
Women for Women International

Goldilocks Toolkit

Innovations for Poverty Action

poverty-action.org/goldilocks



Right-fit monitoring and evaluation (M&E) systems embody the principles of Credible, Actionable, Responsible, and Transportable, or CART. In the Goldilocks case study series, we examine the M&E systems of several innovative organizations and explore how the CART Principles can work in practice.

Acknowledgements: The Goldilocks toolkit was authored by Mary Kay Gugerty, Dean Karlan, and Delia Welsh. Lead authors for this toolkit resource also include Sana Khan and Tetyana Zelenska, with editing and design by the IPA Communications Team (David Batcheck, Laura Burke, Jennifer Cowman, Heidi McAnnally-Linz, Megan McGuire).

Women for Women International: Monitoring and Evaluation in Conflict and Post-Conflict Settings

Women survivors of war and conflict are disproportionately affected by acts of violence, displacement, poverty, and loss of property and relatives. Conflict disrupts familial and community networks, compelling women to assume greater responsibility for generating household income and supporting their dependents and community. Women for Women International (WfWI) works in countries affected by conflict and war and addresses these issues by supporting women to earn and save money, improve health and well-being, influence decisions in their home and community, and connect to networks for support.

This case study examines WfWI's collection and use of data in conflict and post-conflict settings to monitor and measure the results of their work. Despite the challenging setting, WfWI has developed a data collection system that produces high quality data and is in the process of making important changes to

their main indicators to ensure that they appropriately capture the work of the organization. Building on current efforts to improve M&E within the organization, we have two primary recommendations: one, that WfWI conduct a rigorous impact evaluation and two, that they modify the IT infrastructure and data collection

processes in order to provide faster and more useful information for operational decision-making in country offices.



What They Do

WfWI is an international non-profit organization that provides marginalized women survivors of conflict with access to training and services to develop income-generating skills, basic knowledge, and support networks to rebuild and transform their lives. Since 1993, WfWI's eight country teams have worked with more than 429,000 women in post-conflict settings in Asia, Europe, and Africa. As of 2016, WfWI has active country operations in Afghanistan, Kosovo, Nigeria, Democratic Republic of Congo, Rwanda, and South Sudan, as well as partnerships with local NGOs in the Kurdistan Region of Iraq and an affiliated locally-registered NGO in Bosnia and Herzegovina.

The foundation of WfWI's work is a 12-month training program for marginalized women. At enrollment, women are grouped into classes of 25, and their training program includes: life skills, such as health, wellness, and rights awareness training; income-generating skills and vocational training to give women basic business knowledge and practical experience in applying market-based skills; and a monthly stipend.

To strengthen and support the ability of women to apply their skills and knowledge at home and in their community, WfWI began a Men's Engagement program in Nigeria in 2001 that now operates in five countries.¹ Tailored to meet local and regional

needs, the program teaches traditional male leaders and influencers in the community about the training program's role in strengthening communities and families. This includes informing men about women's rights and enlisting them to support efforts to combat gender-based violence.

In addition to this men's engagement, WfWI has developed partnerships with local businesses and organizations to provide women graduates who have started individual or group businesses, cooperatives, and associations with additional resources and networks to advance their skills and reach.

Theory of Change

In 2014, WfWI undertook a program review and is currently updating its theory of change through a participatory process that engages country office leadership and field staff. At the core of its work is the belief that marginalized women living in conflict-affected areas need a combination of psychosocial and economic support to improve their lives. The year long program focuses on four domains:

1. Earning and saving money.

Participants receive training in basic business skills and an income-generating activity in one of five key sectors: agriculture, handicrafts and manufacturing, trade and services, livestock, and food processing. The program also provides a \$10 monthly stipend to participants and encourages them to save.

2. Health and well-being.

The program provides health training, on topics such as family planning, nutrition, and stress management.

3. Decision-making in their homes and communities.

The program helps participants understand their rights and the importance of civic participation and negotiation.

4. Networks for support.

The program trains women on how they can organize themselves into groups for advocacy, social support, and cooperative economic activity, starting with their class of 25 peer program participants.

The program aims to generate the following intermediate outcomes:

- startup or growth of a business, self-employment, or employment
- new health knowledge
- increased confidence in one's ability to influence decisions
- improved access to social and community support networks

The ultimate goal of the program is to enable participants to sustain an income, be physically and psychologically well, have strong social networks and safety nets, and be decision-makers.

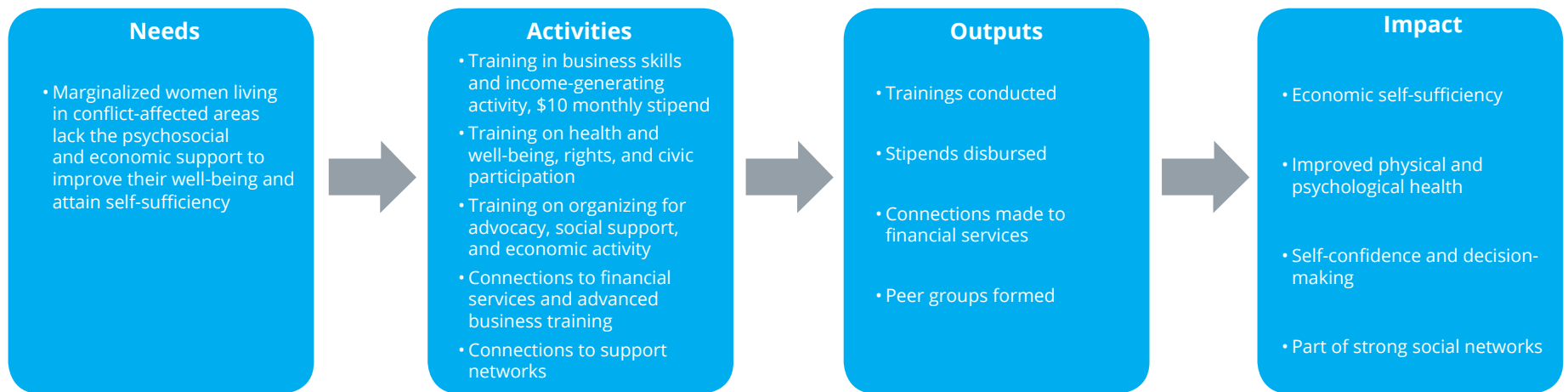


FIGURE 1. THEORY OF CHANGE*

* Organizations use a variety of methods to present their theories of change. To standardize our discussion of these cases, we present our own simplified version of Women for Women International's theory of change here.

Data Collection

Setting up and maintaining an M&E system in conflict and post-conflict settings is particularly difficult. In fragile security environments with damaged infrastructure and/or displaced populations, identifying and following up with program participants takes a lot of time and effort. This additional time and effort is costly. Even after locating participants, they may be very sensitive to detailed questions about their income, well-being, and personal lives, making accurate measurement of key outcomes difficult.

Despite these challenges, WfWI successfully collects data from participants at various points during the lifecycle of the program and also

conducts follow-up surveys one and two years after program completion. Recently, the organization has made significant efforts to improve its data collection processes and quality, shifting from paper to electronic data collection and reducing the amount of data collected per class to a more manageable size. For example, data are now only collected for a representative sample of the total population of program participants.

The switch to electronic data collection and representative sampling has allowed M&E teams across the organization to build in more robust quality assurance checks in both the field and at headquarters. Although country offices

do not have access to the full historic database maintained on a server at headquarters, they have access to recent data collected electronically in the field, in real time, through an online survey data management tool.²

While electronic data are reviewed locally and then sent to headquarters for aggregation with the master backend database, IT constraints currently prevent country office teams from directly accessing the master database (which contains current electronic records and previous records collected through paper forms) from their locations. As a result, local teams cannot easily analyze historical trends in the data and must rely on the headquarters analysis staff for

support. To address this challenge, WfWI is seeking funds to invest in the development of a dashboard tool that will allow country offices to access the full historic database through a simple interface and conduct various types of analyses.

M&E staff are primarily responsible for data collection activities, collecting baseline and endline data before and after the training sessions. When multiple data collection exercises are scheduled at the same time and the work exceeds the capacity of the M&E team, implementation staff may help out with data collection.

The close coordination between the two teams has obvious advantages in terms of efficiency, especially in fragile contexts, but does create an unintended risk: participants may report what they think will benefit them or please program staff, rather than the truth. For example, when program staff are involved in data collection, participants might report less income than they actually have to improve their likelihood of eligibility for future programs, or they may over-report

their income or learning at the end of the program to show appreciation for the program. This is a challenge that most field-based organizations face in collecting data from participants. Organizations often have to make a difficult tradeoff: hiring independent surveyors may reduce bias but it costs more.

Activity Monitoring

The organization tracks four main categories of outcomes and uses select indicators in each outcome area to produce external reports. Headquarters determines the indicators and standard survey instrument that all country offices use and report on an annual basis.³ The current core survey instrument and guidelines were developed in 2009. Select measures in the focal outcome areas include the following:

Women earn and save money

- measured by average reported monthly personal earnings and savings, and the number of women earning a minimum income of \$1.25 per day.

Women improve health and well-being - measured by the number of women who report having and being able to maintain the level of physical and psychological health necessary to actively participate in the family and community.

Women influence decisions in the home and community - measured by the number of women who report contributing to decisions in the family and the community.

Women connect to networks for support - measured by the number of women who, by themselves and in solidarity with others, aid the cause of women by promoting and

advocating shared needs and desires.

The programs team also collects operational data (such as the attendance and performance of trainers at trainings) but this information is not formally part of the M&E system. As needed, WfWI also collects data beyond this core system, depending on specific requirements in grant agreements or men's engagement activities.

Along with efforts made to improve data collection and quality, WfWI is reviewing and revising its performance indicators to better measure the progressive steps in the program's theory of change. Country offices have a lot of discretion in deciding what activity monitoring data should be

gathered and how it should be collected. In the Rwanda office, we observed that M&E staff were well-trained and familiar with the steps they should follow in program monitoring.

We also observed practices that were not explicitly noted in the written documents and protocols available to M&E staff. For example, the M&E team conducts regular monitoring visits to check on the quality of the training sessions. On these visits, the team is primarily looking for deviations from the program, for example, if attendance is exceptionally low and has not been reported by the trainer, or if the trainer should be delivering a particular module but is instead focusing on something else. The typical outcome from the monitoring visits includes a report submitted to the supervisor on the findings from the visit, a conversation with the trainers if there is a discrepancy between the standard program framework and guidelines and practice, and potentially a discussion with the program team managers.

While the process seems to be quick and allows managers to address

implementation issues immediately, we recommend reinforcing this practice with protocol documentation on the specific steps to be followed in a monitoring visit and required information to be collected. This would likely strengthen the data collection and allow managers to analyze the program data alongside the M&E data to identify trends and bigger picture operational questions.

Measuring Impact

WfWI's impact measurement system uses a basic before-and-after comparison of the participants. The organization conducts four surveys on a sample of women participating in the program (baseline, endline, one-year follow-up, and two-year follow-up).

Participants are asked the same set of questions in all four rounds, with the addition of questions related to women's participation in the program after the baseline (e.g. satisfaction with specific areas of targeted change through the training, use of vocational skills gained in the training, use of stipends, etc.). The questions focus on the four main program outcomes: income and saving, health knowledge and practice, decision-

making in the household and community, and connection to support networks. These indicators are particularly valued by donors and supporters to assess the position of women before and after participation in the program. However, because they are simple before-and-after measurements without a valid counterfactual, they do not credibly estimate the program's causal impact on these outcomes.

Given the absence of a valid comparison group, the before-and-after method of assessing impact has important limitations in assessing the 'average causal impact' of a program. Because we do not know what would have happened to individuals had they not been in

the program, we cannot estimate the impact of the program on the outcomes measured. This is an issue that WfWI recognizes well and as a first step to addressing this issue, language that states the limitations of the before-and-after approach is now used alongside the use of M&E data. Unfortunately, before-and-after reporting seems to be a standard requirement in most donor reporting templates, which makes incorporating a counterfactual difficult as a standard practice.

The political and physical insecurity in WfWI countries makes it challenging to use randomized control trials (RCTs) or other methods to create a rigorous counterfactual. Tracking participants over

time is more difficult given the instability in some of the areas where WfWI works. Security concerns may limit the number of areas for program implementation, restricting the potential sample for an RCT. Non-participants recruited to serve as a counterfactual group for an RCT may have higher levels of unease/suspicion of survey activities, compared to participant communities and individuals who have greater familiarity with WfWI's work and goals. And for staff already dealing with security issues, adding a complex evaluation framework can impose too much of a burden on them.

Nonetheless, these challenges are not insurmountable, and WfWI is currently seeking donor support for conducting a rigorous evaluation, possibly an RCT.⁴ Since evaluation costs are often part of project budgets, the support of donors who are well-versed in different evaluation methodologies and can prioritize including larger budgets for more rigorous evaluations within grants is important.

Goldilocks Recommendations

Women for Women International has a well-established program and has made significant improvements to its data collection system and activities in recent years. The organization is continuing its efforts to refine indicators and make data more usable by the country offices. **Our main recommendation is to pursue a randomized evaluation to provide lessons that can help improve the model and contribute to a growing evidence base for social protection and women's empowerment programs.**

Credible: Collect high quality data and accurately analyze the data.

There are two aspects to consider for WfWI's impact data and analysis: whether the evaluation design establishes a causal link of the program to important outcomes, and the quality of the data collected.

On evaluation design, the current method of impact measurement, a before-and-after comparison of outcomes for participants, absent a randomized evaluation, does not provide the organization with enough data to estimate causal impact. We recommend that WfWI pursue a randomized evaluation to validate the model, measure which aspects are the most vital for impact, and contribute to a growing evidence base for social protection and

women's empowerment programs. Running a randomized evaluation in several sites would allow WfWI to compare the program in different environments. However, constrained M&E resources suggest that starting with one site may be the most feasible, and would produce valuable lessons that could inform other country programs.

On data quality, the organization has implemented electronic data collection with supports for ensuring data quality in the baseline and follow-up surveys. To build on the current system, we made two recommendations for data collection practices in the field:

1. Update data collection guides and

survey instruments to more fully capture all of the practices survey teams employ in administering a survey. For example, survey instruments ask respondents whether they have knowledge of important health practices. We observed that surveyors use a test to gauge knowledge, which is a much more reliable measure. However, the test itself is not documented in the data collection guides and instruments. Improving the documentation of field procedures are likely to improve data consistency over time and may help standardize measurement across country programs.

2. Notify respondents of the separation between program implementation and data collection for M&E. In responding to surveys implemented by WfWI staff, program participants may be tempted to provide answers related to the criteria for program participation. While collecting data at the same time as training visits has operational efficiencies given available resources, respondents may not distinguish between the teams responsible for implementing the program and M&E.

We recommend developing a script and other procedures to clarify that data collection is independent from program implementation.

Actionable: Commit to act on the data you collect.

WfWI has undertaken improvements to strengthen how the country offices and headquarters work together in preparing donor reports, budgeting and planning, and designing and adapting programs. We recommend that as part of this process, WfWI creates a system to integrate learning and communication around the data reports into the process. This would ensure systematic and regular feedback between managers in the field and at headquarters who are using the data from the M&E system.

We also suggest a helpful exercise for focusing data collection on the most important indicators: develop a decision rule for each measure and a threshold that would trigger a change in the program if it were crossed.⁵ For example, the standard survey tracks the portion of women who practice family planning. To

make this indicator actionable, it would be helpful to decide what a country office would do if this number doesn't change or goes down after participating in the program.

Responsible: Ensure the benefits of data collection outweigh the costs.

WfWI incurs a high cost in tracking a large range of indicators for the women that participate in their program, even though the M&E team has created systems to reduce the burden of data collection. We recommend that the M&E team work closely with other departments at headquarters and country offices to map its current data collection plan to the theory of change, and consider reducing or revising data collection that is less relevant to program operations.

Transportable: Collect data that will generate knowledge for other programs.

Conducting a rigorous impact evaluation

of WfWI's program in one or multiple sites could help inform program design and allow for comparing the program in different environments. A study of the program's impact is likely to add knowledge to growing evidence on social protection programs and would constitute a public good. In this, the principle of responsibility (collecting data whose benefits outweigh costs from the perspective of internal organizational learning) needs to be balanced with the principle of transportability (collecting data for learning that can influence program sites in other countries where the organization works, or further, to influence other organizations' programming). For example, rigorous evaluations of an ultra-poor graduation model⁶ show strong positive impacts on household welfare, including income, consumption, savings, and health in the short and medium term. Some components of the graduation model overlap with the WfWI program, such as vocational skills training and consumption support. However, there are also some key differences: the WfWI program does not include a transfer of a productive asset (such as livestock), but

has more training and information on individual and property rights, decision-making, and building support networks.

Lessons for Others

1. Test and implement M&E processes that improve data collection efficiencies.

Collect data from a representative sample of participants in more depth, rather than conducting light-touch and high-cost surveying of all program participants. Focus data collection on key indicators. Investing in a robust electronic data collection infrastructure will also deliver gains in terms of cost, quality, and speed of data use.

2. Track participants beyond program completion.

Within the bounds of available resources, conducting panel tracking surveys of

program participants can be helpful in viewing participant trajectories post-program and identifying areas in which future participants might benefit from stronger or new programming.

3. Document M&E protocols and regularly train staff in them.

Clearly documenting M&E protocols and training all staff to collect data in the same way help ensure the consistency and reliability of data across program sites. Encouraging M&E staff to maintain their independence from program implementation, raise issues they have with data collection, or to propose new ideas for collecting data may also boost the credibility of the data they collect.

4. Design impact evaluations that measure overall causal impact and test programmatic tweaks.

When designing a rigorous impact evaluation, the study should measure the impact of the overall program, but also ideally allow for some testing of specific program components. This provides more learning for WfWI and for others that are implementing similar programs.

Endnotes

1. WfWI defines men's engagement as the process that aims to recruit men as allies in women's empowerment by improving their knowledge about health, social, and economic issues that negatively affect the situation of women.
2. The process includes several data quality checks: checks in the programmed survey, checks by the M&E Managers in the Country Offices (who view every batch of electronic data using an online database tool that temporarily stores the records), and finally checks when data are imported from the online database tool to the organization's backend SQL database.
3. Country Offices can also design complementary instruments for supplemental data collection or as required by specific donors. Instruments that are designed in the country offices are sent to headquarters for review and feedback, and vice versa.
4. Discussed in the organization's Research & Evaluation Strategy paper, and they have been engaging with researchers to move this agenda forward.
5. Visit Goldilocks Toolkit: Monitoring for Learning and Accountability, available at: <http://www.poverty-action.org/publication/goldilocks-toolkit-monitoring-learning-and-accountability>, which spells out the decision rule idea in more detail.
6. The **Ultra-Poor Graduation Model** is a livelihood program that consists of livelihood training, productive asset transfers, consumption support, savings plans, and healthcare. Rigorous impact evaluations in seven countries found large and lasting effects on a range of outcomes for program participants. J-PAL and IPA Policy Bulletin. (2015). Building Stable Livelihoods for the Ultra-Poor. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab and Innovations for Poverty Action.