : The Case of Panama's PSPV Program [Internet]. World Bank; Available from: http://hdl.handle.net/10986/25126

32. De Naeyer L, Baptist Habaguhirwa J, Ndizeye C. Report of Audit on: Quantity verification and client satisfaction Quality counter verification and Performance Based Financing system & procedures [Internet]. Management Sciences for Health; 2011 Apr [cited 2016 Dec 28]. Available from: http://siteresources.worldbank.org/EXTPBFTOOLKIT/Resources/7364043-1386179756147/Round4_May11.pdf

33. Cordaid International [Internet]. [cited 2021 Feb 1]. External Verification Performance Based Financing in Healthcare Sierra Leone. Available from: https://www.cordaid.org/en/publications/external-verification-performance-based-financing-healthcare-sierra-leone/

34. Turcotte-Tremblay AM, Gali-Gali IA, De Allegri M, Ridde V. The unintended consequences of community verifications for performance-based financing in Burkina Faso. Social Science & Medicine. 2017 Oct 1;191:226–36.

35. Institute for Health Metrics and Evaluation [Internet]. [cited 2024 Feb 26]. GBD Results. Available from: https://vizhub.healthdata.org/gbd-results

36. Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. Randomized, Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk: The ANRS 1265 Trial. PLOS Medicine. 2005 Oct 25;2(11):e298.

37. Bailey RC, Moses S, Parker CB, Agot K, Maclean I, Krieger JN, et al. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. The Lancet. 2007 Feb 24;369(9562):643–56.

38. Gray RH, Kigozi G, Serwadda D, Makumbi F, Watya S, Nalugoda F, et al. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. The Lancet. 2007 Feb 24;369(9562):657–66.

39. RDQA Spreadsheet Tool — MEASURE Evaluation [Internet]. [cited 2019 Jun 27]. Available from: https://www.measureevaluation.org/resources/training/capacity-building-resources/hiv-english/session-24-field-visit/Routine%20Data%20Quality%20Assessment%20Tool.xls/view

40. Kuunibe N, Lohmann J, Schleicher M, Koulidiati JL, Robyn PJ, Zigani Z, et al. Factors associated with misreporting in performance-based financing in Burkina Faso: Implications for risk-based verification. Int J Health Plann Manage. 2019 Oct;34(4):1217–37.

41. Grover D, Bauhoff S, Friedman J. Using supervised learning to select audit targets in performance-based financing in health: An example from Zambia. PLOS ONE. 2019 Jan 29;14(1):e0211262.

42. Turcotte-Tremblay AM, Gali Gali IA, Ridde V. An Exploration of the Unintended Consequences of Performance-Based Financing in 6 Primary Healthcare Facilities in Burkina Faso. Int J Health Policy Manag. 2022 Feb 1;11(2):145–59.

43. Turcotte-Tremblay AM, De Allegri M, Gali-Gali IA, Ridde V. The unintended consequences of combining equity measures with performance-based financing in Burkina Faso. Int J Equity Health. 2018 Sep 24;17(1):109.

44. Zitti T, Ridde V, COULIBALY A. From Measuring User Satisfaction to Improving Quality of Care: Community Verifications and Performance-Based Financing in District Hospitals in Mali. 2022.

45. Bates MA, Glennester R. The Generalizability Puzzle (SSIR). Stanford Social Innovation Review [Internet]. 2017 [cited 2017 May 24]; Available from: https://ssir.org/articles/entry/the_generalizability_puzzle

46. Field B, Booth A, Ilott I, Gerrish K. Using the Knowledge to Action Framework in practice: a citation analysis and systematic review. Implementation Science. 2014 Nov 23;9(1):172.

47. Kalk A, Paul FA, Grabosch E. “Paying for performance” in Rwanda: does it pay off? Trop Med Int Health. 2010 Feb;15(2):182–90.

48. Piattifuenfkirchen M, Hadley S, Otto MA, Mathivet B. Alignment of Performance-Based Financing in Health with the Government Budget : A Principle-Based Approach. Health, Nutrition, and Population Discussion Paper Washington, DC : World Bank Group [Internet]. 2021 Aug 1 [cited 2023 Jun 6]; Available from: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/935821632462316181/alignment-of-performance-based-financing-in-health-with-the-government-budget-a-principle-based-approach

49. Paul E, Bodson O, Ridde V. What theories underpin performance-based financing? A scoping review. Journal of Health Organisation and Management. 2021 Jan 20;ahead-of-print.

50. Kingori P, Gerrets R. Morals, morale and motivations in data fabrication: Medical research fieldworkers views and practices in two Sub-Saharan African contexts. Soc Sci Med. 2016 Oct;166:150–9.

51. Kingori P, Gerrets R. The masking and making of fieldworkers and data in postcolonial Global Health research contexts. Critical Public Health. 2019 Aug 8;29(4):494–507.

52. Biruk C. Cooking Data: Culture and Politics in an African Research World [Internet]. Duke University Press; 2018 [cited 2023 Jun 7]. Available from: https://library.oapen.org/handle/20.500.12657/30253

53. Bouter LM. Commentary: Perverse Incentives or Rotten Apples? Accountability in Research. 2015 May 4;22(3):148–61.

54. NSF Fellows' perceptions about incentives, research misconduct, and scientific integrity in STEM academia | Scientific Reports [Internet]. [cited 2023 Jun 6]. Available from: https://www.nature.com/articles/s41598-023-32445-3

55. Chimhutu V, Lindkvist I, Lange S. When incentives work too well: locally implemented pay for performance (P4P) and adverse sanctions towards home birth in Tanzania - a qualitative study. BMC Health Services Research. 2014 Jan 18;14(1):23.

56. Feldacker C, Murenje V, Makunike-Chikwinya B, Hove J, Munyaradzi T, Marongwe P, et al. Balancing competing priorities: Quantity versus quality within a routine, voluntary medical male circumcision program operating at scale in Zimbabwe. PLOS ONE. 2020 Oct 13;15(10):e0240425.

57. An Exploration of the Unintended Consequences of Performance-Based Financing in 6 Primary Healthcare Facilities in Burkina Faso - PMC [Internet]. [cited 2022 Sep 8]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9278611/

58. Paul E, Sossouhounto N, Eclou DS. Local stakeholders' perceptions about the introduction of performance-based financing in Benin: a case study in two health districts. Int J Health Policy Manag. 2014 Sep 26;3(4):207–14.

59. Oxman AD, Fretheim A. Can paying for results help to achieve the Millennium Development Goals? A critical review of selected evaluations of results-based financing. Journal of Evidence-Based Medicine. 2009;2(3):184–95.

60. Fritsche G. Performance-Based Financing Toolkit | RBF Health [Internet]. The World Bank; [cited 2016 Nov 13]. Available from: https://www.rbfhealth.org/resource/performance-based-financing-toolkit-0

61. EvalCommunity. Effective Budgeting for MEAL in Development Projects [Internet]. 2023 [cited 2023 Aug 10]. Available from: https://www.evalcommunity.com/hire/budgeting-for-meal/

62. SAGE Publications Inc [Internet]. 2023 [cited 2023 Nov 29]. Utilization-Focused Evaluation. Available from: https://us.sagepub.com/en-us/nam/utilization-focused-evaluation/book268943

63. Vergeer P, Heard AC, Josephson E, Fleisher L. Verification in results-based financing for health [Internet]. The World Bank; 2016 Dec [cited 2019 May 14] p. 1–22. Report No.: 112342. Available from: http://documents.worldbank.org/curated/en/504661485897909610/Verification-in-results-based-financing-for-health

64. Health Systems 20/20 U. Verification in Performance Based Incentive Schemes [Internet]. 2012 [cited 2016 Oct 27]. Available from: https://www.hfgproject.org/wp-content/uploads/2015/02/Verification-in-Performance-Based-Incentive-Schemes1.pdf

65. Kuunibe N, Lohmann J, Schleicher M, Koulidiati JL, Robyn PJ, Zigani Z, et al. Factors associated with misreporting in performance-based financing in Burkina Faso: Implications for risk-based verification. Int J Health Plann Manage. 2019 Oct;34(4):1217–37.

66. Turcotte-Tremblay AM, Gali-Gali IA, De Allegri M, Ridde V. The unintended consequences of community verifications for performance-based financing in Burkina Faso. Social Science & Medicine. 2017 Oct 1;191:226–36.

67. Wiley.com [Internet]. [cited 2023 Nov 29]. Multiple Imputation for Nonresponse in Surveys | Wiley. Available from: https://www.wiley.com/en-us/Multiple+Imputation+for+Nonresponse+in+Surveys-p-9780471655749

68. Honaker J, King G, Blackwell M. Amelia II: A Program for Missing Data [Internet]. [cited 2024 Feb 29]. Available from: https://gking.harvard.edu/Amelia

69. Chesnaye NC, Stel VS, Tripepi G, Dekker FW, Fu EL, Zoccali C, et al. An introduction to inverse probability of treatment weighting in observational research. Clinical Kidney Journal. 2022 Jan 1;15(1):14–20.

70. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value Health. 2010;13(2):273–7.

71. Xiao Y, Bochner AF, Makunike B, Holec M, Xaba S, Tshimanga M, et al. Challenges in data quality: the influence of data quality assessments on data availability and completeness in a voluntary medical male circumcision programme in Zimbabwe. BMJ Open [Internet]. 2017 Jan 27 [cited 2019 Feb 18];7(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5278271/

72. Yourkavitch J, Prosnitz D, Herrera S. Data quality assessments stimulate improvements to health management information systems: evidence from five African countries. J Glob Health. 2019 Jun;9(1):010806.