

Coaching for quality improvement under performance-based contracting: a theory-of-change evaluation in Honduras

Short title: Coaching for quality improvement in Honduras

Authors and affiliations:

Wolfgang Munar, MS, MD¹

L. Esther Aranda, DrPH, MPH, MCP²

Molly E. Lauria, MPH¹

Pedro Bernal, PhD⁴

Cinzia Innocenti, PhD⁵

Marvin Rodríguez, BA, MS⁶

Affiliations:

¹ Department of Global Health, Milken Institute School of Public Health, The George Washington University, Washington, DC, USA.

² Independent Researcher, Bethesda, MD, USA

³ Department of Global Health, Milken Institute School of Public Health, The George Washington University, Washington, DC, USA.

⁴ Health, Nutrition and Population Division, Inter-American Development Bank, Washington, DC, USA

⁵ Independent Researcher, San Salvador, El Salvador

⁶ Independent Researcher, San Jose, Costa Rica

Corresponding author:

Wolfgang Munar, MS, MD

Associate Professor and Director, Gill-Lebovic Centre for Community Health in the Caribbean and Latin America - Department of Global Health

Milken Institute School of Public Health, The George Washington University

950 N. Hampshire Ave, NW, Washington, DC 20552, USA

Email: wolfgangmunar@gwu.edu

Telephone: +1-202-995-1015

ORCID: <http://orcid.org/0000-0002-9234-987X>

ABSTRACT

Introduction. Practice coaching is increasingly used to strengthen quality improvement (QI) capacity in primary healthcare (PHC) systems in low- and middle-income countries (LMICs), yet the causal pathways through which it shifts provider behaviour, and the systemic conditions that enable or constrain those pathways, remain under-theorised. Using a theory-based qualitative evaluation, we examined how and why a practice-coaching intervention influenced QI in cervical cancer screening (CCS) and antenatal care (ANC) within Honduras's decentralised PHC system during the third phase of the Salud Mesoamerica Initiative (SMI).

Methods. We conducted a within-case explanatory case study. A programme theory was reconstructed before data collection and iteratively refined against evidence. Data comprised semi-structured interviews with 11 mid-level managers, 6 PHC team medical leads, and 2 regional managers, complemented by direct observation and document review. We applied combined deductive–inductive coding, thematic analysis, and pattern matching, and reporting per COREQ.

Findings. We identified four causal patterns that refined the initial programme theory. Three were activated pathways: (1) novel professional identity among participating managers; (2) collective efficacy and data-driven learning, sustained through verifiable progress on observable indicators, strong for CCS but null for ANC, where outcomes were less attributable to teams' actions; and (3) relational coordination, psychological safety, and trust, which provided the interpersonal basis for the first two. A fourth, unanticipated pattern showed structural misalignment between coaching's enabling, learning-based logic and the directive, punitive logic of Honduras's performance-based contracting environment, confining gains to localised "enabling bubbles."

Conclusion. Coaching can activate meaningful QI pathways in LMIC primary care, but sustained, equitable impact requires deliberate alignment between coaching's learning-oriented principles and the institutional performance-management architecture, and matching of coaching investment to clinical processes with observable, attributable outcomes.

Keywords: Performance management; Practice coaching; Primary Health Care; Honduras; Theory of Change Evaluation; Quality improvement; Accountability.

KEY MESSAGES

What is already known on this topic

Practice coaching can improve clinical processes in primary care, but the evidence base is concentrated in high-income settings, effectiveness varies across contexts, and evaluations rarely test explicit causal hypotheses about how coaching produces change in LMIC health systems.

What this study adds

We identified four causal-link patterns: three activated pathways: novel managerial identity, data-driven collective efficacy, and relational coordination; and a fourth unanticipated pattern of misalignment between coaching's enabling logic and the directive, punitive logic of performance-based contracting, which limited gains to localised "enabling bubbles" of improvement.

How this study might affect research, practice or policy

Sustained impact of coaching in decentralised LMIC PHC systems requires deliberate alignment between coaching's learning-oriented principles and the performance verification and the system's accountability architecture. Coaching investment should be matched to clinical processes whose outcomes are observable and attributable; field mentoring should be maintained during scale-up; and longitudinal, comparative research should characterise how directive and enabling performance-management logics interact over time.

INTRODUCTION

Expanding access to PHC in LMICs has been a major achievement of global health investment over the past two decades, yet a persistent quality gap undermines these gains (1). Suboptimal provider practices remain a leading cause of preventable mortality and morbidity in LMICs, limiting population-level effective coverage. The Lancet Global Health Commission on High-Quality Health Systems estimated that poor-quality care accounts for a substantial share of excess mortality in low-income settings and called for system-level strategies that go beyond access (2).

Over the past four decades, LMIC health systems have increasingly adopted performance management (PM) strategies to align provider behaviour with clinical standards and organisational priorities. However, an evidence gap map of such strategies in LMIC primary care found that the literature is dominated by individual-level financial incentive schemes such as pay-for-performance and results-based financing, while organisationally focused strategies and outcomes remain understudied and poorly theorised (3).

Practice coaching has emerged as one such organisationally focused strategy. Defined as a collaborative, relationship-based intervention, it is designed to build organisational capacity for QI by empowering frontline teams to use data to make decisions and solve problems (4). Systematic reviews show that coaching can improve clinical processes and, in some instances, patient outcomes, but the evidence base remains concentrated in high-income settings and effectiveness varies substantially across contexts (5–7). Supportive performance-improvement interventions, including supervision, mentoring, and coaching, have been implemented in LMICs (8), but evaluations rarely test explicit causal hypotheses about how and why coaching produces change (9). This theoretical thinness makes it difficult to distinguish between theory failure, where the intervention's causal logic is itself flawed, and implementation failure, where the logic is sound, but execution was inadequate (10).

Conceptual framework

We adopt a complexity-informed conceptual framework (11) as our organising lens for understanding how PM operates within complex health systems. The framework distinguishes two ideal-typical orientations to PM. Directive approaches seek to control provider behaviour through hierarchical answerability, targets, routine monitoring, and extrinsic incentives. Rooted in agency theory (12,13), such approaches assume that providers are self-interested actors whose behaviour diverge from organisational goals unless externally monitored and rewarded; such is the logic underpinning pay-for-performance schemes, whose results in LMICs have been mixed (14). Enabling approaches, by contrast, aim to promote collective responsibility, intrinsic motivation, trust, and team self-organisation through shared sensemaking and iterative learning. The latter draws on stewardship theory's (15) premise that providers are intrinsically motivated professionals broadly aligned with organisational goals (16).

The directive–enabling distinction is analytic rather than descriptive: real-world health systems contain both elements, and their interaction shapes how PM strategies evolve. Practice coaching sits within the enabling orientation; collaborative, relationship-based, and designed to strengthen frontline teams' capacity to identify problems, interpret data, and develop locally owned solutions rather than enforce compliance with externally imposed targets (8). What remains less well understood is what happens when an enabling intervention is introduced into a PHC system whose context reflects a different, or mixed, set of management logics. The latter is the focus of the evaluation presented here.

Context and intervention

Honduras is one of Latin America's poorest and most unequal countries. By 2024, 49.3% of the population lived below US\$6.85 per day (2017 PPP), exceeding the Central American average of 37.5%; rural poverty reached 63%, with half of rural inhabitants in extreme poverty (17,18). These conditions shape both the demand for and the delivery of PHC, particularly in the remote municipalities where the intervention under study was implemented.

Honduras's PHC system combines decentralisation with New Public Management principles, including performance-based contracts and pay-for-performance incentives tied to maternal and child health targets (19–21). The Ministry of Health (*Secretaría de Salud* - SESAL) retains central stewardship, setting policy, establishing clinical standards, and allocating resources, while delegating service delivery to a diverse set of PHC operators: municipal governments, associations of municipalities, non-governmental organisations, and SESAL-run facilities. SESAL delegated oversight and technical coordination of operators to its regional offices (21). Operators receive flexible budgets tied to maternal and child health targets, with financial penalties for unmet performance. This contractual architecture corresponds to a directive PM environment. The diversity of operator types functioning under distinct institutional arrangements within a single national programme creates natural internal variation, making Honduras an instructive setting for evaluative research on PM strategies such as coaching.

The SMI was a multi-phase, multi-country public–private partnership started in 2011. It was established to improve health system performance and reproductive, maternal, and child health outcomes in the poorest rural municipalities in Mesoamerica (22). Beginning during SMI's third phase (2019-2022) and extending into one additional year, a practice coaching intervention was introduced to strengthen QI capacity of mid-level managers and frontline teams among participating operators. A concurrent impact evaluation (23) found a positive effect on cervical cancer screening (CCS) and improved QI knowledge among managers, but no significant effect on antenatal care (ANC) guideline adherence and no measurable change in organisational culture.

This study addresses the overarching learning question: How and why does a practice coaching intervention influence QI in the Honduran PHC system? It is operationalised through three specific questions: (1) To what extent were the hypothesised causal links between coaching activities and the intermediate outcomes of provider capability, opportunity, and motivation realised? (2) To what extent were the underlying causal-link assumptions supported by the empirical evidence? (3) How did the structural diversity of Honduras's decentralisation context shape the activation of the coaching theory?

METHODS

Evaluation design

We conducted a theory-of-change (ToC) evaluation as a within-case explanatory case study of the practice coaching intervention implemented during SMI Phase 3 in Honduras. ToC evaluation, a type of theory-driven evaluation (24,25), treats a programme's articulated theory as a series of testable causal hypotheses and assesses them against empirical evidence, asking how and under what conditions causal pathways are activated, partially realised, or constrained rather than only whether the programme worked. ToC evaluation is suited to addressing the gaps described earlier. By making causal pathways and programme actors' assumptions explicit, it enables researchers to move beyond asking whether a programme worked to explaining how and why its effects occurred or did not (26–28).

The choice of a within-case design was appropriate because the evaluation focuses on a single bounded case, the coaching intervention as implemented in Honduras in SMI-supported municipalities, while exploiting the natural internal variation created by the country's diverse operator arrangements to examine how the same intervention landed in different organisational environments. The unit of analysis is the coaching intervention as embedded in the Honduran PHC system during SMI Phase 3.

This study operates from a post-positivist integration of scientific realism and pragmatism (29) and is complemented by complexity and critical realism perspectives to help understand how program resources trigger psychological and social mechanisms in specific contexts (25,30,31). It assumes that real causal processes exist within complex health systems and can be identified through systematic, theory-driven inquiry, while acknowledging that knowledge of those processes is constructed through both participatory dialogue and empirical testing (32). Accordingly, causal-link assumptions embedded in the programme theory are treated as testable hypotheses rather than fixed truths, and the evaluation seeks coherence between observed evidence and a broader body of theory.

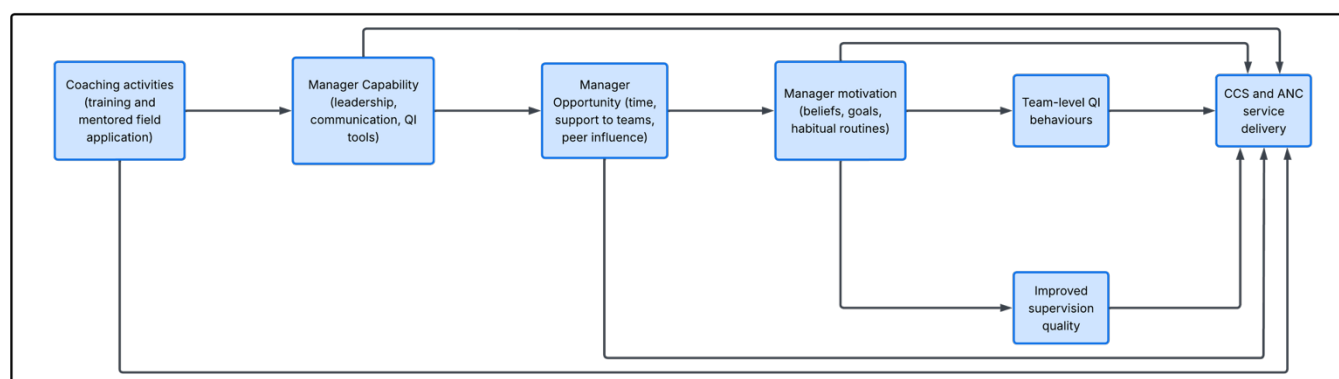
Programme theory reconstruction

Prior to primary data collection, we reconstructed the programme theory following procedures for surfacing programme logic (33), drawing on programme design and policy documents, participatory workshops with stakeholders, semi-structured interviews with programme actors at multiple system levels, and a targeted scoping of social and behavioural science literature. The reconstructed theory was supported by self-determination theory (34) and adult learning theory (35) at the individual level, social cognitive theory (36) at the interpersonal level, and situated learning theory (37) at the organisational level. Theory reconstruction was also informed by the COM-B model of behaviour change (38) and a systematic process of assumption surfacing (39). The theory was refined after data collection and analysis.

The preliminary programme theory in Figure 1 is represented as a linear causal chain in which coaching activities, including in-person training and field mentoring, would build mid-level managers' capability (leadership, communication, and QI tool use), which would in turn expand their opportunity to apply new practices (protected time, support to teams, and peer influence), thereby strengthening their motivation (beliefs, goals, and habitual routines). Motivation was hypothesised to translate into two proximal behavioural outcomes: team-level QI behaviours and improved supervision quality, which would jointly drive the distal service-delivery outcomes of CCS and ANC adherence. Each arrow represents a testable causal-link assumption.

Participants were recruited through a purposive, maximum-variation strategy designed to maximise the information power of the sample with respect to the causal hypotheses being examined (39).

Figure 1. Preliminary Programme Theory



Recruitment proceeded in two stages. First, at the operator level, we invited mid-level managers from every decentralised PHC Operator engaged in the intervention, ensuring that the full institutional range of contracted providers was represented. Second, within each Operator we drew a smaller, nested sub-sample of PHC team medical leads, selected for being supervised by the managers sampled and by their direct, first-hand experience of the coaching activities at the team level. Two regional SESAL managers were included to triangulate operator- and team-level accounts against their perspective.

An interview guide (Appendix A) was developed from the programme theory and the above theoretical framework; it covered eight thematic areas, including training history, supervision and monitoring practices, knowledge of national CCS and ANC standards, responses to the coaching intervention, QI implementation experience, and perceptions of external influences on national-standard implementation. Secondary data comprised SESAL and Honduran Government policy documents; sector diagnostics from regional development banks and other multilateral organisations; programme documents and IHME performance verification surveys; and targeted peer-reviewed and grey literature.

Data analysis

Analysis followed a structured, iterative theory-testing strategy. We built chronological narratives, then three coders applied combined deductive coding derived from the programme theory and theoretical framework, and inductive coding to identify emergent themes, working iteratively among the articulated theory, the impact evaluation findings (23), and the primary data. Pattern matching (24,40) was used to compare hypothesised and observed causal patterns, and alternative explanations were systematically assessed. Coding was conducted in two cycles by a three-member team in ATLAS.ti, with intercoder consistency established on shared transcripts in the two cycles of coding; an audit trail documented coding decisions and the evolution of the analytical framework. The codebook is presented in Appendix A.

Table 1. Participant characteristics by respondent type

Respondent type	Total	Women	Men
Mid-level managers / supervisors and medical coordinators	11	8	3
PHC team medical leads	6	5	1
Regional SESAL Managers	2	1	1
Total	19	14	5

Operational and budgetary constraints impeded member checking. Credibility was instead strengthened through triangulation across interviews, programme documents, and the impact evaluation, as well as through systematic searches for disconfirming evidence. Dependability was supported by collaborative two-cycle coding and independent coding by team members not previously involved in the programme. Reporting follows the COREQ checklist (41) presented in Appendix A.

Positionality

Four of the authors have prior involvement with the SMI. Authors A, B and E have participated in evaluations of the SMI since 2016, providing deep contextual knowledge of the programme and its evolution across phases. Author D is the principal investigator of the impact evaluation of the coaching

intervention and works at the IDB, one of the two funding sources for this evaluation. These positions provide interpretive depth but also create a risk of confirmation bias toward a favourable reading of the intervention. We mitigated this risk through three strategies: (i) two team members (Authors C and F) without prior SMI involvement participated in data analysis; (ii) disconfirming evidence was actively sought throughout analysis, helping identify patterns not present in the initial program theory; and (iii) all evolving interpretations were subject to full-team review before finalisation.

Patient and public involvement

Patients and members of the public were not involved in the design, conduct, reporting, or dissemination plans of this research. The study focused on the experiences and practices of mid-level managers, primary healthcare team medical leads, and regional health authority managers; no patient-level data were collected.

FINDINGS

We identified four causal-link patterns through which the coaching intervention interacted with Honduras's PHC system. Three describe activated pathways linking coaching activities to changes in mid-level managers' and frontline teams' capabilities, opportunities, and motivation. A fourth, unanticipated pattern describes the systemic conditions that constrained those pathways and confined their effects to localised "enabling bubbles" of improvement. Table 2 summarises the patterns and their evidential support; Table 3 maps illustrative quotes to each pattern by respondent type. Additional supporting evidence is presented in Appendix B.

Pattern 1. Professional identity transformation and enhanced confidence

Mid-level managers consistently reported that their prior medical and nursing education had not equipped them for supervisory roles, leading them to default to the directive, command-and-control styles prevailing in their organisational environment. The coaching intervention combined structured workshops on QI tools with mentored field application of soft skills. Participants framed this combination

as filling a gap not addressed by university training. One manager reported, "We received training on soft skills like assertive communication and decision-making... tools for continuous improvement... and finally, data analysis for timely decision making."

Table 2. Causal-link patterns, theoretical foundations, and evidential support.

Pattern	Causal-link assumption tested	Theoretical foundation	Evidential support
1. Professional identity transformation	If managers receive integrated soft- and hard-skill training plus mentored field practice, then they will shift from directive to facilitative supervision.	Adult learning theory (Knowles); social identity theory (Hogg & Terry).	Strong. Consistent across 11 manager interviews; partial fidelity loss in cascade training (1 case).
2. Collective efficacy and data-driven learning	If teams apply QI tools and receive timely data feedback, then collective efficacy will emerge and behaviour change will be sustained.	Social cognitive theory (Bandura); communities of practice (Lave & Wenger).	Partially supported. Strong for CCS where indicators are observable; weak for ANC; gaming hypothesis raised by 1 regional director.
3. Relational coordination, psychological safety, and trust-building	If coaching strengthens communication and supportive leadership, then teams will experience increased psychological safety and improved coordination.	Self-determination theory (Ryan & Deci); psychological safety (Edmondson).	Strong. Cited across manager and provider interviews; rapid catalytic shifts triggered by external validation.
4. Incentive misalignment and systemic constraints	Not in initial theory. Emergent: enabling coaching logic conflicts with directive, punitive performance verification and resource constraints, confining gains to localised enabling bubbles.	Performance management in complex systems (Newton-Lewis et al., 2021); stewardship vs agency theory.	Strong. Convergent across PHC managers, providers, and regional directors viewed from complementary angles.

Managers described a deliberate shift from enforcer to facilitator behaviours, repositioning themselves alongside their teams: "As licensed nurses, we are often made to command, and almost all the operators practically command. *'I am the boss. I do this. And I want this and that.'*" (Supervisor). The shift cascaded into routine supervision and peer development with PHC team leads. One PHC team lead described receiving only second-hand training through a one-day cascade and acquiring a partial version of the content.

Pattern 2. Collective efficacy and data-driven learning

The second pattern describes a reinforcing feedback loop through which teams sustained behaviour change by verifying their own progress against measurable indicators. PHC teams moved from passive compliance with reporting requirements to active use of data for problem-solving. Visible progression on indicators triggered satisfaction and motivated application of QI tools. One manager noted, "I think it helped us a lot to reach the goal. For example, with cervical cancer. We were in yellow, then we went to green. It has helped us organise our work better."

The pattern was strongest in CCS, where process indicators made progress immediately observable and where teams could trace concrete clinical impact. A PHC team lead noted, "Now we are more organised. The cytology is done on the proper day. The sample is sent to the lab. The patient is called. She is told that her results are here."

One PHC team lead described the current CCS target as "only 13 cytologies per month, which is very low," producing what was perceived as hollow rather than motivating achievement. This raises the question of potential gaming in the respondent's specific operator. However, the extent to which such CCS targets were softened ex-post to ensure compliance with the performance targets cannot be addressed with the present data.

Pattern 3. Relational coordination, psychological safety, and trust-building

The third pattern served as the interpersonal "connective tissue" enabling previous patterns to take hold. Managers reported that coaching addressed gaps in collaborative capacity that technical QI tools alone could not bridge. Supervision was increasingly experienced as helpful rather than punitive; as one manager stated, "My team views me as a co-worker; we are on the same level. I tell them, 'We are working toward the same goal.'" Reframed as partners, managers described modelling open

relationships and teams reportedly redistributing workload, reducing overtime, and improving work–life balance: "We improved our interpersonal relationships, distributed the work more effectively, and [work] was less of a burden."

External validation appeared to serve as a catalyst for psychological safety. One team lead described initial resistance reversed by the first mentoring visit, when the international team "started congratulating us. The team was motivated and totally changed." This suggests that supportive external feedback can trigger emotional reorientation, complementing the slower identity transformation of Pattern 1.

Pattern 4. Incentive misalignment and systemic constraints

A fourth pattern, unanticipated by the initial programme theory, emerged from respondents' accounts of the broader institutional environment. Participants described a structural mismatch between the coaching intervention's enabling, learning-based logic and the directive, inspection-oriented logic that characterised SESAL's performance verification. Regional offices' audits were experienced as unidirectional: "Regarding the 'region,' they do a type of follow-up. They do something called technical audits, where they come in, identify the problem, and leave a lot of requirements that sometimes people don't even understand. And they leave and never come back. Or they come back one or two years later. Or six months later. And there is no review of previous commitments or follow-up of the process." Punitive consequences are direct: "They monitor compliance with monitoring indicators; they do this every three months. They choose a health facility at random. And that is punitive for us. If we don't have something, they take away funds."

Regional managers described the same monitoring architecture in a markedly different manner, framing it as the principal mean for generating accountability and noting that "if you do not have a goal, you are just going to be there and give the minimum you can give." The two perspectives are complementary but disconnected (Box 1). Layered on top of this, chronic resource shortages, including the absence of basic supplies such as cytology gloves, staff turnover, and delayed operator payments, the cumulative

effect was that improvements remained confined to localised enabling bubbles whose boundaries were set by the surrounding directive logic.

Box 1. Two views of one monitoring system

Frontline experience (Mid-level manager): "They do something called technical audits where they come, they identify the problem, and they leave a lot of requirements here that sometimes people don't even understand. And they leave and never come back. They come back one or two years later... There is no review of previous commitments or follow-up of the process."

Regional managerial framing (Regional Director): "With constant monitoring of the health centres, people work a little bit based on a goal... If you do not have a goal to establish, if you do not know your indicators, then you are just going to be there and give the minimum you can give."

Frontline PHC team directors and Regional Health Authority directors articulated behavioural outcomes across various levels of engagement. PHC directors offered ground-level, experientially detailed accounts that included negative outcomes and demoralisation, alongside specific positive routines, such as patient sign-in books to verify delivery of CCS results. Regional directors described aggregated, system-level outcomes framed within the logic of compliance and oversight. These complementary yet disconnected perspectives show that the concrete demotivators most salient to operators, such as funding delays, stock-outs, and job insecurity, were largely absent from regional directors' accounts, consistent with Pattern 4.

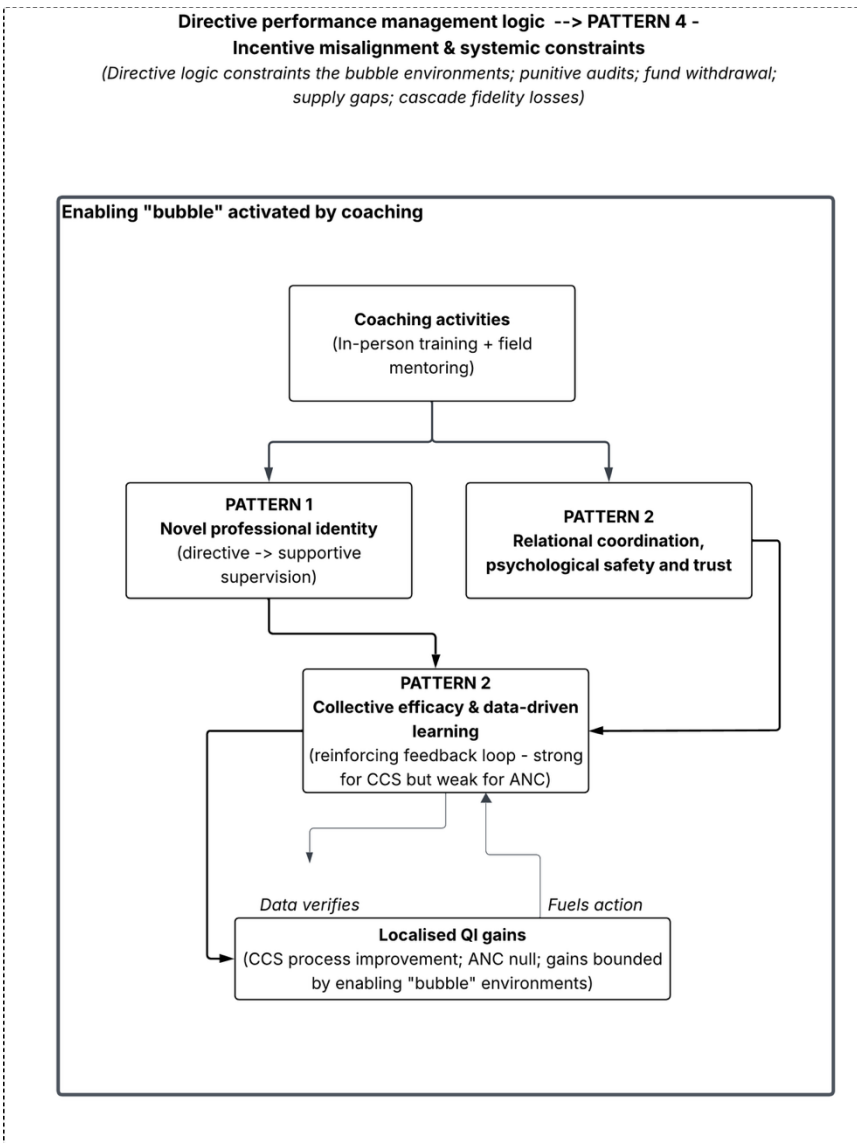
Table 3. Illustrative quotes by causal-link pattern and respondent type

Pattern	Mid-level manager / Coordinator	PHC Medical Director / Regional Director
1. Identity transformation	"Instead of accusing or pointing fingers, we first focus on the positive aspects... allowing the provider to guide us on how to improve, rather than dictating to them what to do." (OP-12)	"They learned... measuring what you do is very important... and learning to analyse the results of what they are getting." (HF-04)
2. Collective efficacy	"It helped us a lot to reach the goal. With cervical cancer, we were in yellow, then we went to green." (OP-08)	"There was improvement, and yes, it can be done." (HF-02)
3. Psychological safety	"My team views me as a co-worker; we are on the same level. I tell them, 'We are working toward the same goal.'" (OP-02)	"The international coaching team... started congratulating us. The team was motivated and totally changed." (HF-03)
4. Incentive misalignment	"They monitor compliance... that is punitive for us. If we don't have something, they take away funds." (OP-03)	"If you do not have a goal to establish... then you are just going to be there and give the minimum you can give." (DIR-02)

DISCUSSION

This theory-of-change evaluation identified four causal-link patterns that, taken together, refine the intervention's initial programme theory. Three describe activated pathways at the individual, team, and

Figure 2. Refined programme theory of the C4QI coaching intervention



interpersonal levels: professional identity transformation from directive to supportive supervision (Pattern 1); collective efficacy and data-driven learning, through which teams sustained behaviour change by verifying measurable progress (Pattern 2); and relational coordination, psychological safety, and trust-building, which constituted the interpersonal basis for the first two (Pattern 3). A fourth pattern, unanticipated by the initial theory, describes a structural misalignment between the intervention's enabling, learning-based logic and the directive, inspection-oriented logic of

Honduras's established performance-based contracting environment (Pattern 4). These patterns are

synthesised in the refined programme theory in Figure 2. In the refined model, three coaching-activated mechanisms generate localised QI gains within an enabling 'bubble', which is constrained by the directive performance-management logic of the wider system (Pattern 4). The CCS/ANC asymmetry is represented in the strength of the Pattern 2 reinforcing loop.

The clearest explanatory contribution of this evaluation is to provide plausible explanation for the asymmetric findings in the concurrent impact evaluation. Cervical cancer screening is a service whose process indicators, such as completion rates and the use of dashboard colour bands, made team performance immediately observable, and whose clinical impact is occasionally vivid and proximate. These features appear to have activated the Pattern 2 reinforcing loop: data-verified success strengthened collective efficacy, which motivated continued application of QI tools, which produced further measurable gains, consistent with social cognitive theory's predictions about performance feedback (36). Antenatal care adherence, by contrast, is a multi-step process whose outcomes are diffuse, delayed, and harder to verify through routine indicators. The reinforcing loop was activated more weakly, providing a plausible mechanism-level explanation for the null ANC effect that an outcome-focused trial cannot supply on its own. Also, provider adherence, while necessary, is not sufficient to address the demand side of ANC and CCS. This may generalise beyond Honduras: coaching interventions are likely to produce stronger, faster gains where clinical processes generate observable, attributable, and frequent feedback, and weaker gains where they do not.

Our findings indicate that coaching's enabling logic, triggered by collaborative problem-solving, data for learning instead of judgement, and supportive supervision, operated within an institutional context characterised by a directive logic: performance-based contracts and pay-for-performance with explicit financial penalties, episodic technical audits experienced as unidirectional and demoralising, and indicator monitoring tied to funding withdrawal. A previously published framework on PM in complex adaptive health systems (11) predicted that such mixed institutional environments would produce outcomes shaped by the interaction of the two logics rather than by the intervention alone, and our

evidence seems to confirm such prediction in behavioural terms. The two logics did not simply coexist; they generated competing demands on frontline managers and teams, who were asked to conduct QI routines while being audited for indicator compliance. They also triggered the experiential demotivators that are most significant to providers, such as funding delays, supply stock-outs, and job insecurity, largely invisible to the regional managers in charge of system oversight. The result was the "enabling bubble" (42) phenomenon: real but spatially and temporally bounded improvement, contingent on local relational conditions and vulnerable to system constraints. This finding is consistent with the conceptualisation of health systems software and hardware (43,44); with Hood's (45) caution that performance-by-numbers regimes can simultaneously enhance and obstruct service performance, depending on organisational culture and the credibility of metrics; and with realist evaluation evidence from other LMIC settings (46–50) that motivational mechanisms are highly context-contingent.

Three implications for policy and practice are outlined below. First, sustaining coaching's gains in LMIC PHC systems requires deliberate alignment between the intervention's enabling principles and the institutional performance-management architecture, not merely co-location. Design considerations include remaking regional performance verification engagement as developmental rather than punitive, decoupling fund-withdrawal triggers from indicators where coaching is in active use and creating shared dashboards that make frontline experiential data legible to regional managers, which we term *supportive accountability*. Second, programme designers should match coaching investment to the feedback observability of the targeted clinical processes; where outcomes are diffuse, complementary measurement infrastructure (proxy process indicators, structured patient-flow tracking) is likely a precondition for the Pattern 2 loop to activate. Third, the train-the-trainer fidelity weakness that surfaced in Pattern 1 suggests that scale-up should preserve field mentoring rather than rely solely on train-the-trainer, cascade approaches, even at the cost of slower geographic reach.

Two research priorities emerge. First, longitudinal mixed-methods studies are needed to characterise how directive and enabling performance-management logics interact over time within single

decentralised health systems, and whether "enabling bubbles" expand, persist, or collapse once external coaching support is withdrawn. Also, comparative work across SMI countries would test whether the four causal-link patterns generalise or whether their relative strength varies with institutional configuration. Such work would inform the design and evaluation of coaching interventions in decentralised LMIC health systems and clarify the conditions under which enabling interventions can reshape, rather than merely coexist with, directive performance-management environments.

Strengths and limitations

Study strengths include the explicit theory-testing design, the iterative theory refinement, the triangulation of interview data with direct observation and programme documentation, and the complementary pairing with an impact evaluation. Limitations should be acknowledged. Sampling was limited to managers, PHC team leads, and regional directors; neither other team members nor patients were interviewed, and their perspectives may modify Pattern 3. The evaluation took place at the end of an established multi-phase programme, and even if the intervention included all operators in SMI areas, it raises the possibility that the most successful PHC teams were over-represented. Given well-known unintended effects of PM strategies (51), the gaming hypothesis raised by the "low cytology target" observation could not be resolved with the available data and requires future work.

CONCLUSION

Practice coaching can activate meaningful causal pathways toward quality improvement in PHC systems of LMICs, transforming supervision from directive to supportive, catalysing data-driven collective learning, and building the relational conditions on which both depend. In Honduras, however, these gains remained confined to localised "enabling bubbles" because the intervention's enabling logic operated within an institutional architecture organised around directive performance management. Sustained system-wide impact, therefore, requires deliberate alignment between coaching's learning-oriented principles and the regional verification and contracting architecture: what we term supportive

accountability. Further longitudinal and comparative research is needed to characterise how directive and enabling performance-management logics interact over time in decentralised health systems, and to inform the design of coaching interventions that can reshape, rather than merely coexist with, prevailing institutional logics.

Contributorship statement

Conceptualisation: WM and PB. Methodology: WM and PB. Programme theory reconstruction: WM, PB, and LEA. Investigation (data collection): CI. Formal analysis: LEA, MR, and MEL. Validation: WM. Writing — original draft: WM and LEA. Writing — review and editing: all authors. Visualisations (Figures 1 and 2): WM. Project administration: WM and PB. Funding acquisition: WM. Supervision: WM. All authors read and approved the final manuscript. WM is the guarantor and accepts full responsibility for the finished work, the conduct of the study, access to the data and the decision to publish.

Funding

This work was supported in whole or in part by [funder names anonymised for peer review]. The funders had no role in study design, data collection, analysis, interpretation of findings, or the decision to submit the manuscript for publication.

Competing interests

All authors have completed the ICMJE uniform disclosure form. The authors declare the following support for the submitted work: funding from a philanthropic organisation, channelled through a multilateral development bank to the authors' academic institution; and funding from a research centre at the authors' academic institution, which supported author participation (including a doctoral student author) and covers article processing charges. One author is employed by the multilateral development bank that funded the preceding quantitative component. One author reports a past grant (within the previous 36 months) from an international non-governmental organisation, under a separately funded award to a consortium at the authors' academic institution focused on routine immunization, unrelated

to the submitted work. No author reports other financial relationships with organisations that might have an interest in the submitted work in the previous three years, and no author reports other relationships or activities that could appear to have influenced the submitted work. Full funder names, author-specific disclosures, and signed ICMJE forms have been provided to the editorial office and will be inserted into the manuscript upon acceptance.

Patient and public involvement

Patients and members of the public were not involved in the design, conduct, reporting, or dissemination plans of this research. The study focused on the experiences and practices of mid-level managers, primary healthcare team medical leads, and regional health authority managers as the relevant programme actors for testing the reconstructed coaching theory of change; no patient-level data were collected.

Ethics

Ethical approval was obtained from the corresponding author's institutional review board [name and reference number anonymised for peer review] and from the relevant national health authority in Honduras. All participants provided written informed consent before participating in the study. Participation was voluntary, and participants were informed of their right to withdraw at any time without consequence. Confidentiality and anonymity were assured throughout data analysis and reporting.

Consent for publication

Not applicable. The manuscript does not contain any individual person's identifying data, images, or details.

Data availability statement

No data are publicly available. The study draws on confidential key-informant interviews conducted with a small, purposively selected sample of mid-level managers, primary healthcare team medical leads, and regional health authority managers. The limited number of identifiable role-holders makes release

of the transcripts, even in de-identified form, a risk to participant confidentiality. The data are therefore not publicly available, in line with the ethical restrictions approved by the Institutional Review Board and the terms of the written informed consent obtained from participants. Reasonable, restricted requests from qualified researchers may be directed to the corresponding author and will be considered on a case-by-case basis, subject to ethical and confidentiality safeguards.

Use of artificial intelligence

The authors used Grammarly Pro solely for grammatical review; all analytical interpretations, theoretical claims, and evidentiary judgements are the authors' own. The authors take full responsibility for the integrity and accuracy of all manuscript content.

Acknowledgements

The authors thank the mid-level managers and primary healthcare team medical leads who participated in the interviews and candidly shared their experiences. We are grateful to colleagues at [institutional partners anonymised for peer review] for their support. Any errors or interpretive judgements are the authors' alone.

Supplementary materials

Appendix A — Interview guidelines, codebook, and COREQ checklist. Appendix B — Indicative quotes by causal-link pattern and respondent type.

REFERENCES

1. Rowe SY, Peters DH, Holloway KA, Chalker J, Ross-Degnan D, Rowe AK. A systematic review of the effectiveness of strategies to improve health care provider performance in low-and middle-income countries: methods and descriptive results. *PloS One*. 2019;14(5).

2. Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. *The Lancet Global Health*. 2018;6(11):e1196–252.
3. Munar W, Snilstveit B, Aranda LE, Biswas N, Baffour T, Stevenson J. Evidence gap map of performance measurement and management in primary healthcare systems in low-income and middle-income countries. *BMJ Global Health*. 2019;4(Suppl 8).
4. Grumbach K, Bainbridge E, Bodenheimer T. Facilitating improvement in primary care: the promise of practice coaching. Vol. 15. 2012. p. 1–14.
5. Ballengee LA, Rushton S, Lewinski AA, Hwang S, Zullig LL, Ricks KAB. Effectiveness of Quality Improvement Coaching on Process Outcomes in Health Care Settings: A Systematic Review. *J Gen Intern Med*. 2022;37(4):885–99.
6. Walunas TL, Ye J, Bannon J, Wang A, Kho AN, Smith JD, et al. Does coaching matter? Examining the impact of specific practice facilitation strategies on implementation of quality improvement interventions in the Healthy Hearts in the Heartland study. *Implementation Science*. 2021 Mar 31;16(1). <https://doi.org/10.1186/s13012-021-01100-8>
7. Wang A, Pollack T, Kadziel LA, Ross SM, McHugh M, Jordan N. Impact of practice facilitation in primary care on chronic disease care processes and outcomes: a systematic review. *J Gen Intern Med*. 2018;33(11):1968–77.
8. Vasan A, Mabey DC, Chaudhri S, Brown Epstein HA, Lawn SD. Support and performance improvement for primary health care workers in low-and middle-income countries: a scoping review of intervention design and methods. *Health policy and planning*. 2017;32(3):437–52.
9. Floate H, Durham J, Marks GC. Moving on from logical frameworks to find the ‘missing middle’ in international development programmes. *J Dev Effect*. 2019;11(1):89–103.

10. Suchman EA. Principles and practice of evaluative research. New York: Russell Sage Foundation; 1967.
11. Reference anonymised for review
12. Mitnick BM. The theory of agency. Public Choice. 1975;24(1):27–42.
13. Eisenhardt KM. Agency theory: An assessment and review. Academy of Management Review. 1989;14(1):57–74. <https://doi.org/10.5465/amr.1989.4279003>
14. De Walque D, Kandpal E, Wagstaff A, Friedman J, Piatti-Fünfkirchen M, Sautmann A, et al. Improving effective coverage in health: do financial incentives work? Washington, DC: World Bank Publications; 2022.
15. Davis JH, Schoorman FD, Donaldson L. Toward a stewardship theory of management. Academy of Management Review. 1997;22(1):20–47. <https://doi.org/10.5465/amr.1997.9707180258>
16. Torfing J, Bentzen TØ. Does stewardship theory provide a viable alternative to control-fixated performance management? Administrative Sciences. 2020;10(4).
17. Bank W. Honduras Poverty Assessment: Toward a Path of Poverty Reduction and Inclusive Growth. Washington, DC: The World Bank; 2024.
18. United Nations Development Programme (UNDP). Human Development Report 2023/24: Reimagining cooperation in a polarized world. New York: United Nations Development Programme; 2024.
19. Zarychta A, Wong J. Decentralization and resource contributions within local governance networks: Evidence from health sector reform in Honduras. Policy Studies Journal. 2024;52(1):113–37.
20. Carmenate-Milián L, Herrera-Ramos A, Ramos Caceres D, Lagos Ordonez K, Ordoñez TL, Valladares CS. Situation of the health system in Honduras and the new proposed health model. Archivo Medico. 2017;9(4):1–8.

21. Root ED, Zarychta A, Tapia BB, Grillos T, Andersson K, Menken J. Organizations matter in local governance: evidence from health sector decentralization in Honduras. *Health Policy and Planning*. 2020;35(9):1168–79.
22. Mokdad AH, Colson KE, Zuniga-Brenes P, Rios-Zertuche D, Palmisano EB, Alfaro-Porras E, et al. Salud Mesoamerica 2015 Initiative: design, implementation, and baseline findings. *Popul Health Metr*. 2015 Feb 7;13:1–16. <https://doi.org/10.1186/s12963-015-0034-4>
23. Reference anonymised for review.
24. Chen HT. Practical program evaluation: Assessing and improving planning, implementation, and effectiveness. 2nd ed. Los Angeles: Sage Publication; 2015. 443 p.
25. Funnell SC, Rogers PJ. Purposeful program theory: Effective use of theories of change and logic models. Vol. 31. San Francisco, CA: Jossey-Bass; 2011. 576 p.
26. Mayne J. Theory of change analysis: Building robust theories of change. *Can J Program Eval*. 2017;32(2).
27. Breuer E, Lee L, Silva M, Lund C. Using theory of change to design and evaluate public health interventions: a systematic review. *Implementation Science*. 2016;11(1).
28. Mason P, Barnes M. Constructing theories of change: Methods and sources. *Evaluation-Us*. 2007;13(2):151–70.
29. Chen HT. Theory-driven evaluation: Conceptual framework, application and advancement. In: *Evaluation von Programmen und Projekten für eine demokratische Kultur*. Wiesbaden: Springer; 2012. p. 17–40. https://doi.org/10.1007/978-3-531-19009-9_2
30. Sayer A. Realism and social science. Thousand Oaks: Sage Publications; 2000. 211 p.
31. Bhaskar R. A realist theory of science. 1st, editor. London, UK: Routledge; 2010. 240 p. (Classical Texts in Critical Realism).

32. Rogers PJ. Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation-Us*. 2008;14(1):29–48.
33. Leeuw FL. Reconstructing program theories: Methods available and problems to be solved. *Am J Eval*. 2003;24(1):5–20. <https://doi.org/10.1177/109821400302400102>
34. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*. 2000 Jan;55(1):68–78.
<https://doi.org/10.1037/0003-066x.55.1.68>
35. Knowles MS SRA Holton III EF. *The adult learner: The definitive classic in adult education and human resource development*. New York: Routledge; 2014.
36. Bandura A. Exercise of human agency through collective efficacy. *Curr Dir Psychol Sci*. 2000;9(3):75–8.
37. Lave J, Wenger E. *Situated learning: Legitimate peripheral participation*. Cambridge university press; 1991.
38. Michie S, Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*. 2011;6(1):1–12.
39. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. *Qualitative Health Research*. 2015;26(13):1753–60.
<https://doi.org/10.1177/1049732315617444>
40. Trochim WM. Outcome pattern matching and program theory. *Evaluation and program planning*. 1989;12(4):355–66.
41. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International journal for quality in health care*. 2007;19(6):349–57.

42. Leonard DK. Where are “pockets” of effective agencies likely in weak governance states and why? A propositional inventory. Brighton, UK: Institute for Development Studies; 2008. p. 33. (Working Papers).
43. Sheikh K, Gilson L, Agyepong IA, Hanson K, Ssengooba F, Bennett S. Building the field of health policy and systems research: framing the questions. PLoS Medicine. 2011;8(8):e1001073. <https://doi.org/10.1371/journal.pmed.1001073>
44. Gilson L, Grandsoult V. Health systems as human systems: reflexivity, relationships, and resilience in the pursuit of the SDGs. Frontiers in public health. 2025;13:1653839. <https://doi.org/10.3389/fpubh.2025.1653839>
45. Hood C. Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses. Public Admin Rev. 2012;72:S85-S92.
46. Vareilles G, Marchal B, Kane S, Petric T, Pictet G, Pommier J. Understanding the motivation and performance of community health volunteers involved in the delivery of health programmes in Kampala, Uganda: a realist evaluation. Bmj Open. 2015;5(11). <https://doi.org/10.1136/bmjopen-2015-008614>
47. Lohmann J, Wilhelm D, Kambala C, Brenner S, Muula AS, Allegri M. “The money can be a motivator, to me a little, but mostly PBF just helps me to do better in my job.” An exploration of the motivational mechanisms of performance-based financing for health workers in Malawi. Health Policy and Planning. 2018;33:183–91.
48. Munar W, Wahid SS, Mookherji S, Innocenti C, Curry LA. Team-and individual-level motivation in complex primary care system change: A realist evaluation of the Salud Mesoamerica Initiative in El Salvador. Gates Open Research. 2018;2.

49. Ebenso B, Mbachu C, Etiaba E, Huss R, Manzano A, Onwujekwe O. Which mechanisms explain the motivation of primary health workers? Insights from the realist evaluation of a maternal and child health programme in Nigeria. *BMJ Global Health*. 2020;5(8).
50. Wahid SS, Munar W, Das S, Gupta M, Darmstadt GL. 'Our village is dependent on us. That's why we can't leave our work'. Characterizing mechanisms of motivation to perform among Accredited Social Health Activists (ASHA) in Bihar. *Health Policy and Planning*. 2019;35(1):58–66.
<https://doi.org/10.1093/heapol/czz131>
51. Franco-Santos M, Lucianetti L, Bourne M. Contemporary performance measurement systems: A review of their consequences and a framework for research. *Management Accounting Research*. 2012;23(2):79–119.