Fidelity of Implementation of Prospera Digital: Evaluation of a Multi-Site *mHealth* Intervention Aimed at Improving Maternal Health Outcomes in Mexico

Pablo Gaitán-Rossi ID,[1] Selene De la Cerda Lobato,[1] Alma Cecilia Pérez Navarro,[1] Arturo Aguilar Esteva,[2] Manett Rocío Vargas García,[2] and Mireya Vilar-Compte[1]

[1]EQUIDE, Iberoamericana University, Mexico City, Mexico and [2]Department of Economics, Mexican Autonomous Technology Institute (ITAM), Mexico City, Mexico

ABSTRACT

Background: Infrastructure and human capital limitations motivate the design of *mHealth* programs, but their large-scale implementation may be challenging in a development context. Prospera Digital (PD) is a pilot *mHealth* intervention aiming to improve maternal and child health and nutrition designed as a randomized controlled trial with 3 treatment arms. It was implemented during 2015–2017 in 326 treatment clinics located in 5 states in Mexico.

Objective: Assess, with an external evaluation, PD's fidelity of implementation using 6 dimensions: adherence, quality, responsiveness, intervention complexity, facilitation strategies, and program differentiation.

Methods: Benchmark for implementation was first established by interviewing PD's developers. Extensive fieldwork in the 5 states was then conducted to assess its fidelity in heterogeneous contexts. The evaluation team visited 39 health clinics to assess the initial sign-up events and conduct a follow-up. Overall, the team made 28 closed observations; conducted 17 focus groups; and interviewed 74 health providers, 10 community leaders, and 92 beneficiaries. Field notes from the implementation team on all clinics were also examined.

Results: Co-ordination between the Health and Social Development ministries was adequate, although some health providers were not informed about PD. Program developers added useful implementation strategies during roll-out to reinforce sign-up events. Key quality facilitators were the clarity and relevance of the messages from the short messages service. Beneficiaries expressed high satisfaction with PD. In contrast, implementation barriers to adherence in some localities might reduce the potential impact of PD. Program differentiation was low between the 3 treatment arms.

Conclusions: PD is a promising strategy to contribute to the promotion of early childhood development in Mexico. Implementation science evaluation can help improve the quality of large-scale *mHealth* interventions by anticipating barriers and providing insights on how to increase performance. This is especially relevant to inform impact evaluation in development contexts. The trial was registered at the American Economic Association's registry for randomized controlled trials with trial registry number 'AEARCTR-0001035'. *Curr Dev Nutr* 2019;3:nzz107.

Introduction

Early life is a key stage of an individual's development and its influence on later outcomes has been documented as part of the research surrounding the fetal origins hypothesis (1, 2). In a context in which high inequality is a primary concern, indicators surrounding access to health early in life (including fetal development) become the first signs of perpetuating disparities. Mexico, for instance, faces important challenges with a rate of maternal mortality of 34.6 deaths per 100 000 live births in 2015, whereas the average in the Organisation for Economic Co-operation

and Development (OECD) is 7 (3). (The coverage of antenatal care during the first trimester in Mexico was 84% among pregnant women aged 20–49 y but 75% for women <19 y (4). According to the 2015 UNICEF Multiple Indicator Cluster Survey (MICS), 10.5% of the infants in Mexico have low weight [i.e. the percentage of children born alive in the previous 2 y who weighed <2500 g at birth]; only 31% receive exclusive breastfeeding before the first 6 mo; 34% get full vaccine immunization [i.e. percentage of children between 24 and 35 mo who received all the vaccines included in Mexico's National Health Plan], and 82% of children presented adequate child development [i.e. the percentage of children aged 36–59 mo who have an adequate development score in 3 out of 4 dimensions on a scale of infant development] (5)). Wide differences in maternal and child health outcomes based on socioeconomic level need to be addressed in order to improve health, human capital, and well-being across the life course (6).

Early childhood interventions target a sensitive period where benefits in development, growth, nutrition, and reductions in morbidity and mortality may be amplified (7). Nonetheless, poor health service provision and limited resources can be important limitations. In such a context, *mHealth* interventions – understood as the use of mobile devices for public health practice – focusing on this period could be effective tools by increasing access, distribution, provision, and surveillance of health services, especially in areas with an insufficient health infrastructure (8). *mHealth* is attractive because relatively small and low-cost interventions can reach large and remote populations (8). Systematic reviews have found positive outcomes on *mHealth* interventions (previous work has found they are an effective tool that empowers pregnant women and healthcare providers (9). They increase antenatal service utilization, skilled attendance at birth, and vaccination rates (10); and have a modest impact on exclusive breastfeeding [11]); however, persistent barriers for their implementation can reduce their potential effects and limit their efficient scale-up (some of these barriers are the cost of mobile plans, the lack of network infrastructure, low digital literacy, nonculturally relevant content, and lack of knowledge on how to implement). Moreover, large-scale implementation of *mHealth* innovations has been limited by a shortage of empirical evidence, (although 87% of the countries in WHO are conducting ≥1 *mHealth* program, only 14% of them reported an evaluation [8]), supporting their value in terms of cost, performance, and health outcomes, as well as by the lack of process evaluations identifying facilitators and bottlenecks to deliver the *mHealth* intervention effectively (12).

Therefore, fidelity of implementation (FOI) becomes a key input for adequate evaluation of evidence-based policies and for replication or scale-up of valuable *mHealth* interventions (13). FOI examines the degree to which an intervention was delivered as prescribed in the original protocol or as it was intended by the program developers (14). FOI allows determining if a lack of impact is due to the program itself or because of poor implementation (15, 16). Research on the FOI explains how and why an intervention is successful by opening the black box of implementation and linking program components with program outcomes (17). Lastly, FOI eases replication, helps explain why some studies generate different results, and provides guidelines for implementation and scale-up (15). Consequently, these evaluations are useful to optimize the program in specific settings, to expand the sustainability of the intervention, and to promote its dissemination to other contexts (18).

This article documents the FOI of a multi-site pilot test of an *mHealth* intervention in Mexico aiming to reduce maternal mortality and improve early childhood nutrition and development. The program targets a group of vulnerable mothers and children who are beneficiaries of the nationwide Social Inclusion Program, PROSPERA, a large conditional cash-transfer program (formerly named Oportunidades, PROSPERA was launched in 1997 aiming to strengthen the social rights and developmental capacities of low-income families in order to break the intergenerational poverty cycle by addressing 3 key components: education, health, and nutrition. In 2017, nearly 7.2 million families were affiliated, including 1,255,869 children aged <5 y). Harnessing the available social infrastructure, Prospera Digital (PD) was launched in 2015 as a pilot study embedded into PROSPERA's health component in a subset of 655 clinics in 5 states. The intervention required a co-ordinated collaboration between 2 state ministries: *1*) PROSPERA delegations – part of the State Ministry for Social Development – provided the trainers, delivered the materials, and used the conditional structure of the program to motivate the participation of pregnant women, and *2*) the State Ministry of Health co-ordinated the health clinics (Mexico has a fragmented health system – according to employment status [19] – with different health outcomes and, therefore, beneficiaries of PROSPERA receive medical services from 2 systems: the State Health Services [Seguro Popular] and the Mexican Social Security Institute [IMSS]-PROSPERA. Both types of clinics were considered for PD) and recruited the target population.

PD operates by inviting and recruiting pregnant PROSPERA beneficiaries through primary clinics. Women interested in participating are trained in the use of the *mHealth* platform, and the short messages service (SMS) text messages are sent during pregnancy and the first 2 y of life of their infants. SMS messages are free and contain health advice on maternal health and childcare, nutrition and breastfeeding, warning signs and how to respond, and reminders of medical appointments and vaccinations (see **Supplementary Material** for an explanation of their design, the target behavior, and examples). Messages are personalized and tailored to each participant based on self-reported information and clinical history. Beneficiaries received 2 types of messages: 1-way messages with general information and 2-way messages, requiring a response before sending a follow-up message. Moreover, beneficiaries can prompt actions like reporting obstetric emergencies and changing appointment dates.

SMS text messages, training materials (i.e. manuals, posters, and visual guides), the implementation, and the qualitative strategy were tested through a prepilot formative study conducted in September 2015 in the State of Mexico. This led to 561 pretested message flows intending to modify health behaviors and were delivered through the RapidPro platform (20).

A randomized controlled trial (RCT) with 3 treatment groups and 1 control group was designed to measure the impact of the pilot *mHealth* intervention. Treatment group 1 (T1) exclusively delivered the SMS messages described above. In treatment group 2 (T2), participants received the same SMS messages as in T1 and community leaders signed a voluntary 'contract' with beneficiaries to commit to 1 of 3 possibilities (chosen at will by the beneficiary): *1*) taking folic acid pills, *2*) attending all the prenatal check-ups, and *3*) replying to all PD 2-way messages. Community leaders also sent predefined messages that reinforced health information to the beneficiaries with whom

TABLE 1   Comparison of the 5 states included in the pilot of Prospera Digital by sociodemographic variables

	Guanajuato	State of Mexico	Hidalgo	Puebla	Chiapas
Multidimensional poverty[1]	46.6%	49.6%	54.3%	64.5%	76.2%
Extreme multidimensional poverty[1]	5.5%	7.2%	12.3%	16.2%	31.8%
Indigenous population[2]	9.1%	17.0%	36.2%	35.3%	36.1%
Affiliation to health services	85.0%	78.7%	82.4%	79.8%	82.3%
Urbanization[2]	61.4%	69.2%	28.8%	45.6%	33.9%
Total population (in millions)[2]	5.8	16.1	2.8	6.1	5.2
Beneficiaries of PROSPERA[3]	4.2%	3.6%	7.8%	8.0%	13.5%

[1]Note: The source is Mexico's official poverty measure in 2014 – the most recent before program roll-out – as reported by the National Council of Evaluation (CONEVAL).
[2]The source is the 2015 intermediate census survey from Mexico's National Institute of Statistics and Geography (INEGI). The indigenous population was measured as self-identification to an indigenous community. Urbanization was measured as the percentage of population living in localities with 15,000 inhabitants or more.
[3]The source of the data is PROSPERA's beneficiary register from the period July–August 2015.

they signed the 'contract' (examples of these messages are in the Supplementary Material). Treatment group 3 (T3) used the same SMS messages as in T1 and allowed beneficiaries to evaluate anonymously, through their mobile phone, the services and medical care received; health providers did not have to do additional activities for PD. Services and medical care evaluation resulted from questions sent to participants about mandatory medical protocols that should have been followed during their routine prenatal visits (e.g. weight, height, and blood pressure measurement and a test for anemia) and perceived quality of the service (i.e. waiting time, receiving a response to their questions and concerns, and general satisfaction). Health providers in the best performing clinics would be awarded additional vacation days as an incentive.

Randomization was done at the clinic level. The inclusion criteria for the clinics in the study was: being primary health centers, being in a locality with connectivity to the mobile network, ≥5 pregnant PROSPERA beneficiaries in the clinic, and ≥80% of the population reporting Spanish as their first language. For the pilot study, 655 clinics in 5 states (Chiapas, Guanajuato, Hidalgo, Puebla, and the State of Mexico), stratified by urban and rural status, were randomly assigned. Most of these states are located in the central region of Mexico but they were chosen because they vary in population, levels of poverty, urbanization, and percentage of self-identified indigenous population; see Table 1 for a comparison of key variables. Program roll-out began in December 2015 and 2 rounds of recruitment were considered (2 rounds of training took place in each participant clinic in each state. The second round was implemented ∼7 mo after the first one with the objective of signing up more beneficiaries, mainly those who started their pregnancies after the first round). All pregnant beneficiaries with <32 weeks of gestation (during the prepilot, only beneficiaries with <28 weeks of gestation were invited, but in the pilot the period was extended up to 32 wk as to include more beneficiaries) were invited to participate in the selected clinics and those who did not own a mobile phone received one from the program (30%). In total, 3236 beneficiaries signed up to PD and received messages. The current qualitative evaluation analyses the FOI of the pilot PD in 5 Mexican states.

Methods

Stemming from implementation science, a qualitative evaluation to assess PD's FOI was conducted (21). The conceptual framework for implementation of Carroll et al. served to guide the qualitative evaluation design and data analysis (15). The framework poses that an intervention achieves high fidelity when adherence and program differentiation is high and moderators – quality of delivery, participant's responsiveness, intervention complexity, and facilitation strategies – positively contribute to achieve the intended outcomes. These elements are not sharply differentiated and there is overlap between them (15). These are the elements considered in the current qualitative evaluation:

1) *Adherence to the intervention* measures 2 aspects, first, who in the target population received the intervention and, second, how much and how often the prescribed content of the intervention was delivered (given and received dosage). Therefore, adherence assesses whether the implementation process succeeded in: recruitment, coverage, exposure of the target population, completion of content, and intended duration of the intervention. Consequently, it examines which obstacles reduced the adherence to the intervention and which population subgroups might experience difficulties using the intervention.

2) *Program differentiation* (the original framework of Carroll et al. defines program differentiation as part of Adherence (15); however, due to its importance in the RCT, we separated them into 2 elements to ease the presentation of key results) refers to the appropriate adaptation of the essential components in an intervention, i.e. the active ingredients required to reach the intended outcomes. Theoretically, an intervention can be meaningfully implemented if the essential components are preserved even if other nonessential elements of the intervention fail.

3) *Intervention complexity* recognizes that complex interventions have more elements and actors involved during their implementation. Therefore, greater complexity implies a greater scope for variation in delivery and are thus more vulnerable to low fidelity. Conversely, in interventions where the core components are simple, and activities specifically designated to a few actors, the heterogeneity of implementation tends to be lower.

4) *Facilitation strategies* refer to support actions or design elements seeking to optimize and to standardize delivery, such as guidelines and feedback loops. These strategies aim to ensure equal training and support. Complex interventions may need extensive support strategies to achieve high fidelity.

5) *Quality of delivery* is related to enabling conditions for the participants during the implementation of an intervention, such

as a comfortable site, appropriately trained facilitators, useful and comprehensible materials, amongst others.

6) *Participant responsiveness* refers to the participant's acceptance of the intervention. It assesses participant's expressions of engagement, satisfaction, and their perception that the intervention is relevant to them. Moreover, higher levels of fidelity are achieved when those participating in the delivery of the intervention (i.e. health providers) are enthusiastic about it.

Successful completion of these 6 elements means implementation has been effective and could yield the expected outcomes. Failure to achieve an element compromises FOI and therefore could reduce the potential effects of the intervention.

The current qualitative evaluation was conducted in 2 steps. First, based on the research protocol, the implementation materials, and discussions with the developers, the evaluation team established a benchmark for implementation. The 2 key parts of the benchmark were the identification of the essential components and a logic model that defined the key processes for the appropriate delivery of the intervention. Logic models establish links between resources, activities, outputs, and outcomes expected to happen in an intervention and help to communicate among stakeholders the intended intervention (22). Second, fieldwork was performed to assess the degree of compliance to the benchmark and the barriers and facilitators that moderated the actual implementation of the PD pilot. This process is guided by the 6 elements of the conceptual framework for implementation analysis of Carroll et al. (15).

Data Collection

During fieldwork, the evaluation team worked independently from the program developers (the evaluation team was co-ordinated by the first author and the corresponding author of the article, both holding a Doctor of Philosophy (PhD) in Social Welfare and Public Policy, respectively, trained in qualitative methods, and with substantive experience using qualitative methods for health and nutrition program evaluation. Both researchers designed the research guides and strategy, tested the materials during the prepilot, trained the research assistants on data collection for this study, and permanently supervised the data collection process. Two female research assistants and co-authors – with a master's degree in Public Policy and in Anthropology, respectively, and with previous experience in qualitative health program evaluations – collected most of the data during the training sessions and follow-up. Nonetheless, the whole team participated in data collection and analysis). The initial sampling unit were clinics, where pregnant women would be recruited and trained to use the SMS platform. Clinics were selected according to a maximum variability approach covering the 5 states, the 3 types of treatments, and rural/urban communities. The evaluation team visited 28 trainings in the health clinics in the 5 states, 57% of them were in rural communities. Clinics with different treatment assignments were included in each state (except Puebla).

Data was collected between January 2016 and September 2017. Data collection included: *1)* field notes from the implementation staff; *2)* closed observations of the trainings based on the manuals designed by developers; *3)* focus groups and face-to-face interviews

with participants in the clinics (i.e. when ≥5 beneficiaries attended the training, all were invited to join a focus group in the health clinic immediately after the session, but when <5 beneficiaries attended or when a focus group with ≥5 people could not be assembled, the evaluation team conducted face-to-face interviews in the health clinics); *4)* semistructured interviews with health providers (physicians and nurses) in the clinic; and *5)* semistructured interviews with community leaders ('vocales', as they are known in Mexico) in clinics linked to T2.

Approximately 4 mo after enrolling in PD, a new sample was selected to collect follow-up data. Eleven clinics which included the 5 states, urban and rural variability, and the 3 treatments (with an oversampling of treatments T2 and T3) were selected. Diverse techniques were used to collect data: *1)* focus groups in the clinics were conducted among women from the selected clinics who had participated in the trainings to use the platform; *2)* telephone semistructured interviews were performed among women who agreed to be contacted for follow-up and who were using PD's platform; *3)* semistructured interviews in the clinics were conducted with health providers directly involved in maternal and infant health. A description of the complete data collection process, by treatment, is in **Table 2**.

Specific guides were designed for each data collection technique. All participants – beneficiaries, community leaders, physicians, and nurses – signed written informed consent forms before providing information. PD was registered at AEA as AEARCTR-0001035 and received Institutional Review Board approval by Bristol University with number '26101'.

Analysis

Interviews and focus groups were audiotaped, transcribed, and anonymized for analysis. Once the data collection was complete, the evaluation team agreed on a codebook based on the 6 elements of Carroll et al. (15). Using grounded theory as the theoretical framework to analyze the data and the health clinic as the unit of analysis, 2 researcher assistants independently coded the material using NVivo 11, and a third researcher reviewed the coded material to establish key findings. By contrasting these elements with the logic model, the expectation was to find the sources of variation in implementation that may have hindered the effectiveness of the intervention (15).

Results

The first step of the FOI was to define a benchmark for implementation, i.e. to identify the essential components and to establish the key delivery processes. The *essential components* were: delivery of the SMS messages and appropriation of information (T1, T2, and T3); signing the contracts and the continuous participation of community leaders sending predefined messages (T2); and the evaluation of the health clinics by beneficiaries and the delivery of incentives by the State Ministry of Health (T3).

In addition, the delivery process was defined in a logic model as a process with 3 phases: *1)* launching and recruitment to the program; *2)* training of the beneficiaries; and *3)* the continuous use of the platform (**Figure 1**). Before the fieldwork began, the developers of the program

TABLE 2 Description of each treatment in the randomized controlled trial and the data collection for the qualitative evaluation of fidelity

	Randomized controlled trial	Qualitative evaluation
Treatment 1	Delivery of SMS messages in 107 health clinics	**Trainings** (7 clinics; 4 rural): —7 closed observations —2 focus groups (13 beneficiaries) —11 interviews to beneficiaries —12 interviews to health providers **Follow-up** (2 clinics; 1 rural): —3 focus groups (20 beneficiaries) —4 telephone interviews to beneficiaries —1 interview to health providers
Treatment 2	Delivery of SMS messages and community leaders sign 'contract' with beneficiaries and community leaders continuously send predefined messages in 111 health clinics	**Trainings** (11 clinics; 7 rural): —11 closed observations —0 focus groups —30 interviews to beneficiaries —10 interviews to community leaders —24 interviews to health providers **Follow-up** (5 clinics; 3 rural): —5 focus groups (35 beneficiaries) —10 telephone interviews to beneficiaries —4 interviews to health providers
Treatment 3	Delivery of SMS messages and beneficiaries anonymously evaluate medical care through their mobile phone and best performing clinics receive additional days of vacation in 108 health clinics as an incentive	**Trainings** (10 clinics; 5 rural): —10 closed observations —3 focus groups (20 beneficiaries) —17 interviews to beneficiaries —26 interviews to health providers **Follow-up** (4 clinics; 3 rural): —4 focus groups (35 beneficiaries) —20 telephone interviews to beneficiaries —7 interviews to health providers
Control clinics	329 health clinics	No control clinics were included

confirmed that these are the essential components and agreed that the logic model accurately represented the intended delivery process. The logic model helped the evaluators identify the key moments and information needed to assess the fidelity framework of Carroll et al. The first phase of the logic model (launching and recruitment) helped identify several parts of adherence, such as recruitment and coverage, and to assess the complexity of the intervention. The second phase of the logic model (training) helped to examine the initial exposure to the treatment, the completion of the content, and the duration of the training. Key inputs of the section were program differentiation during the trainings, quality of delivery, and immediate participants' response and health provider's buy-in. The third phase (continuous use of the platform) complemented information about the exposure, completion, and duration to the program, confirmed insights of program differentiation and completed responsiveness of beneficiaries and health providers. The 3 phases offered information about the facilitation strategies. The main results of the fieldwork are summarized in the following 6 elements.
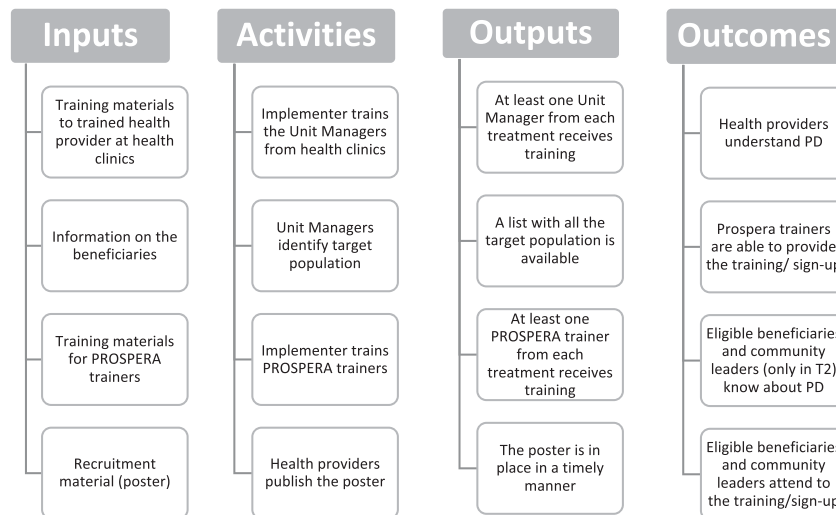
Adherence to the intervention

The implementation team trained ≥1 health provider per intervention clinic and, according to the field notes and closed observations, the recruitment of the target population was adequate in most health clinics. Examination of the field notes from the implementation showed that

PD was delivered to 90% of the intended clinics during the first round and to 86.7% during the second round. The most common causes of nondelivery were: absence of the target population (i.e. clinics with no pregnant women at the time of recruitment), missing material, lack of mobile signal, and road blockages. Health clinics had lists with the expected number of participants for each training. However, the closed observations revealed that this number coincided with actual attendants only in 11% of the visited clinics. In the rest of the clinics, the coverage was lower: 20–80% of the expected participants arrived at the initial training.

The training was divided into 3 parts: theory, practice, and sign-up. It was mostly delivered completely, but a few trainers modified the order of the sessions. The theory part was designed to last 35 min; however, trainers usually shortened this section by simplifying the program description. They always mentioned it was free, voluntary, nonpartisan, and the benefits of joining. Signing the informed consent forms to participate in the program and providing the phones took longer than expected (30 min on average).

The expected duration of the practice session was 40 min. Trainers struggled in this part because the ability to use the phone varied among participants and thus the pace of the session differed; young participants were often faster and those who received a new phone were slower than the rest. Differences in the phones forced the trainer to omit the section on how to use them (i.e. adding PD as a frequent number). Nonetheless, participants helped each other to catch-up with the session. Another

First implementation phase: Launching and recruitment to the program

Inputs	Activities	Outputs	Outcomes
Training materials to trained health provider at health clinics	Implementer trains the Unit Managers from health clinics	At least one Unit Manager from each treatment receives training	Health providers understand PD
Information on the beneficiaries	Unit Managers identify target population	A list with all the target population is available	Prospera trainers are able to provide the training/ sign-up
Training materials for PROSPERA trainers	Implementer trains PROSPERA trainers	At least one PROSPERA trainer from each treatment receives training	Eligible beneficiaries and community leaders (only in T2) know about PD
Recruitment material (poster)	Health providers publish the poster	The poster is in place in a timely manner	Eligible beneficiaries and community leaders attend to the training/sign-up

Second implementation phase: Training of the beneficiaries

Inputs	Activities	Outputs	Outcomes
Training materials at health clinics	PROSPERA trainers provide the training and key information for each treatment	Beneficiaries receive the full training and community leaders and health providers receive key information per treatment	Beneficiaries know what is PD and how to interact with PD
Eligible beneficiaries and community leaders attend to the training/ sign-up	PROSPERA trainers deliver mobile phones (when necessary)	All beneficiaries have a mobile phone	T2 beneficiaries sign contract with community leaders and receive their messages
RapidPro platform is fully functional	Beneficiaries practice in the platform	All beneficiaries practice in the platform	T3 beneficiaries understand the evaluation and health providers the incentives
Adequate message flows are available	Beneficiaries and community leaders sign-up and sign contracts	All beneficiaries receive SMS messages and answer back	Beneficiaries start interacting with PD

Third implementation phase: Continuous use of the platform

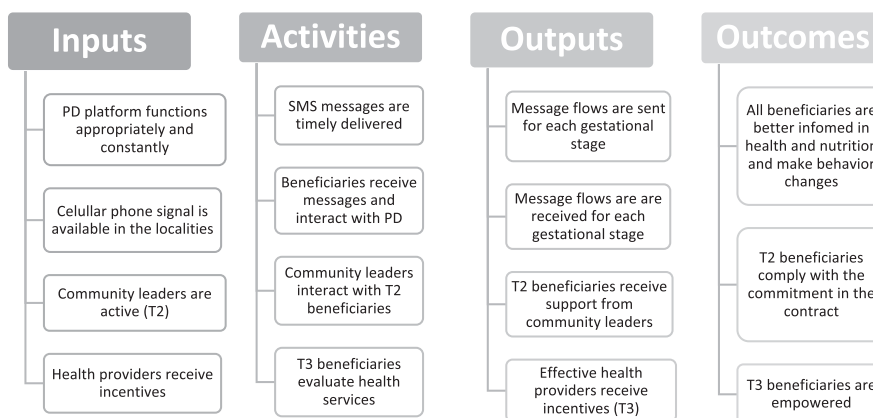| Inputs	Activities	Outputs	Outcomes
PD platform functions appropriately and constantly	SMS messages are timely delivered	Message flows are sent for each gestational stage	All beneficiaries are better infomed in health and nutrition and make behavior changes
Celullar phone signal is available in the localities	Beneficiaries receive messages and interact with PD	Message flows are are received for each gestational stage	T2 beneficiaries comply with the commitment in the contract
Community leaders are active (T2)	Community leaders interact with T2 beneficiaries	T2 beneficiaries receive support from community leaders	T3 beneficiaries are empowered
Health providers receive incentives	T3 beneficiaries evaluate health services	Effective health providers receive incentives (T3)	

FIGURE 1 Logic model for the 3 implementation phases of Prospera Digital (PD).

omission was the explanation on how to activate the trigger words for emergencies and appointments.

After the training, the lack of mobile signal inside some clinics or zones within the locality prevented the beneficiaries from practicing with PD's platform; according to the field notes, this happened in 38 clinics during both rounds. Unexpectedly, slight differences among carriers on how to sign-up delayed the session and excluded some participants. The program is free, but participants needed to have minimum credit on their phone (1 Mexican peso, roughly 0.05 Unites States dollars [USD]) to receive the first SMS and sign-up to the platform. They often did not know about the need for funds or always carried the phones without money because participants use them exclusively to answer calls. These problems divided the group between those who got into the platform and those who did not, leading to a low sign-up rate by the end of the trainings: 59% in the first round and 55% in the second. In order to solve this issue, the developers of PD added the possibility to sign-up remotely. Because of these difficulties, trainers frequently did not provide the motivational closure message.

Once participants started using PD, they reported reading, understanding, and responding to the SMS messages. Since the intervention was personalized, the number of messages the participants received once enrolled varied by gestational stage. Beneficiaries mostly remember receiving 2–3 message flows per week and they were sensitive to message fluctuation, noting the moments when the number of messages decreased and increased. They did not mind the fluctuation, but they wanted to receive more, rather than less messages. Administrative problems that led to late payments for the service resulted in a temporary suspension of the service. A group of participants in Puebla recognized this gap in the incoming flow of messages and reported that the program had "*abandoned them*". Importantly, illiterate beneficiaries and women for which Spanish is not their first language were instructed to ask family members for help – especially teenagers – in order to keep interacting with the platform. Follow-up interviews confirmed that they were able to ask for help and use the platform.

Even though most participants mentioned they were able to interact seamlessly with the platform, some network issues reduced the exposure to the intervention. One locality in Chiapas and one in Puebla lost the mobile signal for almost a month. Also, one locality in Chiapas near the international border received the signal from Guatemalan mobile carriers, which resulted in dropping it from the sample. Another barrier was that mobile carriers occasionally sent "*No Service*" or "*Invalid*" messages and beneficiaries took them as failures in PD. Another barrier resulted from default messages from phones indicating that sending messages could result in fees to the users. Beneficiaries interpreted messages from the carriers about the costs of their service as the cost for participating in PD. Some decided to add credit to their phones, but others stopped answering the 2-way messages, "*just in case*". Likewise, others reviewed their credit and confirmed the messages were free of charge. The result was that some read the 1-way messages but did not respond to the 2-way messages. The developers reacted to this limitation by designing mostly 1-way messages and ensuring that even if the message was 2-way, the beneficiaries would receive the relevant piece of information without needing to answer the message.

A worrisome exposure challenge was that several beneficiaries reported their pregnancy in the second or third trimester, which halved the exposure to the intervention during pregnancy. Moreover, another key challenge for full delivery was that beneficiaries frequently lost or changed mobile phones but they did not know how to get back into PD once they got a new one. They ignored the mechanism explained in a sheet with Frequently Asked Questions handed out during the initial training.

Program differentiation

Delivery of the messages (T1, T2, and T3).

The main function of the messages was to provide key health information aimed at promoting healthy behaviors. Participants valued the messages, as one of them highlighted: "*I learned a lot, I realized I was forgetting things, but now I know about the pregnancy, the babies, how to feed him, about diseases*". Qualitative evidence suggests participants did not use the messages to substitute their medical appointments and the reminders might even increase them; as exemplified in the following quote: "*I did not bring my other daughter to the clinic, but now I bring her each month, to get her vaccines and all*". The messages were useful to solve immediate problems, as another participant expressed: "*Oftentimes, I have no idea what to do, but now I have some guidance on how to take care of her*". Trigger words to change an appointment or for an emergency call were underused, probably because the trainers of PD frequently omitted the explanation. Even though an impact evaluation is needed to ascertain the effects of PD, participants reported behavior changes associated with the messages in several domains, such as improved experience with their pregnancy, preparations for delivery, medical care, nutrition, breastfeeding, and childcare.

Treatment 2: community leaders.

Implementation records show that community leaders were recruited inconsistently. They did not assist 5 out of the 10 observed clinics and 3 left early, partially because of the voluntary nature of the task. According to the closed observations, only 2 community leaders received the special training to perform their task. A trainer explained that the clinics "*Did not recruit the leaders, as they did not know they were supposed to do so*". The community leaders that did attend the trainings faced the same problems signing up to the platform as those described previously. Records indicate the sign-up rate was 60.8% on the first round and 47% in the second. Community leaders were also allowed to sign-up remotely.

The evaluation team recorded in their observations a lack a motivation from the community leaders. Interviews with the community leaders often confirmed this and highlighted that using the platform was difficult: "*The time to understand it is too short; if most people are slow learners, such as myself, then we'll face difficulties*"; "*I had trouble sending messages, dialing the asterisk, and with all the numbers; it was tough*". Moreover, interviews revealed that they did not know about their key activities: "*I understood that they will receive support to manage their pregnancy and nothing else*"; "*They called us because we are also mothers, for our testimony, so we can explain them how to take care and feed the baby*". None of the community leaders recalled the contract they had to sign with beneficiaries as a commitment with PD. One community leader explained that she was able to sign-up but it wasn't until the third month that she received a reminder and started sending messages; then, "*I sent some during 2 weeks but then I had to stop because PD was not available anymore*".

Unsurprisingly, the community leaders were not present in the beneficiaries' discourse about the intervention. None of the participants in the follow-up interviews reported interactions with their leaders; as a beneficiary mentioned: "*I never received her messages, zero, I received the others, though*"; "*I saw her (the community leader) and told me to wait for the messages, she did not know what to do, but expected to find out soon*". Notably, participants did not feel they needed help from the community leaders.

Treatment 3: evaluation of clinics.
The trainers omitted or superficially explained the evaluation element to the beneficiaries in 7 out of the 10 observed clinics. Some participants were unaware of the possibility to evaluate but the ones who understood the component expressed it would be useful so "*They can improve, provide a better service*". Importantly, they were not worried about a backlash due to negative evaluations: "*I do not think so, it is important for us to express how they provide the service and how they treat people*".

In only 3 of the observed clinics did the health providers receive specific information about the evaluation and the available incentives. Some trainers provided a general explanation of PD, as if they were in the messages-only treatment (T1) but did not mention the beneficiaries feedback nor the incentives for them. Nonetheless, health providers that did receive the full training had a positive opinion about the intervention. They mentioned that "*I would appreciate knowing what we are doing wrong, maybe, inadvertently, I use a very technical language, and this could help me change it*"; "*It is a good personal incentive showing that we are doing a good job and also the recognition to our unit would make me proud*". Overall, health providers in this treatment were more knowledgeable about PD.

During follow-up, health providers ignored when the evaluation occurred or which clinics were the winners – an award ceremony had already taken place. Beneficiaries recognized they were unable to differentiate information messages from evaluation feedback. They recalled the evaluation messages but did not realize they were evaluating.

Intervention complexity
Even though the co-ordination of 2 state ministries introduced complexity to the intervention, the implementation staff were able to organize most of the trainings. PROSPERA proved effective because all the staff were trained, the material was received on time, and training sessions were mostly adequate. Timely recruitment by the Health Ministry at the clinics worked with beneficiaries because they relied on previous medical records with contact information. Therefore, the basic procedures for the deployment of PD were placed correctly and on time.

The implementation became too complex when additional actors were included – especially health providers and community leaders – and the additional training for T2 and T3 was insufficient. Most health providers ignored what PD was, had misconceptions about it, and only those who were directly trained understood their responsibility. The health providers who knew about PD recognized that they did not know how the triggers were supposed to work and mentioned they had not received information related to any emergency. Responses suggest that participants did not trigger the centralized system of notifications or that the system was not functioning properly. In the T2

clinics, the recruitment of community leaders was inadequate, and they lacked incentives to participate. Thus, their role was not noticeable to beneficiaries.

Facilitation strategies
Trainers used detailed implementation manuals that guided them throughout the training and posters aided communicating key messages during the session. However, several health providers in charge of recruitment complained about the absence of communication strategies to reach the target population. Moreover, they did not have clear guidelines to help illiterate women and participants whose first language was not Spanish (i.e. indigenous languages, especially in Chiapas).

Beneficiaries received a simple and visual guide with basic information that explained the program, how to access the platform, and the type of interactions. They also received a separate sheet with Frequently Asked Questions, but the trainers often forgot to distribute them. During program roll-out, additional messages were sent to reinforce the training. These messages were especially useful to explain that the intervention was free.

An unexpected support strategy was needed during the implementation of PD. Lack of signal and difficulties with the commands to sign-up forced the implementation team to add a second option to get into the intervention. Developers created a format and instructed trainers that if sign-up was not possible, the contact information would be recorded in that format and beneficiaries would automatically be subscribed to PD. Once the list was compiled, the implementation staff enrolled them remotely. This was a successful solution to avoid an important number of drop-outs. However, several women were still unable to sign-up because trainers could not do it if the participants' mobile phones had insufficient funds.

Quality of delivery
The setting for several trainings was inadequate because it was not private and was uncomfortable for pregnant women (i.e. broken chairs, insufficient space, and bathrooms out of service). The trainers were aged between 30 and 40 y and 50% were male. Beneficiaries described the trainers (trainers had field experience dealing with matters related to PROSPERA with the communities. Thus, they were already known to some participants and knew how to acknowledge and speak to a crowd) as "*nice*", "*warm*", "*respectful*", "*trustworthy*", and highlighted that they spoke "*clearly*" and knew what they were talking about. Observers noted how most of the time the trainers were able to correctly answer the participants' questions. When the implementation team was present, the trainers ceded authority and asked them to respond to the questions, which distorted the regular session.

In 87% of the observed trainings the material was delivered fully and on time. Developers were concerned that providing a mobile phone for free to beneficiaries who did not have one would create conflict within the training session. For that reason, they chose to give the cheapest nonsmart phone available in the market (worth 20 USD). Beneficiaries usually had better mobiles and, on average, 42% received one (Chiapas 43%, Guanajuato 37%, Hidalgo 64%, Puebla 29%, and the State of Mexico 37%). Few conflicts were reported with the mobile phones.

Participants considered that the messages were "*clear*" and the language was "*simple and easy to grasp*". Messages that required numbers, such as dates or medical information, were the hardest to

respond to for them. Moreover, they felt the messages were timely and matched the developmental changes they were experiencing: it felt "*like magic, like they could foresee what was going to happen*". However, for the program to transition between pregnancy and infant messages, mothers needed to report the infant's birthdate in order to get precise information. This initially did not work and the developers had to design a transition mechanism that would insist asking mothers and, if a response was not sent, a date would be assumed.

Participants preferred to receive messages throughout the day – not in a bundle – so they could "*understand each one*" and "*avoid confusions*" of receiving several messages at the same time, especially when they received two 2-way messages, since they ignored how to answer the first one. Receiving them did not interrupt their daily activities; however, most liked to answer the 2-way messages at night, once they finished their activities.

Participant responsiveness

Interviews and focus groups with participants consistently showed accurate expectations with PD and considerable satisfaction with the training. The training was "*nice*" and "*interesting*" and the information was generally perceived as "*important*" and "*useful*". They appreciated the reminders, the health and nutrition information, and the possibility of changing the appointment and asking for help during emergencies.

After the training, most beneficiaries were able to point out the benefits of the program and what was expected from them. They felt that practicing with the tool was "*easy*" and most of them felt confident about using it correctly. The ones who could not join the program immediately felt "*frustrated*" and excluded. One participant was unable to send the sign-up message due to its cost and expressed that: "*I could not send the message and did not receive the information, as other ladies did*".

Once they were using PD regularly, they valued 1-way messages because they considered them to be the most "*informative*" and the 2-way messages because they showed PD is "*interested*" in them. Their favorite topics were nutrition during pregnancy, breastfeeding, and infant development. Participants expressed some frustration with the 2-way messages because they would want to extend "*the conversation*" beyond 2 responses and would like to be able to send their own questions. Messages were perceived as useful to solve problems and anticipate different situations. Even among women with previous pregnancies, the messages were "*helpful reminders of symptoms and circumstances, a useful update*". Some beneficiaries even reported that PD "*cared*" about them and felt "*abandoned*" when they did not receive messages. Participants expressed they would like the intervention to last until their infants were aged 2 y.

Discussion

The evaluation provided a nuanced description of how PD was used and implemented, and how the low FOI in the essential components of T2 and T3 could be affecting its expected outcomes. The results offered key details on how beneficiaries interact with these programs, which in turn shape their satisfaction and engagement. For instance, why they prefer 1-way messages and the reasons they read the 2-way SMS messages but might not have answered them. Furthermore, the responses of the participants about the usefulness of the messages, an

essential component for the 3 treatments, suggest that the intervention could yield valuable behavior changes, as has been observed in previous research (8, 9). Nonetheless, limitations of adherence and inadequate implementation of the essential components of T2 and T3 might have hindered behavior changes.

Important contextual and implementation barriers were related to the adherence to the intervention. The implementation team was able to recruit target participants and to deliver the program in the majority of health clinics; however, the coverage was limited because training session attendance fluctuated between 20–80%. A contextual barrier that reduced the intervention exposure period was the practice from pregnant women to delay prenatal services until the second or even third trimester of the pregnancy. More communication materials like posters and specific strategies for recruitment could help health providers improve this fundamental step. In addition, the requirement of having minimum credit on the mobile phone to send/receive the first message precluded eligible participants signing up to PD. Insufficient or absent phone signal hampered the trainings, excluded participants from sign-up, and limited the reception of messages. Moreover, once in the program, unexpected automated error messages from the carriers confused the participants and suggested that the interaction with the platform had a cost. An additional obstacle was the high rate at which participants lose or change mobile phones and their unfamiliarity with the process of resubscribing. These barriers reduced the exposure to the intervention, which may cause a "voltage drop", limiting population impact and sustainability (23). The implementation team had low leverage over carriers to solve such problems but SMS messages with reminders of issues such as errors, false fees, or ways to re-enroll can be useful and simple strategies to avoid high attrition.

In contrast, several facilitators strengthened the implementation of PD. Beneficiaries were highly responsive to the trainers who conveyed key information in culturally accepted ways. The training materials were simple, visually attractive, and included all the basic information. Even though the trainings were not always delivered in full, participants helped each other, especially with older participants who were not familiar with the mobile phones. This highlights the relevance of adapting contexts when implementing complex and large-scale interventions. Similarly, illiterate women and women with an indigenous background were able to ask for help, frequently from teenagers. Identifying these informal strategies was important because they may explain heterogenous effects and suggest ways of adapting the intervention to other contexts. One of these strategies may be to increase the practicing exercises in the platform during the first months and to send additional reminders of the information covered in the trainings.

Another important facilitator was the quality of the program. The messages were relevant, easy to understand and respond to, and participants even demanded a higher frequency of SMS messages. Likewise, the information was timely, useful, and perceived as precise according to their gestational and development stages. Notably, beneficiaries reported how the messages triggered an emotional response in which they felt taken care of by the program or abandoned in the absence of messages. Health providers showed support for the intervention and believe it could benefit participants. Both beneficiaries and health providers expressed satisfaction with PD, which may facilitate the buy-in of important stakeholders and effective implementation (24). Large interventions are complex, in part, because of the co-operation involved

among institutions and policies in the outer setting of the program (18). PD mostly succeeded in co-ordinating 2 state agencies and managed to deliver the majority of the trainings on time. Nonetheless, the outer setting should not be taken for granted in a new political environment, such as the one Mexico is facing in the light of a new administration.

A key finding of the evaluation is that the inadequate implementation of some essential components may hinder the effectiveness of 2 treatments in the experimental evaluation. PD tested 1 main treatment –messages only – and 2 additional treatments; a contract and motivation from community leaders (T2) and an evaluation of health services with incentives for health providers (T3). The main treatment had a high correspondence with the intended design. Adequate training, precise recruitment, quality of delivery, and adoption of the intervention suggested how program components may lead to key behavioral changes, such as additional antenatal visits, improved nutrition, breastfeeding, and quality childcare. Program fidelity, however, was low in T2 and T3 when compared with the intended design. T2 failed to recruit enough community leaders and trainers omitted its specific content, especially the contract. In addition, community leaders had difficulties using the platform and had low motivation to participate because they did not receive incentives. Beneficiaries rarely reported an interaction with the leaders and did not understand how they could help them. Therefore, the evidence suggests that the T2 was implemented in a similar way as the main treatment. T3 faced difficulties as well. Trainers did not fully explain its essential features to the beneficiaries and to the health providers. Untrained health providers had misconceptions about the program, so the effect of the evaluation was not well understood by all healthcare providers in participating clinics. In addition, a key assumption of T3 was to provide incentives to healthcare providers to improve their performance, while empowering pregnant women. But incentives were unclear to healthcare providers and evaluation mechanisms were unperceived by participating women, so T3 might not clearly differentiate from the main treatment. A new implementation strategy should provide a more intensive immersion of health providers and a more engaging mechanism to community leaders in order to fully observe their potential contributions.

PD is an example of how complex interventions need additional facilitation strategies to ensure the adequate implementation of essential components. These findings can help improve the quality of the design and the implementation of similar large-scale interventions through anticipating potential barriers and providing insights on the way to increase their performance. This is particularly relevant in a development context, where improving health system quality is paramount to improve health outcomes (25).

In conclusion, maternal and infant health and nutrition remain key challenges in middle-income countries. *mHealth* programs, such as PD, are attractive interventions that can contribute to build up mothers' knowledge and empowerment. Moreover, these interventions may help improve the quality of universal healthcare systems (25). Program evaluations based on implementation science offer useful designs to strengthen important programs, expand its sustainability, and promote its dissemination to new contexts (26).

The FOI showed key aspects that might reduce the impact of PD and that need revision before scale-up or replication into other settings. It also revealed that PD might have important population health and nutrition effects in a vulnerable population in Mexico. Therefore, the adequate implementation of PD is a promising strategy to contribute to the Sustainable Development Goals in Mexico and other middle-income countries.

References

1. Barker DJP. Mothers, Babies and Health in Later Life. London, UK: Elsevier Health Sciences; 1998.
2. Conti G, Heckman J. Economics of Child Well-Being. In: Ben-Arieh A, editor. Handbook of Child Well-Being: Theories, Methods and Policies in Global Perspective. Dordrecht, The Netherlands: Springer Publishing Company; 2014. p. 363–401.
3. OECD. Health at a Glance 2015. OECD Indicators. Paris: OECD; 2015.
4. Gutierrez J, Rivera-Donmarco J, Shamah-Levy T, Villalpando-Hernandez S, Franco A, Cuevas-Nasu L, Romero-Martinez M, Hernandez-Avila M. Encuesta Nacional de Salud y Nutricion 2012. Resultados Nacionales. Cuernavaca, Mexico: Instituto Nacional de Salud Publica; 2012.
5. INSP, UNICEF. Encuesta Nacional de Niños, Niñas y Mujeres en México 2015 - Encuesta de Indicadores Múltiples por Conglomerados 2015, Resultados Principales. Mexico City: Instituto Nacional de Salud Pública and UNICEF México; 2016.
6. Richter LM, Daelmans B, Lombardi J, Heymann J, Boo FL, Behrman JR, Lu C, Lucas JE, Perez-Escamilla R, Dua T, et al. Investing in the foundation of sustainable development: pathways to scale up for early childhood development. Lancet North Am Ed 2017;389(10064):103–18.
7. Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, Perez-Escamilla R, Rao N, Ip P, Fernald LCH, et al. Nurturing care: promoting early childhood development. Lancet North Am Ed 2017;389(10064):91–102.
8. WHO Global Observatory for eHealth. Global Diffusion of eHealth: Making Universal Health Coverage Achievable. Geneva: World Health Organization; 2016.
9. Tamrat T, Kachnowski S. Special delivery: an analysis of mHealth in maternal and newborn health programs and their outcomes around the world. Matern Child Health J 2012;16(5):1092–101.
10. Sondaal SFV, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, Klipstein-Grobusch K. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):1–26.
11. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: systematic review and meta-analysis. J Glob Health 2016;6(1):010401.

12. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013;1(2):160–71.

13. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, Labrique A; WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174.

14. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, Hensley M. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2010;38(2):65–76.

15. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci 2007;2:40.

16. Muralidharan K. Field experiments in education in developing countries. In: Banerjee A Duflo E, eds. Handbook of Economic Field Experiments. Hand Book of Field Experiments. Vol. 2. Oxford, UK: Elsevier; 2017.

17. Saunders RP, Evans MH, Joshi P. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. Health Promot Pract 2005;6(2):134–47.

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

19. Dantés OG, Sesma S, Becerril VM, Knaul FM, Arreola H, Frenk J. Sistema de salud de México. Salud Pública de México 2011;53:s220–32.

20. UNICEF. RapidPro: UNICEF; 2018 [Available from: https://community.rapidpro.io/about-rapidpro/].

21. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015;10:53.

22. McLaughlin JA, Jordan GB. Logic models: a tool for telling your programs performance story. Eval Program Plann 1999;22(1):65–72.

23. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013;8:117.

24. Jacobs SR, Weiner BJ, Bunger AC. Context matters: measuring implementation climate among individuals and groups. Implement Sci 2014;9:46.

25. Kruk ME, Gage AD, Joseph NT, Danaei G, García-Saisó S, Salomon JA. Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. Lancet North Am Ed 2018;392(10160):2203–12.

26. Tumilowicz A, Ruel MT, Pelto G, Pelletier DL, Monterrosa EC, Lapping K, Kraemer K, De Regil LM, Bergeron G, Arabi M, et al. Implementation science in nutrition: concepts and frameworks for an emerging field of science and practice. Curr Dev Nutr 2019;3(3):1–11.