Optimizing Audit and feedbaCk To Implement eVidence-based prAcTices in primary health carE in Nepal, Mozambique, Tanzania and China (ACTIVATE trial): rationale and design of a factorial randomized trial

Huanyuan Luo[1#], Hamada Kidanga Mussa[2,3,4#], Xiaoqing Zhu[5,2,3#], Qing Zhao[2,3], Yiyuan Cai[6], Dadong Wu[7], Jiangyun Chen[8], Xiaohui Wang[9], Changchang Li[10], Duolao Wang[11], Mengjun Zhang[1], Run (Sherry) Wang[10], Pramesh Koju[12], Sony Shakya Shrestha[13], Rajeev Shrestha[13], Biraj Man Karmacharya[13], Eusébio Eugênio Chaquisse[14,15], Rosa Marlene Cuco[14,15], Elsa Luís Kanduma[14,15], Isaias Ramiro[15], Archana Shrestha[16], Sophia Siu Chee Chan[17], Dong (Roman) Xu[8,10,5]*

*Corresponding author. No 1023, South Shatai Road, Guangzhou, 510515, China
Correspondence: romanxu@gmail.com; romanxu@i.smu.edu.cn

#Authors contribute equally to this work.

1, Southern Medical University Institute for Global Health (SIGHT), Dermatology Hospital, Southern Medical University, Guangzhou, China
2, Acacia Lab for Implementation Science, School of Health Management, Southern Medical University, Guangzhou, China
3, School of Health Services Management, Southern Medical University, Guangzhou, China
4, School of Health and Medical Sciences, the State University of Zanzibar, Zanzibar, Tanzania.
5, School of Public Health, Southern Medical University, Guangzhou, China
6, Department of Epidemiology and Health Statistics, School of Public Health, Guizhou Medical University, Gui'an, China
7, Shenzhen Maternity and Child Healthcare Hospital, Southern Medical University, Shenzhen, China
8, Acacia Lab for Implementation Science and Center for World Health Organization Studies, School of Health Management, Southern Medical University, Guangzhou, China
9, School of Public Health, Lanzhou University, Lanzhou, China
10, Shenzhen Hospital of Southern Medical University, Guangzhou, China
11, Biostatistics Unit, Liverpool School of Tropical Medicine, Liverpool, UK
12, Dhulikhel Hospital, Kathmandu University Hospital, Dhulikhel, Nepal
13, Kathmandu University School of Medical Sciences, Dhulikhel, Nepal
14, Department of Primary Health Care, the National Directorate of Public Health, the Ministry of Health of Mozambique, Maputo, Mozambique
15, Health Commission of Mozambique, Maputo, Mozambique
16, Department of Chronic Disease Epidemiology, Center of Methods for Implementation and Prevention Science, Yale School of Public Health, New Haven, USA
17, School of Public Health, The University of Hong Kong, Hong Kong, China

Abstract

Background The ability of primary health workers (PHWs) to practice in accordance with evidence-based guidelines and norms is a critical component of improving the quality of primary healthcare. Implementation science seeks to promote the routine use of evidence-based practices in healthcare by identifying barriers and facilitators to their implementation, and by developing strategies such as audit and feedback (AnF) to overcome these barriers. However, the effects of AnF show significant heterogeneity across studies, so this research focuses on the validity of AnF components and the optimal combinations of these components through head-to-head comparisons.

Methods During the preparation phase, we will conduct a preliminary exploration to optimize AnF intervention aimed at improving the quality of primary healthcare. This will involve identifying key candidate components and levels of the AnF intervention, understanding the mechanisms of action, and assessing the resource constraints associated with implementing AnF in primary healthcare (PHC) facilities. This will be achieved through expert consultation and a Best-Worst Scaling (BWS) questionnaire survey. In the optimization phase, we will use the gold standard method of Unannounced Standardized Patients (USP) to assess the quality of primary healthcare for diabetes and hypertension, focusing on the accuracy and standardization of consultation, examination, diagnosis, and treatment procedures in comparison to evidence-based practices, across Nepal, Mozambique, Tanzania, and China. A factorial design randomized controlled trial will be conducted to determine how feedback on care quality can be delivered to PHWs in order to optimize its impact on healthcare quality improvement. To achieve this, the factorial trial will incorporate four key AnF components, each at two levels, resulting in a total of sixteen intervention groups. PHWs will be randomly assigned to these groups.

Discussion This study will provide an empirical foundation for using AnF to improve the quality of primary healthcare in developing countries, while also enhancing and complementing the existing AnF theory. This study focuses on the optimization of the important implementation strategy of AnF in implementation science, contributing a universal research paradigm to the field.

Trial registration ClinicalTrials.gov, NCT06480487. Registered 27 December 2024, https://clinicaltrials.gov/study/NCT06480487.

Keywords

Audit and Feedback, Evidence-Based Practice, Primary Healthcare, Quality

Introduction

Background and rationale

Numerous studies have highlighted the need to improve the quality of primary healthcare (PHC) in low- and middle-income countries (LMICs). Since 2017, our research team has been conducting the *Primary heAlth Care quAlity Cohort In ChinA* (ACACIA) study, using Unannounced Standardized Patients (USPs) to covertly visit more than 2,000 randomly selected PHC facilities across seven provinces of China, in order to gather accurate information on the quality of services. Interim results from the ACACIA study indicate that PHC services in China perform moderately well in terms of efficiency and timeliness of service delivery. However, there are significant concerns regarding the technical quality (the efficacy and safety of diagnosis and treatment) and the quality of patient-centered services. For eleven common diseases, the correct diagnosis rate among primary health workers (PHWs) was only 27.3%, and the standardized rates for consultation, examination, and treatment were only 16.0%, 10.0%, and 23.0%, respectively[1]. These findings suggest that the quality of PHC in China is comparable to that in other LMICs[2-4].

Implementation science seeks to promote the routine use of evidence-based practices in healthcare by investigating the barriers and facilitators to their implementation, and developing strategies to overcome these barriers. Over time, implementation science has evolved to introduce new concepts aimed at advancing evidence-based practices in PHC, such as radical incrementalism (including Multi-phase Optimization Strategy, or MOST)[6], Consolidated Framework for Implementation Research (CFIR)[7], and Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM)[8]. The MOST is an innovative, systematic approach designed to assist researchers in the development, optimization, and implementation of complex interventions. Its goal is to identify the most effective combinations of intervention components to achieve optimal implementation outcomes, while considering resource constraints. In general, compared to traditional health interventions, there are currently few randomized controlled trials (RCTs) focused on implementation strategies, and even fewer conducted in LMICs.

There are five major types of supply-side implementation strategies aimed at improving healthcare quality: (1) audit and feedback (AnF), (2) training and education, (3) infrastructure improvement (e.g., updating electronic medical record systems), (4) financial incentives, and (5) clinical decision support systems (e.g., implementing physician-specific "reminder" systems)[9, 10]. Training and education for medical staff are among the most commonly used strategies to improve healthcare quality[11, 12]. However, due to the well-known "know-do" gaps in healthcare training and education, the knowledge and skills acquired are not always reflected in actual practice[13, 14]. Infrastructure improvements and financial incentives, while effective in some cases, are not always feasible[15-17]. Clinical decision support systems, including "reminder" mechanisms, have been widely studied for their efficacy but are often of limited value for replication[18]. Therefore, the AnF implementation strategy is the primary focus of this research.

Through AnF, professionals are encouraged to modify their clinical practices to align with evidence-based standards[19]. An audit involves assessing professional performance against explicit criteria or standards, such as evidence-based practices and clinical guidelines. The results of this assessment are then systematically communicated back to the professionals in a structured manner. AnF is a crucial implementation strategy for improving PHC. However, the effects of AnF show significant heterogeneity across studies. This variability may stem from the complexity of AnF components or inconsistencies in its implementation due to vague definitions. Meta-analyses have indicated that

the overall effect size of AnF as an intervention was established after the first 30 RCTs[20]. However, comparisons of different AnF component combinations primarily rely on indirect comparisons across studies. These comparisons often yield weak evidence-based conclusions due to variations in study designs, settings, intervention types, target behaviour complexities, and outcome indicators[14]. Thus, future research need to prioritize evaluating the validity of individual AnF components and determining the optimal combination of these components through direct, head-to-head comparisons.

Additionally, there is a lack of theoretical foundations guiding the establishment of components in existing AnF research. According to the literature, few AnF studies explicitly utilize theory to inform their development, implementation, and evaluation[21]. Instead, the majority of studies rely primarily on the researchers' experience or intuition.

The accuracy of the audit is the foundation of effective feedback. Chart review is the most commonly used audit method. While chart review audits are straightforward to implement, they are often imperfect in PHC settings in LMICs. Information may be omitted, over-recorded, or inaccurately documented, meaning the results of chart reviews do not necessarily reflect the true quality of PHC. In recent years, Unannounced Standardized Patients (USPs) have been increasingly adopted for assessing PHC quality[1]. USPs are highly trained individuals who can realistically and consistently simulate the symptomatic and emotional characteristics of specific conditions. By following standard healthcare service procedures without informing the healthcare providers of their role, USPs can capture a realistic and comprehensive picture of the quality of care delivered. However, to our knowledge, no studies to date have utilized USPs in the context of AnF.

Up to 140 RCTs on AnF have been conducted[22], with the majority taking place in developed countries such as the United States, Canada, the United Kingdom, Sweden, Australia, and New Zealand[23-25]. To the best of our knowledge, only a few studies have been conducted in China and LMICs[26-29] and none in Mozambique, Tanzania, and Nepal. It is necessary to develop and validate the optimal AnF interventions for LMICs in PHC settings.

Objectives

This study aims to identify the most effective combination of AnF components, taking into account resource limitations, to enhance the implementation of clinical guidelines and improve the quality of PHC in Mozambique, Zanzibar (Tanzania), Nepal, and China. Specifically, the study seeks to determine how the results of quality of care obtained from the audit process can be effectively fed back to healthcare providers in the context of PHC in LMICs, thereby optimizing improvements in care quality. Guided by the MOST strategy, we aim to accomplish the following three goals: (1) To assess the quality of the consultation process, examination, diagnosis, and treatment provided by PHC facilities in the four LMICs using USPs tool; (2) To explore the components and resource constraints (or optimization criteria, such as cost and time) for the AnF intervention in the four countries; and (3) To identify the optimal AnF intervention packages tailored to local contexts by evaluating the main and interaction effects of components, considering optimization criteria specific to PHC settings.

Methods

This protocol is reported according to the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) statement (Additional file 1).

Preparation for AnF intervention

During the preparation phase, we will conduct a preliminary exploration of how AnF for quality of care can be optimized for healthcare workers, with the goal of enhancing healthcare quality. This will involve identifying key candidate components and levels of the AnF intervention, understanding the mechanisms of action, and assessing the resource constraints associated with implementing AnF in PHC facilities. The findings from this preparation phase will serve as the foundation for the optimization phase, where further investigation will focus on refining the effectiveness of AnF for quality improvement. The details of the preparation for the AnF intervention are as follows.

AnF component ranking by expert meetings and Best-Worst Scaling (BWS)

Researchers and experts will convene to identify the most important AnF intervention components (such as the seven key components) and levels derived from the literature (Additional file 2). The research team will present the study using PowerPoint slides outlining the potential AnF components, providing experts with an overview of their tasks. These tasks will include discussing the importance, feasibility, and mechanisms of the identified components and levels, as well as ranking their importance and feasibility. The meeting will be recorded using Tencent Conference Meeting software.

Based on the most important AnF components identified during expert meetings, a Best-Worst Scaling (BWS) survey will be used to invite a broader range of stakeholders to rank the relative importance of the AnF candidate components and identify the key components. The BWS survey is a screening experiment grounded in random utility theory. By prompting participants to identify the best and worst options among a series of components, the questionnaire triggers a trade-off mechanism, allowing for the quantification of the relative importance of each component and helping to identify the most crucial one. Its efficiency in reducing the number of paired comparisons, coupled with its strong ability to maintain consistency in judgments, has made it a widely used and effective multi-criteria decision-making technique.

The Balanced Incomplete Block Design (BIBD) is an experimental design used in BWS to enhance results by organizing items into blocks and balancing the frequency of item presentations across participants. This approach allows for efficient comparison of a set of items, ensuring that each item is presented an equal number of times across blocks. The BIBD minimizes bias and improves the statistical precision of the ratings. It is particularly useful when dealing with a larger number of ranked items. For example, BIBD could be implemented in the standard manner by dividing seven AnF intervention components into seven blocks, each containing three components (Table 1). Each participant will receive a questionnaire with all seven blocks. Participants will be asked to assess which component from each block they consider most effective and which they consider least effective.

Participants in the BWS survey will include the study population of the trial, as well as administrators. While there are some guidelines, we find no universal standard exists for determining an adequate sample size for BWS surveys[33]. The sample sizes in BWS studies have ranged from 15 to 803 participants, with a median of 175[34]. Based on this, we have decided to invite 180 participants to complete the formal survey. To administer the survey, we will recruit four investigators responsible for distributing and collecting the questionnaires, either online or onsite. The Principal Investigator

(PI) and a research assistant will train the investigators on the study objectives and how to administer the questionnaire. The investigators will coordinate the survey logistics, including the time and location for the onsite survey.

Both the counting approach and logit modelling approach will be used for statistical analysis. The counting approach will be employed to calculate the best-minus-worst (B-W) scores, representing the difference between the number of times a given component is selected as "most effective" and "least effective". A positive B-W score indicates that the component is chosen as "most effective" more frequently than as "least effective". Logit models will be used to estimate the coefficients for each component, identifying the most and least effective components in promoting greater completion of guideline-recommended items and ensuring accurate diagnosis and treatment by practitioners. The key AnF components will be selected for the optimization of the AnF intervention.

Study design - factorial RCT

In the optimization phase, a 2×2×2×2 factorial design RCT will be conducted, incorporating four two-level components decided from BWS survey and consensus meeting, resulting in a total of sixteen AnF intervention groups. The PHC workers, policymakers, and health system administrators will be invited to discuss the feasibility and operational details of implementing the sixteen AnF interventions for PHC facilities. Participants will be randomly assigned to one of these sixteen AnF intervention groups. The audited quality of care results will be fed back to the healthcare workers, and outcome indicators reflecting the effectiveness of AnF will be measured. We will estimate the main and interaction effects of AnF components on improving PHC quality via statistical analysis. The optimal AnF combination will be determined by evaluating both the effects and resource constraints in local implementation settings. We assume that the four AnF components in Table 2, each with two levels, will be selected from the BWS survey. The intervention setup is presented in Table 2.

Study sites

The study sites will include Maputo City and Maputo Province in Mozambique; the North, South, and Urban West regions in Zanzibar, Tanzania; the Kathmandu, Bhaktapur, and Lalitpur districts in Nepal; and Henan Province, Guangdong Province, and other regions in China.

Study population

Inclusion and exclusion criteria

In Mozambique, the inclusion criteria are nurses, general medicine technicians, and physicians in PHC facilities; and the exclusion criteria are healthcare providers from other departments, such as tuberculosis, malaria, human immunodeficiency virus/acquired immune deficiency syndrome, or emergency services, as well as students currently on internship. In Zanzibar, Tanzania, the inclusion criteria are clinical officers, nurses, and other healthcare providers involved in consultation, examination, diagnosis, and treatment in PHC facilities. In Nepal, the inclusion criteria are medical officers, health assistants, and senior auxiliary health workers in PHC facilities; and the exclusion criteria are interns or students present during the visit. In China, the inclusion criteria of PHC facilities are primary and secondary hospitals, community health centers (stations) and clinics, as well as township health centers and village health clinics; and the inclusion criteria of research participants

244 are practicing (assistant) physicians and rural physicians working in healthcare facilities that meet the
245 above criteria, with a scope of practice only including general practice and internal medicine.

246 **Sample size**

247 The sample of the 2×2×2×2 factorial design study is calculated based on the primary outcome: the
248 completion rate and adherence to key consultation items in guidelines by workers at PHC facilities
249 (measured as a continuous variable). As stated in the introduction, the ACACIA project, conducted by
250 our research group in China, reported a mean completion rate of key consultation items of 16.0%
251 (standard deviation = 12.7%) from interim data, which is comparable to that observed in other LMICs.
252 Preliminary expert consensus suggests that an increase of approximately 5 percentage points in the
253 completion rate (effect size: unstandardized mean difference) is considered clinically significant. The
254 parameter for the standard deviation of the response variable after treatment, within each
255 treatment condition (i.e., adjusting for treatment but not for post-test), is not available in the
256 literature. Therefore, we use the available data from the ACACIA project. We consider a significance
257 level of 0.05 (two-sides) and a power of 0.9. Using the "FactorialPowerPlan()" command in the MOST
258 package of R software (version 4.4.2), the total sample size required is 274. To ensure equal numbers
259 in each intervention group, we plan to recruit 18 participants per group. Considering potential
260 subject attrition and other factors that may result in a dropout rate of approximately 5%, the
261 adjusted sample size for each group is 19 participants, resulting in a total sample size of 304.

262 **Variables, data collection instrument and outcome**

263 Variables collected at baseline include facility characteristics (i.e., country, diseases managed by USP
264 [diabetes or hypertension]), and PHC workers characteristics (i.e., age, gender, education level, years
265 of work experience).

266 The study will focus on Type II diabetes and hypertension, as these are prevalent conditions in
267 primary care. Outcome indicators will be systematically collected following the RE-AIM framework
268 and Proctor's Implementation Outcomes Framework[8]. Among these indicators, the quality of care
269 indicators will be assessed using the USP methodology and subsequently provided as feedback to
270 PHC workers. The quality of care indicators will align with the quality framework developed by the
271 Institute of Medicine (IOM)[30], which consists of six domains: effectiveness, efficiency, equity, patient-
272 centeredness, safety, and timeliness.

273 - **Effectiveness** (avoiding underuse and misuse) and **safety** (avoiding harm) represent the
274 traditional technical goals of quality care. These encompass aspects such as consultation,
275 examination, diagnosis, and treatment, and will include both the primary and some
276 secondary outcomes.
277 - **Patient-centeredness** (respecting and accommodating individual preferences) will be
278 evaluated using the *Patient Perception of Patient-Centeredness Rating Scale* (PPPC)[31]. The
279 PPPC employs a 4-point Likert scale to measure attitudes across three dimensions: exploring
280 the illness and the experience of being ill, understanding the whole person, and finding
281 common ground.
282 - **Timeliness** will be assessed through measures such as facility opening hours, waiting times,
283 and consultation durations.

Efficiency (avoiding waste) can be evaluated by analyzing the costs incurred for care provided to the USPs during their interactions with PHC workers. Equity (ensuring no disparities in quality due to individual characteristics) cannot be fully evaluated through a single USP visit and will require a separate study. Due to funding and time limitations, both the efficiency and equity will not be assessed in this study.

The primary outcome is the proportion of completed guideline-recommended quality checklist items for consultation of hypertension and Type II diabetes cases of the PHC providers among all of the items. This outcome will be expressed as a continuous score ranging from 0% to 100%.

Secondary outcomes include:

- Quality of care indicators: The proportion of completed guideline-recommended quality checklist items for physical and laboratory exams of hypertension and Type II diabetes cases of the PHC providers among all of the items. This is a continuous score ranging from 0 to 100%.
- Quality of care indicators: Correctness of diagnosis of hypertension and type II diabetes cases by PHC providers. This will be categorized into 2 categories: correct and incorrect diagnosis.
- Quality of care indicators: Correctness of treatment of hypertension and type II diabetes cases by PHC providers. This will be categorized into 2 categories: correct and incorrect treatment.
- Quality of care indicators: Timeliness of hypertension and type II diabetes services in PHC settings.
- Quality of care indicators: Patient-centered quality of healthcare in PHC settings.
- Implementation outcome: Adoption of AnF intervention by study participants. It is a dichotomous variable, consisting of two categories: adopted, not adopted. It will be self-reported by participants using team-developed questionnaire.
- Implementation outcome: Costs to researchers of developing and implementing AnF intervention. It is a continuous variable and will be assessed using project final account of expenditure.
- Implementation outcome: Participants score of acceptability of AnF intervention. It will be self-reported by participants using Generic Theoretical Framework of Acceptability (TFA)-based questionnaire.

Procedure of randomization and data collection

Step 1: Informed consent/basic information/audit. Study participants will be invited to complete the informed consent form and basic information questionnaire on line or offline. USPs seek medical care without disclosing their identity as USPs, following standard procedures to gather information on quality of care indicators. All data will be entered into REDCap[32], a free and robust cloud-based tool for data collection, storage, and management. The auditing content (i.e., quality of care indicators) will be collected using the *Quality of Care Assessment Form* for later feedback. This data will be documented either by the USPs themselves or by investigators after conducting interviews with the USPs completing their visit to the participants.

Step 2: Randomization and intervention. Following the completion of the basic information questionnaire by all study participants, randomization will be performed. The randomization lists will

be generated using block randomization with a block size of sixteen. Eligible participants will be randomly assigned to one of the sixteen AnF intervention groups, with stratification by country. The randomization process will be carried out by two statisticians who are independent of the trial. They will perform the assignment sequentially to ensure accuracy and minimize bias. Given the nature of the intervention content (e.g., the intervening physicians would receive feedback reports from the program coordinator), it is not possible to blind the study coordinator or participating physicians to the intervention assignments. However, outcome assessors and data analysts are not aware of these assignments.

Step 3: Collection of outcome indicators after intervention completion. For the outcomes related to quality of care indicators, we will use USP tools for data collection, following a method similar to the baseline audit. For implementation outcomes, including the adoption of the AnF intervention by study participants and participants' acceptability scores of the AnF intervention, we will conduct a survey to gather the data.

Data analysis and management

Statistical significance will be assessed using a two-sided alpha of 0.05. All statistical analyses will be performed using SAS software version 9.4. The participant baseline characteristics, and primary and secondary outcomes, will be shown as means (standard deviation) or number (proportions) as appropriate. We will conduct the intention-to-treat (ITT) analysis, i.e. the analysis will be conducted based on the originally assigned groups. We will use generalized linear models (GLMs) to evaluate the effectiveness of the intervention. In this study, the primary outcome is the proportion of completed guideline-recommended quality checklist items for consultation of hypertension and Type II diabetes cases of the PHC providers among all of the items, which is a continuous score. For the primary outcome, we will employ a GLM with a normal distribution and a identity link function, from which mean differences and its 95% confidence interval (95% CI) will be derived.

For sensitivity analysis, the missing values of the primary outcome will be imputed using multiple imputation method to verify the robustness of the primary analysis. We will also conduct subgroup analyses, using GLMs to examine the intervention's effectiveness on primary outcome across various characteristics, including country, diseases managed by USP [diabetes or hypertension], and PHC workers characteristics (i.e., age, gender, education level, years of work experience). The economic evaluation of the sixteen interventions will be conducted using the incremental analysis method of cost-effectiveness analysis.

For continuous secondary outcomes, the same statistical analysis methods for analysis of the primary outcome will be applied. For dichotomous secondary outcomes, we will use GLMs with a binomial distribution and an identity link function, from which risk differences and its 95% CI will be derived.

Discussion

This study will provide an empirical foundation for using AnF to improve the quality of primary healthcare in developing countries, while also enhancing and complementing the existing AnF theory. This study focuses on the optimization of the important implementation strategy of AnF in implementation science, contributing a universal research paradigm to the field. Unlike previous studies that explore entirely new interventions, or those that seek various solutions to barriers in implementation research, this study applies theories and frameworks of implementation science to

optimize the AnF implementation strategy. The expected outcome is that this research paradigm will be applicable to the optimization of implementation strategies with similar research problems in the future, enhancing the intervention effects of these strategies, and thus contributing to the development of research paradigms in implementation science. This, in turn, represents both the challenge and difficulty of this study. Specifically:

First, our study contributes to the existing literature by examining [research topic] across four countries, all classified as LMICs, with some categorized as low-income. While most previous research in this field has been conducted in high-income settings, studies from LMICs remain limited. By addressing this gap, our findings will provide valuable insights into the implementation challenges and strategies within resource-constrained environments. Moreover, the cross-country analysis enhances the generalizability of our results, providing a comparative perspective that highlights both common and context-specific factors influencing AnF effectiveness. These insights are particularly relevant for informing policy and practice in similar settings globally.

Second, regarding the concept of intervention development, the study follows the guidance of the radical incremental approach. The radical incremental approach is a commonly used concept in intervention development in implementation science, emphasizing the gradual and stepwise improvement of intervention measures during the development process to achieve a "radical" enhancement of intervention effectiveness. For complex intervention packages, which consist of multiple components aimed at facilitating implementation (e.g., AnF strategy), combining the most effective components can yield the most efficient overall intervention package. This approach reduces the risk of developing ineffective interventions compared to the development of entirely new interventions.

Third, in terms of the development process, the study adopts the principle of "co-production". By closely collaborating with stakeholders (using the Delphi expert consultation method and BWS survey), the study jointly develops and tests the AnF intervention. This approach promotes communication and improves the development efficiency, effectiveness, feasibility, and sustainability of the intervention.

Fourth, in terms of research focus, the study centers on the components and combinations of intervention components. Previous studies on the AnF effectiveness have shown considerable variation, potentially due to the differences in the components and combinations of interventions under the general term of AnF. There has been a lack of investigation into whether the AnF components and their combinations are effective or exhibit synergistic effects. Therefore, this study focuses on identifying which components of the AnF implementation strategy are effective in improving the quality of primary healthcare of diabetes and hypertension (primarily concerning the accuracy and standardization of procedures for consultation, examination, diagnosis, and treatment), as well as the effects of these components when combined.

Finally, in terms of research design, a Multi-phase Optimization Strategy is employed. This research framework is used for the development, optimization, and evaluation of complex intervention packages. In contrast to previous studies that used parallel designs, comparing multiple groups with different feedback methods or comparing them with a control group, the advantage of the multi-phase optimization strategy is as follows: 1) It uses an efficient factorial design (with a smaller sample size to validate more components) to explore whether each component of the intervention is effective and to examine potential synergistic effects between components, such as whether the

411 combined effects strengthen or weaken the overall impact; 2) It emphasizes "optimization under
412 constraints", taking into account local resource limitations to achieve the most effective AnF
413 implementation strategy.

414

415 **Reference**

416 1. Xu, D. R.; Cai, Y.; Wang, X.; Chen, Y.; Gong, W.; Liao, J.; Zhou, J.; Zhou, Z.; Zhang, N.;
417 Tang, C.; Mi, B.; Lu, Y.; Wang, R.; Zhao, Q.; He, W.; Liang, H.; Li, J.; Pan, J., Improving Data
418 Surveillance Resilience Beyond COVID-19: Experiences of Primary heAlth Care quAlity Cohort In
419 ChinA (ACACIA) Using Unannounced Standardized Patients. *Am J Public Health* **2022,** *112* (6), 913-
420 922.
421 2. Daniels, B.; Kwan, A.; Satyanarayana, S.; Subbaraman, R.; Das, R. K.; Das, V.; Das, J.; Pai,
422 M., Use of standardised patients to assess gender differences in quality of tuberculosis care in urban
423 India: a two-city, cross-sectional study. *Lancet Glob Health* **2019,** *7* (5), e633-e643.
424 3. Daniels, B.; Kwan, A.; Pai, M.; Das, J., Lessons on the quality of tuberculosis diagnosis from
425 standardized patients in China, India, Kenya, and South Africa. *J Clin Tuberc Other Mycobact Dis* **2019,**
426 *16*, 100109.
427 4. Kwan, A.; Daniels, B.; Saria, V.; Satyanarayana, S.; Subbaraman, R.; McDowell, A.;
428 Bergkvist, S.; Das, R. K.; Das, V.; Das, J.; Pai, M., Variations in the quality of tuberculosis care in
429 urban India: A cross-sectional, standardized patient study in two cities. *PLoS Med* **2018,** *15* (9),
430 e1002653.
431 5. Halpern, D.; Mason, D., Radical incrementalism. *Evaluation* **2015,** *21*, 143-9.
432 6. Guastaferro, K.; Collins, L. M., Achieving the Goals of Translational Science in Public Health
433 Intervention Research: The Multiphase Optimization Strategy (MOST). *American Journal of Public*
434 *Health* **2019,** *109* (S2), S128-S129.
435 7. Damschroder, L. J.; Reardon, C. M.; Widerquist, M. A. O.; Lowery, J., The updated
436 Consolidated Framework for Implementation Research based on user feedback. *Implement Sci* **2022,**
437 *17* (1), 75.
438 8. Reilly, K. L.; Kennedy, S.; Porter, G.; Estabrooks, P., Comparing, Contrasting, and Integrating
439 Dissemination and Implementation Outcomes Included in the RE-AIM and Implementation
440 Outcomes Frameworks. *Front Public Health* **2020,** *8*, 430.
441 9. McHugh, S.; Presseau, J.; Luecking, C. T.; Powell, B. J., Examining the complementarity
442 between the ERIC compilation of implementation strategies and the behaviour change technique
443 taxonomy: a qualitative analysis. *Implement Sci* **2022,** *17* (1), 56.
444 10. Waltz, T. J.; Powell, B. J.; Matthieu, M. M.; Damschroder, L. J.; Chinman, M. J.; Smith, J. L.;
445 Proctor, E. K.; Kirchner, J. E., Use of concept mapping to characterize relationships among
446 implementation strategies and assess their feasibility and importance: results from the Expert
447 Recommendations for Implementing Change (ERIC) study. *Implement Sci* **2015,** *10*, 109.
448 11. Fischer, F.; Lange, K.; Klose, K.; Greiner, W.; Kraemer, A., Barriers and Strategies in
449 Guideline Implementation-A Scoping Review. *Healthcare (Basel)* **2016,** *4* (3).
450 12. Spoon, D.; Rietbergen, T.; Huis, A.; Heinen, M.; van Dijk, M.; van Bodegom-Vos, L.; Ista, E.,
451 Implementation strategies used to implement nursing guidelines in daily practice: A systematic
452 review. *Int J Nurs Stud* **2020,** *111*, 103748.
453 13. Soong, C.; Shojania, K. G., Education as a low-value improvement intervention: often
454 necessary but rarely sufficient. *BMJ Qual Saf* **2020,** *29* (5), 353-357.

14. Jamtvedt, G.; Young, J. M.; Kristoffersen, D. T.; O'Brien, M. A.; Oxman, A. D., Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *Qual Saf Health Care* **2006,** *15* (6), 433-6.
15. Dopp, A. R.; Kerns, S. E. U.; Panattoni, L.; Ringel, J. S.; Eisenberg, D.; Powell, B. J.; Low, R.; Raghavan, R., Translating economic evaluations into financing strategies for implementing evidence-based practices. *Implement Sci* **2021,** *16* (1), 66.
16. Dopp, A. R.; Narcisse, M. R.; Mundey, P.; Silovsky, J. F.; Smith, A. B.; Mandell, D.; Funderburk, B. W.; Powell, B. J.; Schmidt, S.; Edwards, D.; Luke, D.; Mendel, P., A scoping review of strategies for financing the implementation of evidence-based practices in behavioral health systems: State of the literature and future directions. *Implement Res Pract* **2020,** *1*, 2633489520939980.
17. Vandergrift, J. L.; Gray, B. M., Physician clinical knowledge, practice infrastructure, and quality of care. *Am J Manag Care* **2019,** *25* (10), 497-503.
18. Arditi, C.; Rège-Walther, M.; Durieux, P.; Burnand, B., Computer-generated reminders delivered on paper to healthcare professionals: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* **2017,** *7* (7), Cd001175.
19. G, J.; S, F.; N, I., Audit and Feedback as a Quality Strategy. In: Busse R, Klazinga N, Panteli D, et al., editors. Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies [Internet]. Copenhagen (Denmark): European Observatory on Health Systems and Policies; 2019. (Health Policy Series, No. 53.) 10. Available from: https://www.ncbi.nlm.nih.gov/books/NBK549284/.
20. Ivers, N. M.; Grimshaw, J. M.; Jamtvedt, G.; Flottorp, S.; O'Brien, M. A.; French, S. D.; Young, J.; Odgaard-Jensen, J., Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. *J Gen Intern Med* **2014,** *29* (11), 1534-41.
21. Colquhoun, H. L.; Brehaut, J. C.; Sales, A.; Ivers, N.; Grimshaw, J.; Michie, S.; Carroll, K.; Chalifoux, M.; Eva, K. W., A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implement Sci* **2013,** *8*, 66.
22. Ivers, N.; Jamtvedt, G.; Flottorp, S.; Young, J. M.; Odgaard-Jensen, J.; French, S. D.; O'Brien, M. A.; Johansen, M.; Grimshaw, J.; Oxman, A. D., Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* **2012,** (6), Cd000259.
23. Landis-Lewis, Z.; Kononowech, J.; Scott, W. J.; Hogikyan, R. V.; Carpenter, J. G.; Periyakoil, V. S.; Miller, S. C.; Levy, C.; Ersek, M.; Sales, A., Designing clinical practice feedback reports: three steps illustrated in Veterans Health Affairs long-term care facilities and programs. *Implement Sci* **2020,** *15* (1), 7.
24. Wright-Hughes, A.; Willis, T. A.; Wilson, S.; Weller, A.; Lorencatto, F.; Althaf, M.; Seymour, V.; Farrin, A. J.; Francis, J.; Brehaut, J.; Ivers, N.; Alderson, S. L.; Brown, B. C.; Feltbower, R. G.; Gale, C. P.; Stanworth, S. J.; Hartley, S.; Colquhoun, H.; Presseau, J.; Walwyn, R.; Foy, R., A randomised fractional factorial screening experiment to predict effective features of audit and feedback. *Implement Sci* **2022,** *17* (1), 34.
25. Quigley, M.; Earnest, A.; Andrikopoulos, S.; Wischer, N.; Green, S.; Zoungas, S., ANDA-Evaluating Facilitated Feedback Enhancement - a Cluster randomised Trial (ANDA-EFFECT): protocol for a cluster randomised trial of audit feedback augmented with education and support, compared to feedback alone, on acceptability, utility and health outcomes in diabetes centres in Australia. *Trials* **2022,** *23* (1), 976.
26. Gong, S.; Qiu, X.; Song, Y.; Sun, X.; He, Y.; Chen, Y.; Li, M.; Luo, R.; He, L.; Wei, Q.; Shen, S.; Liu, Y.; Zhang, L.; Zhou, W.; Huang, P.; Mai, J.; Liu, L.; Xu, Y.; Liang, H.; Xia, H., Effect of Financially Punished Audit and Feedback in a Pediatric Setting in China, within an Antimicrobial

Stewardship Program, and as Part of an International Accreditation Process. *Front Public Health* **2016,** *4*, 99.

27. Shen, X.; Lu, M.; Feng, R.; Cheng, J.; Chai, J.; Xie, M.; Dong, X.; Jiang, T.; Wang, D., Web-Based Just-in-Time Information and Feedback on Antibiotic Use for Village Doctors in Rural Anhui, China: Randomized Controlled Trial. *J Med Internet Res* **2018,** *20* (2), e53.

28. Wu, Y.; Li, S.; Patel, A.; Li, X.; Du, X.; Wu, T.; Zhao, Y.; Feng, L.; Billot, L.; Peterson, E. D.; Woodward, M.; Kong, L.; Huo, Y.; Hu, D.; Chalkidou, K.; Gao, R., Effect of a Quality of Care Improvement Initiative in Patients With Acute Coronary Syndrome in Resource-Constrained Hospitals in China: A Randomized Clinical Trial. *JAMA Cardiol* **2019,** *4* (5), 418-427.

29. Zhen, L.; Jin, C.; Xu, H. N., The impact of prescriptions audit and feedback for antibiotic use in rural clinics: interrupted time series with segmented regression analysis. *BMC Health Serv Res* **2018,** *18* (1), 777.

30. Pongsupap, Y.; Van Lerberghe, W., Choosing between public and private or between hospital and primary care: responsiveness, patient-centredness and prescribing patterns in outpatient consultations in Bangkok. *Trop Med Int Health* **2006,** *11* (1), 81-9.

31. Ryan, B. L.; Brown, J. B.; Tremblay, P. F.; Stewart, M., Measuring Patients' Perceptions of Health Care Encounters: Examining the Factor Structure of the Revised Patient Perception of Patient-Centeredness (PPPC-R) Questionnaire. *J Patient Cent Res Rev* **2019,** *6* (3), 192-202.

32. Harvey, L. A., REDCap: web-based software for all types of data storage and collection. *Spinal Cord* **2018,** *56* (7), 625.

33. Louviere J, F. T., Marley A, *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge, United Kingdom: Cambridge University Press: 2015.

34. Cheung, K. L.; Wijnen, B. F.; Hollin, I. L.; Janssen, E. M.; Bridges, J. F.; Evers, S. M.; Hiligsmann, M., Using Best-Worst Scaling to Investigate Preferences in Health Care. *Pharmacoeconomics* **2016,** *34* (12), 1195-1209.

Abbreviations

PHWs	Primary health workers
AnF	Audit and feedback
PHC	Primary healthcare
BWS	Best-Worst Scaling
USP	Unannounced Standardized Patients
LMICs	Low- and middle-income countries
ACACIA	*Primary heAlth Care quAlity Cohort In ChinA*
MOST	Multi-phase Optimization Strategy
CFIR	Consolidated Framework for Implementation Research
RE-AI	Reach, Effectiveness, Adoption, Implementation, and Maintenance

RCTs	Randomized controlled trials
BIBD	Balanced Incomplete Block Design
ITT	Intention-to-treat
GLMs	Generalized linear models
CI	Confidence interval

Authors' contributions

HL, HKM, XZ, and DX designed the study. HL and DX secured funding. HL drafted the manuscript. QZ, YC, SSS, RS, BMK, EEC, RMC, ELK, and IR contributed to the design of the intervention and the development of the implementation approach. DW, JC, XW, CL, DL, MZ, RW, PK, AS, and SSCC contributed to the development of the study methodologies and protocols. The authors reviewed and approved the final manuscript.

Funding

This study is funded by the National Natural Science Foundation of China (72404118), China Postdoctoral Science Foundation (2023M731536), Guangdong Medical Science and Technology Research Foundation (A2023133), and Swiss Agency for Development and Cooperation (81067392).

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

This study obtained ethics approval under a protocol approved by the relevant authorities in China, Nepal, Mozambique and Zanzibar, Tanzania. All participants will provide written informed consent.

Competing interests

The authors declare no competing interests.

Table 1: BWS questionnaire example

In order to improve the quality of care, we plan to organize an intervention to measure and evaluate key health quality indicators for practitioners (completion rate of key consultation and examination items, diagnosis and treatment, etc specified in medical guidelines) and then provide some form of feedback to practitioners on the results of the evaluation, i.e. an "audit and feedback" intervention. **Which** of the following components of the intervention do you think would be the **most or least effective** in promoting greater completion of guideline items and correct diagnosis and treatment by practitioners **(please tick the appropriate component)**?		
Components	Most effective	Least effective
Both verbal and written feedback	o	o
Feedback reports incorporate comparisons with peer data	o	o
Feedback reports provide practitioners with suggested actions to overcome barriers (Action implementation toolkit: e.g. customized training manuals)	o	o

585

586 Table 2: Intervention groups

Group	Component 1 (Delivery Method of Feedback)		Component 2 (Source of Feedback)		Component 3 (Feedback with Peer Comparison)		Component 4 (Feedback with Goals)	
	Level 1	Level 2	Level 1	Level 2	Level 1	Level 2	Level 1	Level 2
	Provided face to face	Provided by electronic mail	Provided by researchers	Provided by authoritative bodies	Provided with peer comparison	Provided without peer comparison	Provided with goals	Provided without goals
1	√		√		√		√	
2		√	√		√		√	
3	√			√	√		√	
4	√		√			√	√	
5	√		√		√			√
6		√		√	√		√	
7								
8	√			√		√	√	
9	√		√			√		√
10		√	√			√	√	
11		√	√		√			√
12	√			√		√	√	
13	√			√	√			√
14	√		√			√		√
15		√		√		√	√	
16	√			√		√		√

587