Evaluation Planning Case Study

Maternity Text-Based Navigation Program

1. Determine Readiness to Evaluate

Why Evaluate as a Health Tech Start-Up? (Section 1a of the Toolkit)

A health tech company developed a text-based navigation program to help connect birthing people to appropriate care and other resources during and after their pregnancy.

- The program had been piloted and was growing.
- The health tech company was interested in assessing outcomes to demonstrate the impact of the program, to understand which subgroups were benefiting, and whether there were gaps or disparities in reach and impact.
- The company also wanted to make the case that the health system should continue to invest in this program beyond the initial pilot and implementation phase.

Bottom line: Evaluation could help ensure the sustainability of the program by demonstrating its impact on patients.

Product Stage of Development (Section 1b of the Toolkit)

The product's stage of development will impact what kind of evaluation will be most appropriate. In this case, the company had piloted the navigation program in two clinics. An initial internal quality improvement study was conducted to help refine the product and workflows.

The company was getting ready to spread the program to additional clinics. There was interest in evaluating the spread efforts and monitoring implementation and impact in different settings.

Bottom line: Evaluation would be useful to help understand how implementation was going in different settings.

Feasibility (Section 1c of the Toolkit)

- **Budget.** A foundation was interested in funding an external evaluation to understand the impact of the navigation program on disparities in birth outcomes to inform the field.
- **Staffing.** The company had a data analyst on staff but decided it did not have the capacity to conduct the evaluation in-house, so it also hoped to invest in an external evaluation to provide an independent perspective.
- **Level of reach and engagement.** The product had been implemented in two pilot clinics and was expanding into several others, with a significant increase in reach and engagement.
- **Timeline.** The company needed to have the results of the evaluation within 12 months to inform contract renewal with the health system.

Bottom line: It felt feasible to move forward with an evaluation at this point.

2. Establish an Evaluation Vision

Prioritize Equity (Section 2a of the Toolkit)

This program was being implemented in a complex ecosystem, across multiple provider organizations, among racially and socioeconomically diverse users. There were significant opportunities for miscommunication. Establishing a multistakeholder planning and implementation group was critical to ensuring that all partners started and remained on the same page.

Early discussions with the partners daylighted the need to create a logic model to ensure that all partners agreed on program implementation. The impacts of COVID-19 on implementation meant that some shifts were being made, so this step helped to ensure clarity. There were also data sharing assumptions and norms that were different between the health tech start-up and the health system. Regular evaluation workgroup meetings helped stakeholders to coordinate and come to an agreement on these and many other details.

Bottom line: Many ongoing discussions were needed to help prioritize equity and collaboration among partners.

Engage Partners (Section 2b of the Toolkit)

The company has a contract with a national health care system to spread the program to five additional facilities. The health care system has a strategic goal to address disparities in birth outcomes, which aligned with the goals of the navigation program.

The foundation hired an external evaluator with input from the tech company and health system.

The external evaluator formed an evaluation workgroup that included representation from the foundation, health system, and tech company. The workgroup included:

- External evaluation team
- Implementation lead from the health system
- Data analyst from the health system
- Implementation leads from the tech company
- Data analyst from the tech company
- Program officer from the foundation

The workgroup met regularly to design the evaluation. Although the external evaluator facilitated this process, there was substantial engagement from the health system, the tech company, and the foundation.

Evaluation Questions (Section 2c of the Toolkit)

The workgroup discussed what each partner wanted to learn from the evaluation. The tech company had done an initial quality improvement study that showed some improvement in patient engagement in prenatal care and improved birth outcomes. Given that this intervention was part of the health system's strategic goal to reduce disparities in birth outcomes, it wanted to understand if it had impacted outcomes specific to underserved patient populations. The workgroup identified these outcome questions:

- What was the impact of the program on health outcomes for birthing people and babies?

- Were there differences in engagement, experience, or outcomes by race/ethnicity or other demographic variables?

In discussing further, workgroup members elevated the importance of understanding patient experience with the navigation program and overall care at the hospital, so they added an evaluation question about patient experience:

- What is the experience of patients who participate in the program?

The workgroup was not initially interested in spending evaluation resources focused on implementation. However, due to some initial challenges with the launch at the new sites, they decided to include an evaluation component about implementation and added the following evaluation questions:

- How was the program implemented? How did implementation vary across facilities?
- What was the experience of the health system staff, providers, and navigators?
- Who was reached by the program? What services and resources were provided and utilized?
- What are potential areas for improvement for the partnership to achieve optimal outcomes?

Bottom line: The process of creating and refining evaluation questions helped partners reach agreement about goals and understand what components would be needed for an evaluation.

Logic Model (Section 2d of the Toolkit)

The evaluation workgroup first worked on developing a logic model for the navigation program to ensure everyone was aligned on the key activities and outcomes for the program.

In the logic model the workgroup developed, the long-term outcomes focused on health outcomes and reductions in disparities. The workgroup discussed whether the outcomes were realistic to expect from the program within the given timeline. Given that the program focused on pregnancy and birth-related outcomes, it seemed feasible that the program could impact birth outcomes within 12 months. However, some of the outcomes the workgroup was interested in looking at were relatively infrequent (e.g., infant mortality), so they expected it may be hard to say anything definitively about that due to small numbers. The workgroup determined it was still worth exploring; however, it was important to ensure that decisionmakers had realistic expectations about what an evaluation might and might not be able to detect.

In addition to identifying key short- and long-term outcomes, it was important for the workgroup to document the program's key inputs and activities to ensure all partners had a shared understanding of what was necessary to implement the program.

The workgroup then discussed how the activities were leading to the long-term outcomes. The shorter-term outcomes helped them identify what would indicate progress and whether the program was on track to influence the desired outcomes.

Inputs	Activities	Participation	Short-Term Outcomes	Intermediate Outcomes	Long-Term Outcomes
Tech company • Navigators	Predelivery	• All pregnant people seeking	• Patients are engaged	• Patients have increased knowledge of	• Increased healthy behaviors,

• Texting platform Hospital system • Leadership buy-in • Financial support • Engagement from clinical and outreach staff • Patient data Patients engaging with hospital system	• Screen patients for risk factors and social needs • Ensure health care access • Connect patients to community resources • Provide patient education (e.g., car seat education) • Listen to patient expectations and concerns • Escalate needs to hospital system, if needed Postdelivery • Assess maternal mental health • Support patients' lactation needs • Confirm pediatric and follow-up obstetric or midwifery appts. • Connect patients to community resources • Listen to patient expectations and concerns • Escalate needs to hospital system, if needed • Collect feedback	prenatal care, prenatal classes, birthing center tours, and those delivering at participating hospitals • Specific focus on Black birthing people	early in pregnancy • Patients have high utilization and engagement with navigation services • Patients have a positive experience with services, feel listened to and comfortable • Patients' short-term needs are met	resources and support • Patients have increased access to resources and support • Patients have increased self-efficacy (i.e., willingness to ask questions, raise concerns) • Patients have increased engagement in care and confidence in the health care system	decreased risk factors • Reduced complications during pregnancy and delivery • Improved infant health outcomes • Reduced maternal mortality • Increased initiation of breastfeeding **Impact** Reduced disparities and improved equity in birth outcomes, particularly for people of color

Inputs	Activities	Participation	Short-term outcomes	Intermediate outcomes	Long-term outcomes
Tech company • Navigators • Texting platform Hospital system • Leadership buy-in • Financial support • Engagement from clinical and outreach staff • Patient data Patients engaging with hospital system	Pre-delivery • Screen patients for risk factors & social needs • Ensure health care access • Connect patients to community resources • Provide patient education (e.g., car seat education) • Listen to patient expectations & concerns • Escalate needs to hospital system, if needed Post-delivery • Assess maternal mental health • Support patients' lactation needs • Confirm pediatric and follow-up OB/midwifery appts • Connect patients to community resources • Listen to patient expectations and concerns • Escalate needs to hospital system, if needed • Collect feedback	• All pregnant persons seeking prenatal care, prenatal classes, birthing center tours, and delivering at participating hospitals • Specific focus on black birthing persons	• Patients are engaged early in pregnancy • Patients have high utilization and engagement with navigation services • Patients have a positive experience with services, feel listened to, and comfortable • Patients' short-term needs are met	• Patients have increased knowledge of resources & support • Patients have increased access to resources & support • Patients have increased self-efficacy (i.e., willingness to ask questions, raise concerns) • Patients have increased engagement in care & confidence in the health care system	• Increased healthy behaviors, decreased risk factors • Reduced complications during pregnancy & delivery • Improved infant health outcomes • Reduced maternal mortality • Increased breast-feeding initiation **Impact** Reduced disparities and improved equity in birth outcomes, particularly for people of color

Bottom line: Having an agreed-upon logic model helped ensure that all partners were on the same page. It was used during the evaluation design process for reference and communication and allowed the evaluator to proceed to selecting an appropriate study design.

3. Identify Evaluation Measures and Design

Identify Measures (Section 3a of the Toolkit)

The evaluation workgroup identified measures for each of their evaluation questions.

Evaluation Question	Indicators	Data Source	Data Collection Method Frequency/Timing
What was the impact of the program on health outcomes for birthing people and babies?	Birthing person outcomes: • Gestational diabetes and hypertension • Pre-eclampsia • Elective c-section • Hospital utilization / length • Complications during birth • Breastfeeding initiation Baby outcomes: • Apgar score • Preterm birth, low birthweight mortality • Length of stay / NICU time	• Health system data / electronic health record	• Data pull at the midpoint and end of the study period
Were there differences in engagement, experience, or outcomes by demographic variables?	• Race/ethnicity of patients engaging with program • Analysis of utilization, satisfaction, and outcome data by race/ethnicity	• Health system data for demographics	• Data pull at the midpoint and end of the study period
What is the experience of patients who participate in the navigation program?	• Patient-reported satisfaction (e.g., needs were met, felt listened to, engagement) • Patient engagement in care and trust of health care system • Patient knowledge of available resources • Net promoter score	• Patient interviews • Navigation program data	• Rolling interviews at 3 months postpartum • Data pull at the midpoint and end of the study period
How was the program implemented? How did implementation vary across facilities?	• Documentation of services provided and perceived effectiveness of implementation • Effectiveness of information flow and escalation process • Community partner engagement	• Staff and provider interviews • Document review of workflows	• Between XX and YY date • ASAP with one more review at study end for changes
What was the experience of the health system staff, providers, and the navigators?	• Satisfaction of staff, providers, navigators with the partnership • Perceived benefits and challenges of partnership • Perceived benefits to patients and care quality	• Staff and provider interviews	• Between XX and YY date
Who was reached by the navigation program? What services and resources were provided and utilized?	• Demographics of people reached • Patient utilization of program's screening and resources by stage of pregnancy (e.g., bidirectional touchpoints, completed screening) • Patient utilization of system resources	• Navigation program data • Health system data	• Data pull at the midpoint and end of the study period

Evaluation Question	Indicators	Data Source	Data Collection Method Frequency/Timing
What are potential areas for improvement for the partnership to achieve optimal outcomes?	• Patient and staff/provider perceptions of areas of improvement • Analysis of challenges and gaps	• Synthesis of all data collected	• Ongoing review and discussion at monthly evaluation meetings • End of study synthesis

As emphasized in the table above, the evaluation workgroup made sure to include mixed methods — building in both qualitative and quantitative data to ensure they would know both what happened and also how and why the program achieved (or didn't achieve) its desired outcomes. The implementation questions were also important to understand the extent to which the program was implemented and how that impacted outcomes.

Bottom line: Establishing indicators and data sources allowed the workgroup to know exactly what they were asking of whom, which was needed to establish data sharing agreements and move forward with planning for data collection. For that, data specs were needed, and below is a snapshot of the one used for this evaluation.

Create Strong Data Specs (Section 3b of the Toolkit)

Data specs lay out details about each measure to facilitate efficient and effective sharing of data. The programmer from each partner and the evaluation team met several times to decide how best to merge the data, clean each variable, and collect needed information for analysis. They created test data sets early on to get a sense of what was in each data set, and then once the data sharing agreements (DSAs) were in place, two iterations of early data sets were requested to finetune further. This allowed for more up-front decisionmaking ahead of the final data pull and analysis.

Domain	Category	Variable	Data Source	Variable Name	Description of Variable	Frequency/Timing of Data Collection	Notes on Inclusion/ Exclusion
Example: Health and demographics	Birthing person	Gestational diabetes (GD)	Electronic health record	DMGest	Checkbox yes	One-time data pull at the end of the study period	Every pregnant person who gave birth between XX and YY date, with a dx of GD
Example: Health and demographics	Baby	Birthweight	Electronic health record	Birthwt	Numeric – kg	One-time data pull at the end of the study period	Every baby born between XX and YY date

Evaluation Design (Section 3d of the Toolkit)

The workgroup reviewed the questions, the general timeline, the logic model, key questions, needed metrics, and the workflow of the organizations. It was determined that a randomized controlled trial would not be

feasible, nor would it be reasonable to identify a control group that would not be offered services. Workgroup members decided that a pre/post nonexperimental design, potentially looking at "dose-response" (comparing level of engagement and outcomes) would be the best choice for their needs. They built in a strong qualitative component to data collection in order to hear directly from staff and patients about their experiences and the impact they felt. The final evaluation design included an evaluation of implementation steps as well as longer-term outcomes to provide information for program improvement in addition to understanding impact.

Bottom line: Having a robust logic model and evaluation question conversations and the evaluation workgroup with key partners at the table meant that conversation around evaluation design was much more cohesive and productive, and the final plan met everyone's needs.

The company had already invested in a small-scale, exploratory study, so it wanted to have as rigorous a design as possible to look at outcomes. The evaluation workgroup thought through the feasibility of different designs and documented challenges that influenced what was feasible. Challenges included these:

- Randomization was not possible because the navigation reached out to all pregnant people that the hospitals interacted with, so a randomized controlled trial could not be pursued.
- No viable comparison group at other facilities existed because the variation in other services available to pregnant people would create too many confounders in understanding the contribution of this navigation program.
- A comparison with a similar cohort the year before could have been compared with the current cohort, but it was participating during the COVID-19 pandemic, and health care utilization had changed so much that the workgroup didn't think looking at a previous year would be comparable.

Due to these challenges with identifying a viable comparison group, the workgroup decided the best design for the intervention was a type of pre/post design looking at the cohort of pregnant people engaging with the health system during an 18-month period. The design would focus on dose-response design, looking at whether those that engaged more with the navigation program or earlier in their pregnancy had better outcomes than those who didn't. The workgroup recognized with this study design that there were a number of potential biases, including that those who actively engage in things like navigation may be more engaged in their health care in other ways, which could impact outcomes.

Bottom line: Selecting the "right" study design involved a lot of back and forth and compromise, but in the end, the evaluation workgroup was able to find a design that worked with the resources available and would allow the evaluation to answer the agreed-upon evaluation questions.

4. Operationalize the Evaluation Plan

Develop a Budget and Timeline Through a Feasibility Review (Section 4c)

Once they had the evaluation study design and questions laid out, the evaluation workgroup did another feasibility review to iron out logistics.

Capacity. The external evaluator would handle data management and analysis. We also had several discussions with all key partners about who could join the regular evaluation meetings, and what data and supports would be needed from the tech company and health care system to conduct the evaluation.

Data access. As soon as we had a design, the evaluation workgroup started the process of an institutional review board (IRB), BAAs and DSAs as required, and it still took three months longer than anticipated.

Budget. The budget was finalized, ensuring adequate resources for all partners to conduct the work.

Timeline. The workgroup then agreed on a timeline and built a workplan for each row of the table. The timeline and workplan were revisited in every meeting so that when work inevitably shifted, implications and needed pivots could be discussed. For example, the DSA took three months longer than expected to finalize, which impacted clinical data collection and patient interviews.

	Planning (3 Months)			Evaluation Implementation (12 Months)											
	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
Coordination with evaluation workgroup	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Evaluation planning															
Logic model development	X														
Evaluation plan		X	X												
IRB & data sharing agreements	X	X	X												
Program familiarization & document review	X	X	X												
Clinical data collection															
Define metrics & data specs	X	X	X												
Validate & refine data pulls				X											
Initial data analysis					X										
Midpoint data pull								X							
Final data pull												X			
Final analysis												X	X	X	
Interviews (primary data collection)															
Develop, test, refine staff & patient interview protocols			X	X											
Staff interviews				X	X						X	X			
Patient interviews							X	X	X						
Qualitative analysis						X				X	X	X	X	X	
Reporting															
Internal learning session						X									
Interim report									X						
Final report														X	X

Bottom line: The workgroup determined that implementing the evaluation as designed was feasible but that it needed to continually revisit capacity, timeline, and budget so it could respond to changes and delays during the evaluation.

Agreements (Section 4d of the Toolkit)

For this evaluation, the health system and the tech company needed to have a business associate agreement (BAA) to share identified data with each other so the data could be matched. Staff at the health system

deidentified the data sets and provided them to the external evaluator for merging and analysis based on the existing data sharing agreement between the health system and evaluator. Due to contracting issues, absences of critical personnel, and legal reviews, it ended up taking six months to get the BAA in place and to begin data collection, which impacted timelines and the ability to conduct data collection.

The external evaluator and the health system both had to submit the evaluation to their IRBs for review. This process went relatively smoothly and quickly, though often this step also can have delays. Both IRBs deemed the evaluation not research because the work was "nonsystematic" and not necessarily generalizable because its main purpose was for internal quality improvement.

Bottom line: It can take months to decide what agreements are needed and then get them place, so timelines need to include significant start-up time and may need to be revised.