



The Western Highlands Integrated Program (WHIP) Evaluation Baseline Survey in Guatemala: *A Case Study in Evaluation Practice*



2014

Acknowledgements



The author with fieldwork staff in San Marcos.

The USAID Bureau for Global Health, through its Office of Health, Infectious Disease and Nutrition (HIDN), provided the financial support that enabled this activity. Lisa Maniscalco served as the technical advisor at USAID and allowed me the freedom to implement key tasks flexibly as the evaluation unfolded. Early planning was also supported by Sally Abbot and Anne Swindale, who helped to develop the case study concept. Tom DiVincenzo and Ángel López at USAID Guatemala were kind enough to share their thoughts for this case study. Several individuals who were unable to participate in interviews, nonetheless provided important technical support during the evaluation that facilitated the development of this report, including: Baudilio López, Ashley Frost, Glenda de Paiz and Josefina Martínez.

Silvia Alayón, working with MEASURE Evaluation at the time, offered me the chance to assume leadership of this activity beginning in mid-2012, just as the idea for an evaluation process case study related to nutrition interventions was taking shape. I remain grateful for the opportunity and the groundwork she laid. Adam Bennett provided assistance with interviewing and related work in Guatemala. Skilled and timely transcription and translation services were provided by Elsie de Johnston.

In May 2012 Gustavo Angeles asked if I would like to dedicate a portion of my time to working with him on the Western Highlands Integrated Program evaluation being planned in Guatemala. That opportunity was invaluable to forming

contacts and ideas for this related report. Edgar Hidalgo and his core administrative team in Guatemala, especially Zulma Rodas and Norma Samol, worked hard to ensure that I had everything necessary for seamless and productive visits, including the one that produced the primary data for this report. All four individuals also offered their own time to participate in key informant interviews for the case study.



Children wait to be weighed and measured.

I owe a debt of gratitude to all of the study staff in Guatemala, but particularly those who helped me travel to and around study sites in San Marcos. Reyna Castillo, Jonás García, Liz Cutuc, Mauricio Flores, and Alejandro Cobar made sure my time in the field was safe, useful, and meticulously planned. Reyna, Jonás, and Liz also took part in qualitative interviews. Rosa Armas, Angélica Rivera, Regina García, José Carlos Fernandez, Mynor Hidalgo, Cesar Calderón, Jorge Meyer, Juan Carlos Coxaj, and Carlos Leal also generously took time out of their busy schedules to share their experiences for this report. Additional thanks to interviewees Amilcar Beltetón and Carlos Che, who welcomed me at INCAP at all times, and to Roberto Molina who not only participated in an interview but also patiently explained the study's complicated sample to me as often as I asked.



The study's technical director at work.

It was truly a pleasure to listen to and learn from you all, and I hope you see reflected in this report all of the important things you shared. The lessons drawn from your work hold significant potential to improve evaluation efforts around the world. Thank you!

- Tory M. Taylor

August 2014

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) under terms of Cooperative Agreement GHA-A-00-08-00003-00 and implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with Futures Group, ICF International, John Snow, Inc., Management Sciences for Health, and Tulane University. The views expressed in this publication do not necessarily reflect the views of USAID or the United States government.

SR-14-106

Acronyms



Study staff prepare for the pilot test.

5DE	Five Domains of Empowerment, a WEAI sub-index	INCAP	Instituto de Nutrición de Centroamérica y Panamá (Nutrition Institute of Central America and Panama)
AGEXPORT	Asociación Guatemalteca de Exportadores (Guatemalan Exporters' Association)	INE	Instituto Nacional de Estadística (National Statistics Institute)
ANACAFE	Asociación Nacional del Café (National Coffee Association)	IRB	Internal Review Board
DHS	Demographic and Health Survey (in Spanish, ENSMI)	LSMS	Living Standards Measurement Survey
ENCOVI	Encuesta Nacional de Condiciones de Vida (National Living Conditions Survey)	M&E	Monitoring and Evaluation
ENSMI	Encuesta Nacional de Salud Materno Infantil (in English, DHS)	MSPAS	Ministerio de Salud Pública y Asistencia Social
FTF	Feed the Future	PPP	Purchasing Power Parity
FFW	Food for Work	PSM	Propensity Score Matching
GHI	U.S. Global Health Initiative	RVC	Rural Value Chain
GIS	Geographic Information Systems	US	United States
GPI	Gender Parity Index, a WEAI sub-index	USAID	United States Agency for International Development
GPS	Global Positioning System	USD	United States Dollars
HIDN	Office of Health, Infectious Diseases and Nutrition	VGD	Vulnerable Groups Development
HIV	Human Immunodeficiency Virus	WEAI	Women's Empowerment in Agriculture Index
		WHIP	Western Highlands Integrated Program

Contents

Executive Summary	5
Introduction	7
Case Study Methods	10
Western Highlands Integrated Program Description	12
WHIP Evaluation Description.....	14
• Overview	14
• Staffing structure	15
• Sampling	16
• Questionnaires	17
• Training and survey implementation.....	19
Findings: Successes and Innovations.....	21
Findings: Challenges and Obstacles.....	30
Findings: Lessons Learned	42
Looking Ahead	49
Appendix	50



Study team members organize their materials.



A driver pauses en route to fieldwork sites.

Cover Photos:
Top right: Partial map of Guatemala.
Left top: Fieldwork staff at work.
Left bottom: Busy central market in Chichicastenango.

Executive Summary



Teams arrange tokens of appreciation for distribution to participant households.

Between July and November 2013, a massive exercise in survey data collection was completed for the Western Highlands Integrated Program (WHIP) Evaluation baseline survey in Guatemala. The program, established in the second half of 2012, integrates the projects and objectives of major agriculture, health and nutrition initiatives supported by the government of Guatemala and USAID. It covers 30 priority municipalities in five Western Highlands Departments where poverty and malnutrition have historically been concentrated. The study was commissioned by USAID and Guatemala's Ministry of Health and Social Assistance (MSPAS) to yield information on health and living conditions among residents of priority municipalities, and non-intervention areas that will serve as quasi-experimental controls, near the start of the program. Follow-up data collection for program performance and impact evaluations is planned in 2015 and 2017. The impact evaluation will provide information on the program's overall effects and the added value of the integrated model versus traditional health and nutrition interventions alone.

planning and a network of support allowed workers at all levels to respond to operational problems that arose, and an integrated system of quality assurance yielded exceptionally high-quality datasets. The baseline survey also benefited from high levels of cooperation among the agencies directly involved in supporting the survey's implementation. Echoing the program strategy itself, donor agency staff from multiple sectors focused their resources on a unified investment, which ensured a broad resource base and limited duplication of effort. Interviewers managed to achieve high response rates even as they implemented lengthy and challenging survey components, some for the first time in Guatemala.

Key informants also articulated the physical and psychological hardships that accompanied the study's implementation. These challenges exceeded what even experienced survey workers had dealt with before. The survey's burden on respondents was also exceptionally high, requiring constant, dynamic effort by the fieldwork team to stem response fatigue and prevent refusals. For the teams, maintaining safety required high levels of vigilance, especially in communities where conflicts over land use between the government and local residents were escalating. The team's initial negotiations with local authorities were often protracted and arduous, and



The study's principal investigator and statistician.

Throughout 2013, concurrent efforts were also undertaken to gather qualitative information about the experience of implementing the baseline study from program document reviews, fieldwork site visits, and semi-structured interviews with key informants. These qualitative efforts contributed to the development of this report, a case study of the WHIP Evaluation baseline. Results of the case study indicate that notable successes in fieldwork planning and implementation were linked to highly experienced leadership and a cohesive, well-trained team of fieldworkers and administrative staff. Strategic



Loading supplies into study vehicles at the project's offices.



Quiché and Mam interviewers in traditional dress.

at the household level many respondents reacted negatively to specific items in the survey. In some cases the difficulty was purely conceptual, but questions about the value of families' assets provoked fears of exploitation and occasionally jeopardized work in whole communities. The research team also encountered a number of other studies being conducted simultaneously in the same areas, and experienced delays in obtaining necessary information about program beneficiaries for the sample.

The WHIP baseline survey experience yielded important lessons about how to implement effective large-scale evaluation approaches. First, enormous resources are required for high quality program evaluation. Impact evaluations in particular are hardly comparable to other monitoring or evaluation efforts, requiring specialized knowledge and skillsets from a team of experts in every case. Planning for a large evaluation

should begin as soon as program planning does, and may take a year or longer to complete. If specific information, such as beneficiary lists, will be required from the program's implementing partners, these needs must be communicated early. The size of the study may actually limit communication across stakeholders, partners, and even individual staff members, the effects of which can be minimized by identifying and fostering the most essential communication channels. In Guatemala's Western Highlands, and undoubtedly elsewhere, enlisting broad community support is perhaps the most important component of implementation success. Lastly, the USAID Evaluation policy and flagship programs like Feed the Future (FTF) are motivating major investments in evaluation. Successful coordinated efforts not only provide high-quality data for decision making but also reinforce commitments to transparency and collaboration.

Introduction



Tortillas for sale at a market.



Highlands view.

Guatemala is the most populous country in Central America, with an estimated 15.1 million people, half of whom identify as members of one of the country's 21 Mayan ethnolinguistic groups.^{1,2} Approximately one in every two Guatemalans is under age 18, and the National Living Conditions Survey (Encuesta Nacional de Condiciones de Vida, ENCOVI) conducted in 2011 estimated that 53%

of Guatemalans were living in poverty.³ The country has the highest rate of chronic malnutrition in Latin America and the Caribbean, and the fourth highest in the world.⁴ Guatemala lies on a major fault zone and is highly vulnerable to natural disasters including earthquakes, hurricanes and volcanic eruptions.⁵ A 2012 earthquake led to dozens of deaths in communities situated between the Pacific coast and Western Highlands.⁶

The service sector is the largest component of Guatemala's GDP at 63%, followed by industry at 24%. Although half of the labor force is engaged in agriculture, this sector contributes just 13% of GDP.^{7,8} Major exports include coffee, sugar,

bananas, vegetables and textiles. Tourism and remittances remain significant and growing sources of GDP, and raw materials for biofuel production (especially sugar cane and palm oil) have assumed an increasing share of the export market since the U.S. and Europe began enacting laws mandating the use of these products in vehicle fuels.^{9,10,11} While Guatemala has experienced significant economic growth in recent years and today is considered a lower middle income country, striking ethnic and regional disparities in wealth and wellbeing remain. Ladinos continue to enjoy markedly better health, greater access to education and medical care, and higher economic status than their indigenous counterparts.



Outside the Ministry of Public Health in Guatemala City.

¹ The World Bank, data by country. <http://data.worldbank.org/country/guatemala> (February 24, 2014)

² Tory M. Taylor, John Hembling and Jane T. Bertrand. "Ethnicity and HIV Risk in Guatemala." *Ethnicity and Health* (forthcoming)

³ Instituto Nacional de Estadística de Guatemala. 2002. Encuesta Nacional de Condiciones de Vida [National Living Conditions Survey]. 2011. http://www.ine.gob.gt/np/encovi/documentos/ENCOVI_Resumen_2011.pdf (February 24, 2014)

⁴ The World Food Program. <http://www.wfp.org/countries/guatemala> (February 24, 2014)

⁵ Index Mundi. Guatemala. http://www.indexmundi.com/guatemala/natural_hazards.html (February 24, 2014)

⁶ "Guatemala earthquake kills dozens in mountain villages." BBC News. <http://www.bbc.co.uk/news/world-latin-america-20243894> (February 24, 2014)

⁷ CIA World Factbook. <https://www.cia.gov/library/publications/the-world-factbook/fields/2012.html> (February 24, 2014)

⁸ Feed the Future. <http://www.feedthefuture.gov/country/guatemala> (February 24, 2014)

⁹ World Travel and Tourism Council. Travel and Tourism Economic Impact 2012. London, UK. http://www.wttc.org/site_media/uploads/downloads/guatemala2012.pdf (February 24, 2014)

¹⁰ Multilateral Investment Fund, a Member of the IDB Group. Washington, DC. "Remittances to Latin America and the Caribbean in 2012: Differing Behavior among Sub-regions." <http://idbdocs.iadb.org/wsdocs/getDocument.aspx?DOCNUM=37735715> (February 24, 2014)

¹¹ Elisabeth Rosenthal, "As Biofuel Demand Grows, So Do Guatemala's Hunger Pangs," *New York Times*, sec. A, January 6, 2013. http://www.nytimes.com/2013/01/06/science/earth/in-fields-and-markets-guatemalans-feel-squeeze-of-biofuel-demand.html?pagewanted=all&_r=2& (February 24, 2014)



Editors at work.

High rates of chronic malnutrition, especially in indigenous communities, have been particularly impervious to change. An estimated 1 in 2 children under five suffers from chronic malnutrition in the country as a whole, the highest prevalence in the western hemisphere and among the highest in the world.¹² In 2008-09, results from the Demographic and Health Survey (DHS) in Guatemala demonstrated a thirty-point difference in the prevalence of stunting between indigenous and ladino children (66% and 36%, respectively).¹³

The country's history has also been marked by serious ongoing political and ethnic conflict. Between 1960 and 1996, civil war between the government and insurgents resulted in the displacement of more than a million people. As many as 200,000 others died or went missing.¹⁴ A Historical Clarification Commission sponsored by the United Nations (UN) in 1999 concluded that the state was responsible for 93% of human rights violations committed during the war, and guerrilla forces for 3%.¹⁴ An estimated 83% of the victims were indigenous Mayans.¹⁴ Throughout the 80s and 90s, much of the worst violence was focused on the Mayan communities of the rural Western Highlands.¹⁴ In May 2013, former president Efraín Ríos Montt was convicted of genocide for his actions during the war; portions of the court proceedings were officially annulled 10 days later on a technicality. The trial is scheduled to resume in 2015.¹⁵

The United States Agency for International Development (USAID) works with the government of Guatemala to implement a development strategy focused on improving health, education, food security, democracy and governance, economic growth, environmental management and disaster response in the country. The Western Highlands Integrated Program (WHIP) is one of the Mission's largest activities. Several Offices within USAID and the Ministry of Health and Social Assistance in Guatemala (Ministerio de Salud Pública y Asistencia Social, or MSPAS) provide combined support for the program, which began in late 2012 and integrates technical support for smallholder agricultural producers with health and nutrition interventions in 30 priority municipalities.



School in San Marcos with damage from the 2012 earthquake; sign reads, "School at high risk."

¹² World Health Organization (WH). Prevalence of stunting (moderate and severe). Data by country. <http://data.un.org/Data.aspx?d=SOWC&f=inID%3A106> (March 19, 2014)

¹³ Encuesta Nacional de Salud Materno Infantil 2008-09 (ENSMI 2008-09). Ministerio de Salud Pública; Universidad del Valle; Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). <http://stacks.cdc.gov/view/cdc/8277> (February 24, 2014)

¹⁴ Guatemala: Memory of Silence. Tz'Inil Na'tab'al. Report of the Commission for Historical Clarification. Conclusions and Recommendations. http://www.aaas.org/sites/default/files/migrate/uploads/mos_en.pdf (February 24, 2014)

¹⁵ Guatemala Rios Montt genocide trial to resume in 2015. BBC News, November 6, 2013. <http://www.bbc.co.uk/news/world-latin-america-24833642> (February 24, 2014)



Along with investing in WHIP, USAID and MSPAS decided to allocate funding for a rigorous evaluation of the program: The Western Highlands Integrated Program (WHIP) Evaluation. This decision aligns with the USAID Evaluation Policy and supports the FTF Learning Agenda, which notes that “many gaps in knowledge still remain about how agriculture and nutrition programs can work together,” and “rigorous evaluations of agriculture programs are urgently needed to

better understand the real potential of agriculture to improve nutrition.”^{16,17,18} The evaluation is designed to examine the integrated program’s performance and overall contribution to changes in outcomes including the percentage of people living in poverty and the prevalence of stunting in children younger than five. In addition, analyses will compare participants receiving support from both the agriculture and nutrition/health components to those exposed to nutrition and health interventions alone. To our knowledge this is the first impact evaluation designed to identify the value added to a traditional health and nutrition program model through targeted agricultural support.

Fieldwork for the evaluation’s baseline phase was conducted between July and November 2013 by an experienced research team led by MEASURE Evaluation. Interviewers and anthropometrists completed more than 24,000 interviews with respondents in over 6,300 households, as well as with local authorities and other leaders in 309 communities and at 266 health facilities. Following fieldwork, a qualitative case study was completed using document reviews and interviews with individuals serving in key implementing roles for the WHIP Evaluation baseline. The case study was designed to identify lessons learned from the WHIP Evaluation baseline that may be applicable to evaluation practice elsewhere. Case study activities culminated in the production of this report. We hope readers from a variety of backgrounds will find it helpful for making decisions about investing in program evaluation and understanding what to expect from the process.



Billboard in Guatemala City promoting a government progress report, “Fewer children with malnutrition.”

¹⁶ Evaluation: Learning from Experience. USAID Evaluation Policy. January 2011, Washington, D.C. <http://www.usaid.gov/sites/default/files/documents/1868/USAIDEvaluationPolicy.pdf> (March 29, 2014)

¹⁷ Douglas Taren and Halima Alaofè. Feed the Future Learning Agenda Literature Review: Improved Nutrition and Diet Quality. Prepared by Feed the Future FEEDBACK for the U.S. Agency for International Development. https://www.securenutritionplatform.org/Documents/usaid_2013_feed-the-future-learning-agenda.pdf?Mobile=1. (February 24, 2014)

¹⁸ Feed the Future Learning Agenda. <http://agrilinks.org/sites/default/files/resource/files/Feed%20the%20Future%20Learning%20Agenda.pdf> (February 24, 2014)

Case Study Methods



Iron supplements are packaged for distribution to anemia-affected households.

This report is the product of qualitative research activities conducted separately by MEASURE Evaluation as part of a case study of the WHIP Evaluation, commissioned in 2012 by technical advisors in the USAID Office of Health, Infectious Diseases and Nutrition (HIDN). The case study was designed to supplement technical study reports, documenting the process of planning and carrying out baseline activities in Guatemala in order to yield lessons learned. It draws from the subjective experiences of 21 purposively selected key informants and several others who agreed to review and comment on draft sections of the document. Interviewees were selected because they served in a decision-making capacity or other key fieldwork role. This included the study's principal investigator, technical director, sampling statistician, fieldwork coordinator, field supervisors and technical advisors at USAID, among others.

Preparation for this case study began with background document reviews in the second half of 2012. A case study report outline was drafted in early 2013, and work began later that

year. In July 2013 the author traveled with baseline study personnel to observe fieldwork at several locations in the Department of San Marcos. Preparation continued following that trip with the creation of a semi-structured in-depth interview guide and preliminary interview participant list. The interviewee list was finalized in consultation with USAID staff, the study's principal investigator and the technical director in Guatemala. Interview dates were established in November as baseline study fieldwork drew to a close, and interviews were conducted December 5-13 inclusive in locations around Guatemala City, including private meeting space adjacent to the project's INCAP offices and at USAID. All interviews were conducted individually and in person, and audiotaped using Olympus DS-2 digital recorders. The average interview took 28 minutes to complete, with the shortest lasting approximately 15 minutes and the longest lasting 57 minutes.

Interviews were semi-structured and designed to capture participants' perspectives on key aspects of study implementation, eliciting information on experiences beyond those typically reflected in quantitative research reports. The standard interview asked participants to identify and discuss aspects of the research process that went well, those that were problematic, systems or supports they saw as crucial to the success of the study, and planning or implementation decisions that might benefit from a different approach in future study rounds or similar research projects.

Digital audio files of the interviews were forwarded to a professional translation and transcription service provider based in Guatemala City, who produced a complete record of each interview in the original language. Of 21 total interviews, 19 were conducted in Spanish and two in English. Following transcription, interview text was reviewed and annotated to identify common themes. Thematic analysis of these interviews forms the primary basis for this report.



Teams prepare for work.



Natural beauty in the Western Highlands.



Readying iron supplements and other supplies.

Direct quotations taken from interviews conducted in Spanish were translated into English by the author for this report. Photographs reproduced here were taken by contributors Liz Cutuc, Jonás García or the author and are used with permission. No identifiable likenesses of research subjects have been included, and we have endeavored to remove place names and other geographic identifiers where the location indicated is smaller than a municipality.

This case study was classed as a quality improvement activity by the Human Research Protection Program at Tulane University, and as such was not required to undergo formal institutional review. All interviewees were over the age of 18 years and provided verbal consent before participating in interviews. The author of this report also served on the study team, providing technical support for planning and implementation from May 2012 onward. This dual role facilitated the process of conducting a qualitative review of the study, through insights that accumulated naturally in the course of work. A second, independent contributor who was uninvolved in the evaluation also assisted with gathering information for the case study.

Table 1: GHI Core Principles

Focusing on women, girls, and gender equity
Country ownership
Health systems strengthening
Promoting global health partnerships
Integration
Research & innovation
Improve metrics, monitoring & evaluation



A young girl allows the anthropometrist to check her weight.

Western Highlands Integrated Program Description



White and red beans for sale.

The current Cooperative Agreement between the Government of Guatemala and USAID reflects three overarching development objectives:

1. Increased security and justice for citizens
2. Increased economic growth and increased social development in the Western Highlands
3. Improved natural resource management to mitigate climate change impacts

The Western Highlands Integrated Program contributes to Objective 2, focusing simultaneously on expanding opportunities to generate income and strengthening food security and access to health services among the country's most vulnerable groups, especially in indigenous communities. The program officially began in late 2012, and by mid-2013 was operating at scale. It aims for close coordination between implementing organizations and regional, departmental and municipal governments, as well as private sector development

"The idea behind WHIP was this: to combine our efforts within the Agency in this region where severe poverty is concentrated, so that the different sectors working together might generate more development over a five-year period." (M&E Advisor, USAID)

partners. Operating in 30 priority municipalities in five departments (Quiché, Huehuetenango, Quetzaltenango, San Marcos and Totonicapán), with a total target population estimated to be just over 1.5 million in 2013, WHIP coordinates the roles of four national health and food security initiatives whose objectives are complementary:¹⁹

1) The Rural Value Chains (RVC) program works to increase the economic power of smallholder producers of coffee, other horticultural and artisanal products through the provision of technical support to affiliated producers' cooperatives. Two organizations currently implement the RVC program: the *Asociación Nacional del Café* (ANACAFE) and *Asociación Guatemalteca de Exportadores* (AGEXPORT). The program engages with smallholder farmers by enrolling producers' cooperatives directly, increasing their member access to markets by negotiating contracts with merchants and helping to improving yields and product quality through expanded access to farming technology and productive inputs such as seeds and tools. The RVC program directly enrolls only a small percentage of families living in WHIP municipalities, estimated at slightly more than 8,600 households in 2013, or around 4%. However, program activities are expected to exert community-level effects resulting in improved economic and health conditions overall, notably for families living in close proximity to member-producer households.



Woven baskets for sale at a market.

¹⁹ Instituto Nacional de Estadística (INE). Population Projections by Municipality: 2008-2020. <http://web.archive.org/web/20131113144755/http://www.ine.gob.gt/np/poblacion/index.htm> (March 5, 2014)



Girls run to wash their bowls after having lunch at a school in San Marcos.

2) **The Nutri-Salud program** is designed to increase access to quality maternal and child health care and reduce malnutrition by promoting a package of essential health and nutrition actions for women during pregnancy and for children in their first two years of life.²⁰ The Nutri-Salud program implements community campaigns to encourage the use of maternal and child health services as well as better feeding practices and diets for reproductive-age women and children under five. Further, it works to establish permanent health centers in rural underserved areas of the Western Highlands, build local capacity to staff these facilities, and address barriers to health service seeking through effective community engagement and the provision of culturally competent care. Nutri-Salud includes the entire population of the 30 WHIP municipalities within its area of influence.

3) **Title II/PL480** provides for the direct donation of U.S. agricultural commodities for emergency relief and development programs. The Title II/PL480 program in Guatemala is one of the western hemisphere's largest, using 10,500 metric tons of donated commodities in food-for-work (FFW) and vulnerable group development (VGD) programming in 2012.²¹ The program works in selected WHIP municipalities and other areas of the country with a focus on pregnant and lactating women and children under two years of age.²²

4) **The local governance project.** As part of the integrated program, USAID also provides ongoing support for projects designed to "increase the capacity of municipal governments to raise revenue, respond to citizen concerns related to violence and security, food insecurity and global climate change

as well as to manage public resources in a participatory and financially secure manner."²³ The local governance program operates in approximately one third of WHIP municipalities.

WHIP is part of two major U.S. Presidential Initiatives: Feed the Future (FTF) and the U.S. Global Health Initiative (GHI). FTF was launched in 2010, with the goal of leveraging USG expertise and investment in the agriculture and health sectors in 19 focus countries to reduce global poverty and malnutrition.²⁴ The results framework for the initiative emphasizes linkages between intersectoral objectives including agricultural sector growth, expanded access to markets and trade, and reductions in social inequality and malnutrition.²⁵ GHI began in 2009 as a way to harness the skills of US government agencies to address the world's most pressing health challenges, and is governed by a core set of operating principles in each of the 34 countries where it works (see Table 1).

Although the integrated program incorporates all four of the components listed above, the WHIP evaluation is specifically designed to measure the performance and impact of the first two: RVC and Nutri-Salud. The Title II/PL480 program and local governance projects are only indirectly captured in the study, principally because their implementation areas and timetables were ultimately not aligned with the other two projects. In planning the evaluation, USAID also opted not to further complicate the already complex study design by trying to enumerate the individual and combined effects of more than two WHIP components. Such an approach could exponentially increase the number of sample domains and control groups in the study, and increase costs accordingly.

²⁰ U.S. Global Health Initiative: Where We Work. Guatemala. <http://www.ghi.gov/whereWeWork/profiles/Guatemala.html#Uxd1Ztix7Dc> (March 5, 2014)

²¹ USAID Fiscal Year 2012 Food For Peace Fact Sheet. <http://www.usaid.gov/what-we-do/agriculture-and-food-security/food-assistance/quick-facts/fiscal-year-2012-food-peace> (March 5, 2014)

²² USAID Office of Food for Peace. Guatemala Bellmon Estimation. Prepared by Fintrac, Inc. for the U.S. Agency for International Development. October 2011. <http://www.usaidbest.org/docs/GuatemalaBellmon2011.pdf> (February 24, 2014)

²³ FedBizOpps.Gov. USAID Solicitation Number SOL-520-13-000006. https://www.fbo.gov/index?s=opportunity&mode=form&id=51e1547cb8135c17c12650861de4357e&tab=core&_cview=1 (March 5, 2014)

²⁴ Feed the Future: The US Government's Global Hunger and Good Security Initiative. <http://www.feedthefuture.gov/> (February 24, 2014)

²⁵ Feed the Future Results Framework. http://feedthefuture.gov/sites/default/files/resource/files/ftf_results_framework_2013.pdf (March 5, 2014)

WHIP Evaluation Description



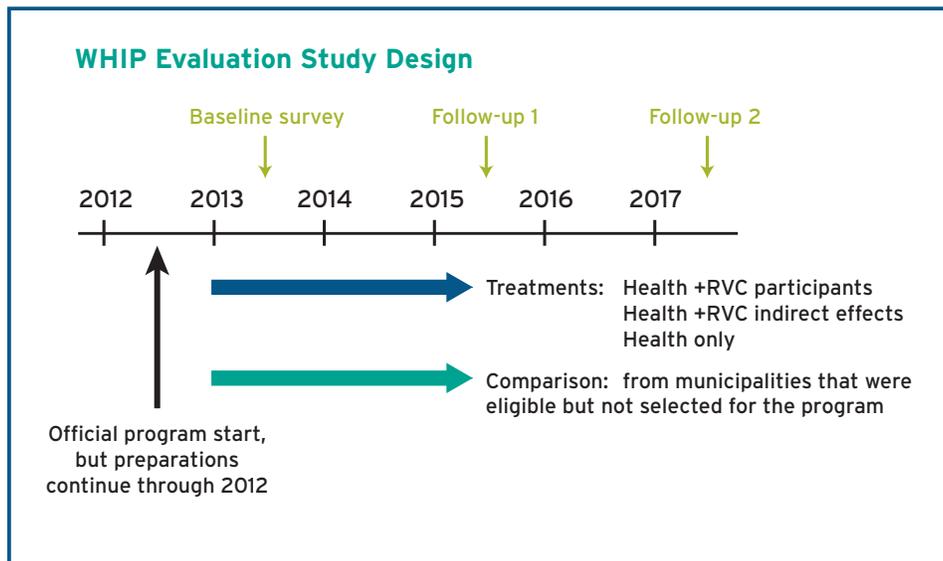
Fieldwork staff take a moment to share a laugh.

OVERVIEW

The WHIP Evaluation is a combined performance and impact evaluation being conducted in three survey rounds; fieldwork completed in 2013 constituted the baseline phase. Follow-up data collection will take place in 2015 and 2017. According to USAID policy, a performance evaluation addresses whether expected changes coincide with program implementation in a target area or group, *e.g. are performance targets being met?* By contrast, impact evaluations include a rigorously defined control group whose members have little to no exposure to the intervention under study. Impact evaluations offer credible evidence for whether changes observed over time in a target community or group are due to the intervention: *e.g. did the program make a difference on selected outcomes? If so, how much?*

The performance evaluation is based on results from each survey round in the program intervention area. Many indicators used in the study are aligned with the objectives of affiliated efforts including FTF, GHI, and the Zero Hunger Pact and Plan, a set of coordinated national policies and programs aimed at significantly reducing malnutrition in Guatemala during the current presidential tenure.^{26,27,28} The impact evaluation will examine program-attributable effects on chronic malnutrition, poverty prevalence, and other key indicators. Beyond comparing results in program areas with those in similar non-program areas, the impact evaluation will explore the effects of the integrated model relative to its health and nutrition component alone, and delineate RVC program effects on participants and others in their communities.

Data entry staff at work.



²⁶ Feed the Future Indicator Handbook: Definition Sheets. http://feedthefuture.gov/sites/default/files/resource/files/ftf_handbook_indicators_sept2013_2_0.pdf (March 5, 2014)
²⁷ The United States Global Health Initiative Guatemala Strategy. Revised December 11, 2010. <http://www.ghi.gov/whereWeWork/docs/GuatemalaStrategy.pdf> (March 5, 2014)
²⁸ Government of Guatemala. Secretary of Health and Nutrition. <http://www.sesan.gob.gt/index.php/ph0> (March 5, 2014)



The study's principal investigator looks over study documents.

"The idea was that MEASURE Evaluation was going to create a technical group led by [the technical director], and he was going to organize a team for the implementation of the survey. But this team needed a base from which to operate, and an institutional infrastructure for accounting, and procurement, and logistics...and the agreement was that INCAP was going to provide that base..."
(Principal Investigator)

The WHIP Evaluation is specifically designed to respond to the following research questions:

1. What changes took place in sectors where one or major components of the integrated program (RVC, Nutri-Salud) were implemented?
2. How much of the change in program sectors is attributable to the integrated program?
3. What was the impact of the intergrated program on families who participated in the RVC intervention, and on those in RVC sectors who did not participate directly?
4. What was the impact of the Nutri-Salud component in sectors without any RVC participants?
5. Was the integrated program more effective than Nutri-Salud alone at improving key population-level outcomes?

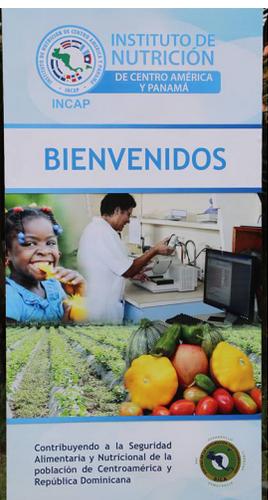
STAFFING STRUCTURE

MEASURE Evaluation, the Global Health Bureau's primary vehicle for supporting improvements in monitoring and evaluation

in population, health and nutrition worldwide, was selected in late 2011 to design the WHIP Evaluation and direct the implementation of its baseline survey. Started 17 years ago, MEASURE Evaluation provides services and technical assistance to USAID, country Missions and counterparts in more than 40 nations. Activities are designed to help stakeholders identify data needs, collect and analyze technically sound data, and use data for effective health decision making. The research team was led by a Principal Investigator (PI) from the University of North Carolina at Chapel Hill with substantial technical and managerial expertise in program evaluation. The PI was assisted by a technical director in Guatemala – a physician and survey implementation expert with extensive experience directing the country's DHS and other population surveys. Technical support was also provided by a monitoring and evaluation specialist from Tulane University, as well as a sampling statistician and regional team of consultants with complementary methodological expertise.

USAID Guatemala allotted USD 3,549,822 for the baseline study, approximately USD 2,296,367 or 65% of which was set aside for a subcontract with a local organization to provide logistics, procurement, and fieldwork staffing support. In October 2012 MEASURE Evaluation advertised a Request for Proposals seeking an organization in Guatemala to provide these services. The study's leadership team identified the proposal submitted by the Institute of Nutrition of Central America and Panama (INCAP) as the preferred candidate, based on the proposal's technical and financial merits. An 11-month agreement went into effect on April 15, 2013 between INCAP and the Carolina Population Center on behalf of MEASURE Evaluation. Guatemala's Ministry of Health and Social Assistance (MSPAS) provided another USD 25,000 towards the study, and additionally donated office space and supplies in Guatemala City for use during training and for project administration.

For the baseline survey, the study's technical director in Guatemala worked under contract with UNC, as did the sampling statistician and other technical advisors who provided



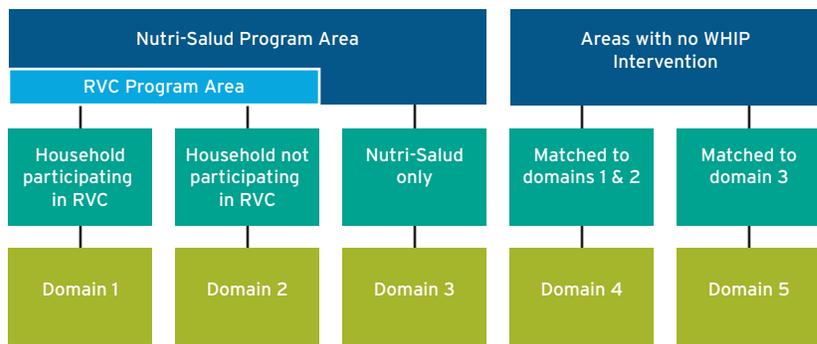
Welcome sign at the entrance to INCAP, Guatemala City.



A team member reviews questionnaires.

support during design and implementation. At INCAP, a small core team of administrators oversaw hiring, purchasing, accounting and other matters related to the local subcontract. Interviewers, anthropometrists, office staff, data entry staff, programmers, coordinators and supervisors were all hired by INCAP. The PI and other MEASURE Evaluation staff working on the study visited Guatemala every few months from the planning stages through preliminary results presentations in early 2014. Because of the need to produce FTF indicator values for USAID’s Bureau of Food Security (BFS) immediately following data collection, Westat through its FEEDBACK project offered to calculate the Women’s Empowerment in Agriculture Index (WEAI) for the WHIP baseline.²⁹

WHIP Evaluation Baseline Sample



0.05) and 80% power ($\beta = 0.20$). These changes reflect five year targets established by USAID Guatemala. Baseline estimates for the indicators used to determine sample size were obtained using data for the 30 WHIP municipalities from the 2008-09 DHS and the 2011 National Living Conditions Survey (ENCOVI), respectively. To facilitate investigations of WHIP impact and the effect of adding the RVC program in Nutri-Salud areas, the baseline study sample was constructed in five distinct parts: three representative domains in WHIP municipalities and two matched control groups.

The sample’s domain one represents RVC program participant households. Implementing partners ANACAFE and AGEXPORT provided beneficiary lists for the sample frame in this domain, and a finite population correction was applied during sample size calculations. Domain two is representative of non-RVC participant residents of the same census sectors as RVC participants, for whom the RVC program is expected to show indirect effects. Sector selection in domains one and two was stratified, with probability proportional to the estimated number of RVC program participants or



The study’s technical director leads a planning meeting.

SAMPLING

Sampling was conducted in stages with geographic census sectors selected first, and subsequently, households within each sector. A total minimum sample size of 7,068 households was established based on desired ability to detect changes in: 1) stunting prevalence from 67.1% to 56.9%, and 2) the percentage of people living in extreme poverty from 17.9% to 14.4%, with 95% significance ($\alpha =$

²⁹ Feed the Future website. <http://agrilinks.org/activity/ftf-feedback-0>



Study staff at the project's closing meeting and celebration.



A woman is weighed as part of the anthropometry module.

non-participants, respectively. Domain three is representative of the residents of census sectors in WHIP municipalities without any RVC program participants. Within the 30 WHIP municipalities, an estimated 4% of households are RVC participants and around 27% fall into domain two (where indirect effects from the RVC program are anticipated). Colloquially, the researchers termed domain three “*solo salud*,” or “health [program] only.” The remaining 69% of households in the WHIP area are included in this third domain, where the RVC program is expected to exert little to no influence.

Control groups in the sample, called group four and group five, were developed using a statistical technique called Propensity Score Matching (PSM). Data from Guatemala’s 2002

national population census and 2004 agricultural census were used to identify 78 census sectors similar on selected demographic and agricultural characteristics to the sectors in domains one and two. These 78 matched sectors constituted the sample frame for group four. Another thirty census sectors were identified as similar to the sectors in domain three, and served as the sample frame for group 5. Household selection in the control groups was conducted randomly using updated area maps. Sampling included twenty households per sector in group four and thirty households per sector in group five.

Taken together, baseline results for domains one, two and three will be reported as WHIP performance outcomes for 2013. The two control groups will be used to assess the program’s impact by comparing changes over time in these

“We arrived in some communities where maybe only 10 people, if that many, spoke Spanish...”
(Community and Services Coordinator)

households to changes in their matched intervention domains. Households in domains one and two, reflecting the direct and indirect effects of the RVC program, will also be compared to households in the health program-only sectors that comprise domain three.

QUESTIONNAIRES

The WHIP evaluation includes a set of four instruments to be completed with eligible respondents in households in the sample, plus community and health facility-level questionnaires. A degree of overlap exists between respondents in the household interview set; women in particular are eligible to respond to multiple questionnaires. Survey instruments administered to individuals in households selected for the study include: household, women’s, a Women’s Empowerment in Agriculture Index module (hereafter, ‘empowerment’), and expenses and consumption. In addition, the baseline included community-level surveys completed with local leaders and facility-level surveys conducted with health facility directors or other key personnel. At the time of this report, preliminary results had been released to USAID for some household survey indicators, but data processing for the community and facility surveys was still underway.

At least five versions of the household questionnaires were developed between January and June 2013, with a final round of revisions made following the pilot test in May. The household and women’s instruments are similar to DHS questionnaires used previously in Guatemala, but also included new content (such as women’s knowledge of key obstetric,



Expenses and Consumption Questionnaires, fresh from the printer.



Fieldworkers participate in a debriefing with the editing coordinator

postpartum, and neonatal risk signs) and omitted some standard DHS questions (such as those designed to assess knowledge about the human immunodeficiency virus, HIV). In addition, while the DHS included one randomly-selected woman of reproductive age in each study household, the WHIP Evaluation baseline included all women ages 15-49 in interviews. Anthropometry and serum hemoglobin measurements were recorded on a standalone form in order to facilitate interviewing efficiency.

The empowerment questionnaire was adapted directly from a template available from FTF as part of comprehensive program survey guidance for implementing partners.³⁰ With USAID approval, the empowerment questionnaire was implemented in the WHIP baseline with a random subsample of approximately 2,000 households in WHIP program sectors, and was not administered in households located in non-program (control group) sectors. The expenses and consumption questionnaire used in the baseline study was adapted from the Living Standards Measurement (LSMS) Survey applied in recent years as part of the ENCOVI in Guatemala.

Questions were read aloud to respondents from paper-based questionnaires. Questionnaires and related fieldwork materials were developed and tested in Spanish. Human subjects approval was obtained through the review board at MSPAS

in Guatemala and the Institutional Review Board at the UNC Chapel Hill Office of Human Research Ethics in the U.S. Although a number of interviewers spoke one or more Mayan languages fluently, in many areas during fieldwork additional Mayan language speakers were hired on site to provide translation services. The translators worked side-by-side with interviewers, orally translating survey questions into the local language during survey administration. Location unit data were recorded on survey forms using the national common coding system so that the household, health facility and community-level datasets may be linked for analysis.

TRAINING AND SURVEY IMPLEMENTATION

Workers carried out a four-week pilot survey in May 2012 in selected municipalities of three Departments: Quetzaltenango, Chimaltenango, and Sololá. A team of 27 people organized into three teams of nine conducted interviews in approximately 250 households over a two-week period; results of this exercise were used to guide changes to fieldwork procedures and questionnaires. By June final versions of all fieldwork materials had been prepared, and five weeks of training began in Guatemala City for potential interviewers and anthropometrists, including those who had worked on the pilot. Training included instruction and practice in interviewing techniques, the use of household location maps, and fieldwork procedures and forms, among other topics.

Measurement standardization training and assessment were conducted separately with anthropometrist candidates. Fourteen individuals underwent training in anthropometry and the use of Hemocue brand portable hemoglobinometers. Thirteen trainees successfully passed the standardization assessment at the end of the course, and seven were ultimately selected to serve as anthropometrists for the baseline study. In all, 89 individuals would complete interviewer training and be assigned to teams collecting household data for the study. Sixteen different manuals designed for study personnel detailed instructions and standards on everything from

³⁰ Feed the Future M&E Guidance Series. Volume 8: Population-Based Survey Instrument with Revised WEAI module. October 2012. http://feedthefuture.gov/sites/default/files/resource/files/ftf_vol8_populationbasedsurveyinstrument_oct2012.pdf (March 6, 2014)



Administrative staff in the project's offices at the Ministry of Health.

Respondent criteria and survey content by questionnaire

Survey	Eligible Respondent(s)	Content Areas
Household	Adult head of household	Household roster, living conditions, food insecurity, membership in agricultural co-operatives, other assistance programs, food production for personal use
Women's	All female residents ages 15-49 years. Sections on children's health and nutrition ask about respondent's three youngest living children age 0-59 months	Demographics, birth history, family planning, antenatal and postnatal care, knowledge about reproductive risks, children's health and nutrition, women's dietary diversity
Anthropometry*	All women (residents or visitors) ages 15-49 and all children (residents or visitors) ages 0-59 months	Weight Height
Serum hemoglobin*	All women (residents or visitors) ages 15-49 and all children (residents or visitors) ages 6-59 months	Current pregnancy and lactation status Serum hemoglobin
Expenses and Consumption	Adult resident most knowledgeable about household assets and consumption patterns, one per household	Household expenses; transport, fuel and energy costs; items for cooking or furnishing the home; purchases related to cleaning and personal hygiene; clothing, recreation, and personal services, travel and tourism; medical expenses, schooling, all other expenditures
Empowerment	Self-identified primary adult male and female decision-makers in dual adult households; primary adult female decision-maker in female-only households (adult male-only households are excluded)	Role in decision-making, access to assets and capital, access to credit, leadership and community influence, motivation for decision-making, time allocation
Community	Self- and community-identified community leaders	Community infrastructure, public transport, water and sanitation services, social development programming, schools, health services, recent natural disasters, markets and commerce
Health Facility	Facility director or key personnel	Services offered, equipment and supply inventory, medication inventory

*Included in supplement to Household questionnaire



Data entry staff at work.

the use of the global positioning systems (GPS) to personnel policies to data entry.

WHIP teams were classified into one of two groups according to the type of instruments they administered: household and community/facility. Each of the seven household teams included 12 people: one team supervisor, two editors, one anthropometrist, six interviewers and two interviewer/drivers. In addition, five general supervisors rotated between the teams in field. The 17 community/facility workers were organized into five teams of three (two interviewers and one supervisor-cartographer) plus one general supervisor and a driver-interviewer. In addition to securing



A member of the central office team participates in a planning meeting

local authorities' permission for study personnel to work in the sectors included in the sample for each study community, the community teams performed four major tasks:

- Update household maps provided by the National Statistics Institute (INE), and where applicable, verify RVC program participants' locations of residence
- Conduct community interviews with local authorities and opinion leaders (including the enumeration of local health facilities)
- Conduct health facility surveys
- Visit local markets and stores to record food prices and weight equivalents (for use by household teams in the expenses and consumption survey)

Both household and community teams worked in commissions of approximately 26 consecutive workdays, with 2 to 3 days off between commissions. Between July and November

The WHIP Evaluation baseline study by the numbers

18 months of planning
139 fieldwork and data processing personnel
24,366 household interviews completed
212,276 kilometers traveled
46,704 liters of fuel used
7,813 hours of data entry

2013, the household teams completed five separate commissions, or approximately 120 working days total (the final commission was only two weeks long). The baseline study included interviews in 309 communities in 54 municipalities of five departments. Back in the project's Guatemala City offices, 23 data entry personnel and 10 central editors worked in two daily shifts to process the incoming questionnaires. Questionnaires were 100% double entered. Data entry also included ancillary fieldwork forms, such as the weights and price equivalency sheets used to calculate standard measures for consumable items in each study community.

Early in the study's development, USAID approved a list of program performance indicators to be measured through WHIP evaluation data collection (see Appendix). In total, this list includes 90 items addressing individual, household, health facility and community-level results across content domains including health status, service utilization, and access to essential infrastructure. Many of the WHIP Evaluation indicators also fulfill monitoring and reporting requirements for component programs such as GHI and FTF. Indeed, all 13 indicators required for FTF program monitoring and evaluation at the household level were incorporated into the WHIP Evaluation questionnaires.³¹ Source questions for non-FTF indicators were carefully matched to the DHS and other national and international standards to ensure comparability wherever possible.

³¹ Feed the Future M&E Guidance Series. Volume 8: Population-Based Survey Instrument for Feed the Future Zone of Influence Indicators with Revised WEAI Module. October 2012 version. http://feedthefuture.gov/sites/default/files/resource/files/ftf_vol8_populationbasedsurveyinstrument_oct2012.pdf

Findings: Success and Innovations



Study staff prepare to leave a town for rural communities.

Skilled leadership and experienced fieldworkers facilitated implementation success. Time and time again, when asked to name factors that were crucial to the study's successful implementation, individuals on the fieldwork team named the technical director. They also referred often to their own and their colleagues' experience with the DHS and other studies in Guatemala. Although the DHS was different from the WHIP Evaluation baseline in a number of significant ways, it provided valuable experience with standardized interviewing and life in the field for the team members, and introduced many of them to one another years ago. This shared familiarity with individuals' particular strengths proved to have important planning advantages. It helped to customize role assignments and training priorities early in the process, informed team configuration decisions, and let workers approach a difficult job with confidence in the study's leadership, trust in their own abilities, and a well-developed support system. Members of the local team displayed high levels of mutual respect and motivation, recognized this as driven

by the director's example, and considered it essential to their success during study implementation. Team supervisors spoke repeatedly of the level of dedication that interviewers and anthropometrists exemplified in their work, and credited the technical director and general supervisors with a management approach that was adept, consistent, and unusually responsive.



Supervisors ask local residents for directions to a study community.

"...my work was hard, but it was nothing compared to what the interviewers did, through sun and rain, and keeping the stress in check after so many hours in a household. They did it, you know? It's a really big achievement." (Fieldwork Coordinator)

"...the people we had were very professional, they had a lot of experience, and I think they were most important factor in getting the job done." (General Supervisor)

"I have a lot of experience with fieldwork for other projects, and the WHIP baseline was really extraordinary...it represented tremendous growth for me both personally and professionally." (General Supervisor)

"[The technical director] was constantly supportive. He was my example for how to act with [my own staff]...I put forth my best effort...to honor his confidence in my abilities." (Central Editing Coordinator)

"[The technical director's] expertise and experience doing this kind of survey in the country is unparalleled. I think he is the person doing this kind of work here, so it was very important for us to have him on our team." (Principal Investigator)

"I think since we had a precedent—even though this study is different from the DHS, it still meant that we had a base. I feel like that made it a little easier..." (Central Editing Coordinator)

"The technical capacity of the team led by [the principal investigator] was really impressive, and [they were] really responsive, and it certainly make our job easier." (Mission Economist, USAID)

"One thing that I really valued was the team members' dedication." (Technical Director)



Wet weather during fieldwork calls for ponchos and boots.

Combining performance and impact evaluations in one design prevented duplication of effort. USAID staff identified the need to be responsive to reporting mandates as a major factor motivating investment in this evaluation. Data from the WHIP baseline will not only contribute to conclusions about program impact over a five year period, but also supply two technical offices at USAID Guatemala (Health and Education, and Economic Development) with performance monitoring information to meet certain annual reporting needs in 2013, 2015, and 2017—eliminating the need to identify and fund a separate mechanism for obtaining these performance indicator estimates. Expanding the level of coordination between USAID Offices involved in the study in future data collection rounds could further expand this major benefit.

“We were trying to figure out a way to get the most bang for our buck...we came up with the idea to do an impact evaluation as part of the population-based survey that BFS was requiring. So we merged the two because they’d be measuring a lot of the same things, and we just needed to add different domains in order to make it an impact evaluation.” (Mission Economist, USAID)

Persistence and innovation in adapting team configurations and materials helped to resolve significant logistical challenges, both in the field and at the office. Basic team configurations for fieldwork and in the project’s central offices were initially largely modeled on previous experience with the DHS and other household-type surveys in Guatemala. Study staff quickly realized that the WHIP baseline had unique qualities necessitating a different arrangement. The high demand for data verification led to two editors being assigned to each of the 7 fieldwork teams instead of one. Anthropometry and anemia testing were removed from the household questionnaire and packaged in a standalone

form, allowing anthropometrists to work separately from interviewers in the same household and to return at different times for follow-up appointments. In Guatemala City, supervisors hired additional staff and instituted a move from one to two work shifts seven days a week, more than doubling the team’s original capacity for central editing and data entry. Overall, adaptability and innovation on the part of the research team allowed a set of lengthy and highly specialized survey instruments to be fielded without any major mistakes.

“We finished on time because we went to the editing and data entry staff and said, look here’s the goal, the challenge, and no one objected to coming in on Saturdays and Sundays. That’s not something that’s easy to find, I mean, these are people who aren’t only there to draw a salary but also to do the job well.” (Technical Director)

“The amount of work we had was more than we planned for... there was pressure from the field, and pressure from the data entry side, because we were like a filter between the two. We had to give feedback to the fieldworkers, so they would know how well forms were being filled out—inconsistencies and everything—and from the other side, data entry. If we didn’t keep the forms coming fast enough they were basically stuck waiting. But like I said, the team...supported me a lot and if we needed something done sooner than we’d planned, they got it done, by just focusing and chipping away at it.” (Editing Coordinator)



The technical director addresses his staff at the project’s closing meeting and celebration.



Pick-up trucks provide a means of public transportation in many highlands communities.

"[Putting anthropometry on a separate form] really sped things up. The anthropometrists calculated that this alone saved two days in every sector. That is, the teams spent three or four days in a sector and under the old system, it would have been five or six days." (Technical Director)

Careful preparation and an integrated system of checks and verification produced data that required very little cleaning at its endpoint. Fieldworker training included instruction in standardized numeral recording, practice following skip patterns on a paper instrument, and numerous other strategies for minimizing interviewer error during questionnaire completion. Information recorded on the questionnaires was checked twice in the field: once by a field editor and then again by a supervisor. The standard review process included a comprehensive series of range and consistency checks; field editors also checked questionnaires for completeness and verified the legibility of all written responses. Conducting the initial data review before leaving a community provided teams with the opportunity to resolve questions through a repeat visit to the household if necessary; other questions were resolved through consultation between the interviewer and editor or supervisor.



A woman is tested for anemia.

Once completed questionnaires were received in the project's central offices, a second review process was initiated, with questionnaires checked and signed off on by another editor and finally, the central editing supervisor. Debriefing meetings held with each interview team at the close of every commission included the dissemination of interviewer-specific reports detailing mistakes made when completing questionnaires and other forms, plus a group discussion of the most common errors and how to avoid them in the future. This integrated system of quality assurance

contributed to unusually clean and complete datasets. FEEDBACK project staff working to calculate the WEAI and related values for Guatemala identified only a single instance of a value in the source data falling outside the prescribed range, and characterized the dataset as exceptionally clean.

"All of the guidelines were focused on data quality...at every level...we were always evaluating, assessing the quality of the information we were getting." (General Supervisor)

"When we found something that was wrong, that shouldn't have been recorded the way it appeared on a form, we communicated with the interviewer to correct it, and also with the team supervisor or the general supervisors to standardize the guidance we were issuing from here." (Editing Coordinator)

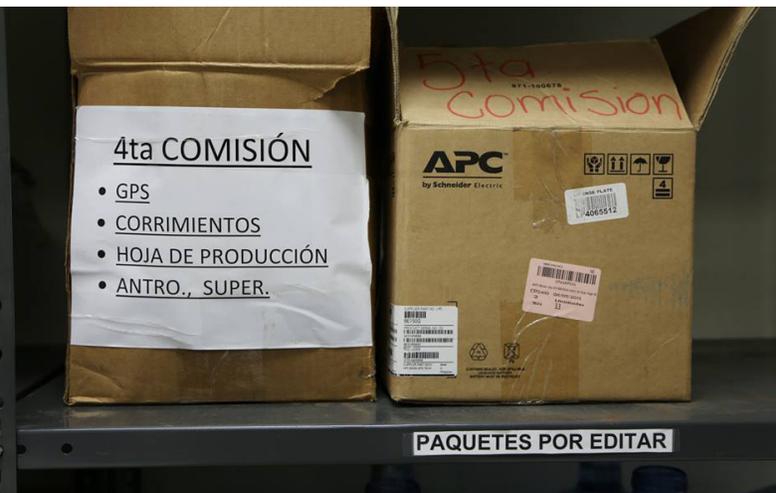
"We had three levels of editing, one that was in the field, the second which was the general supervisors, and another that was done here in the office. And truly, all three levels worked. At the first level, another innovation was using two field editors. We'd planned on having one but there was just so much work to do; it was really complicated." (General Supervisor)

"You can see the [attention to quality] in the databases. When I received the databases, I could tell...they are really, really high quality." (Income and poverty analyst)

Coordination and resource transfer between the central office and fieldwork teams went smoothly, even when the teams were located in very remote areas. Staff members based at the central office were responsible for coordinating the ongoing transfer of a large number of supplies to the field, including but not limited to questionnaires and associated paperwork, cartography materials, and participation incentive items which included a plastic basin and household



A height board towers over a young boy as anthropometrists prepare to work.



Boxed supplies for the project's fourth commission sit on shelves in the central office.

cleaning supplies. General and team supervisors invariably reported that the necessary materials flowed efficiently and largely on time throughout the study's extended fieldwork phase. The technical director, general supervisors, team supervisors and support staff kept in regular contact via cell phone, and reported only rarely finding themselves in locations without signal access.

Twenty weeks of interviewing were organized into five fieldwork commissions of approximately 26 days each, and between commissions teams returned from the field to participate in a series of meetings with central office staff. These meetings took a half day and were designed to allow returning fieldworkers to inventory unused supplies, debrief with the technical director, discuss mistakes identified on completed questionnaires with the editing supervisor, receive the balance of their reimbursements and other pending payments, and obtain new team assignments for the next commission. This routine served to reassure workers that payments and

supplies would be furnished on schedule, guaranteed multiple opportunities to share problems and suggestions, and reinforced the cycle of continuous quality assessment that characterized the study.

"On some projects, I've seen people doing whatever they feel like doing, but not here. Here we were constantly aware of what was happening...we monitored...per capita output on a daily and monthly basis. People realized, hey they're looking at what I'm doing, I have to get my work done..." (General Supervisor)

"Something I introduced for the first time was a system of monitoring from the minute the teams left until they arrived in a sector. We could locate them by GPS...and the GPS would tell us when they were supposed to arrive...we'd call and ask, 'why were you late?'... 'oh, we had to stay in [community] overnight because there was a landslide.'" (Technical Director)

"From my point of view, each survey creates a culture and a way of working and fosters certain habits...it's something that's hard to convey to people sitting on the outside: how important it is for the team working on the inside to function cohesively." (Technical Director)

"Look, I've managed a variety of projects-big ones, small ones...from ten thousand dollar projects up to multimillion dollar ones. The WHIP baseline has been one of the most organized." (Finance Manager)

"I really liked the way the work was coordinated, and the sense of responsibility that the groups had and that individuals had. Because at other organizations or companies, they just send the interviewer out and there's no support at all, no coordination...[this study] showed great work on the part of the [fieldwork] coordinators." (Supervisor-Cartographer)



A man is interviewed outside his home.



A dirt road leading out of one highlands community.

Careful training and a flexible approach prepared every team member to respond effectively to participant concerns, minimizing refusals and nonresponse. In the vast majority of communities selected for the sample, local authorities granted the team permission to work, and cooperation rates within these areas were impressively high. The community and facility team successfully converted a number of community-level refusals through sharing additional information about the study and responding systematically to local leaders' questions and concerns. In sectors where teams were given permission to work, a household cooperation rate of 97% or greater was achieved in every study domain. Approximately 97% of women and 92% of men asked to participate in the empowerment questionnaire did so. All other

cooperation rates by questionnaire type were above 95%.

Women also overwhelmingly agreed to anthropometry and anemia testing for themselves and their young children. When women were reluctant to consent to participate in study activities without their husbands' permission, interviewers waited or returned to the household when men were more likely to be home. Female interviewers conducted women's interviews and anthropometry, and stepped in to do additional interviews when men reacted with suspicion to the idea of allowing male interviewers inside their homes.

Section 2 of the empowerment questionnaire, regarding household assets, posed the only notable exception to otherwise high response rates. Eighteen percent of eligible respondents interviewed skipped this section entirely. Both community leaders and household survey respondents expressed concern that information on the value of their property and possessions might be used by the study team or affiliates to seize these assets. These fears were founded on similar incidents that had taken place in their communities. Interviewers felt strongly that had they pushed harder to convert nonresponse on these questions, many respondents would have refused to continue with the interview process entirely. Additional information on the unique difficulties posed by the empowerment questionnaire can be found under the Challenges and Obstacles section of this report.

*"Instead of focusing on the total time [interviewing] took, we focused on doing it in stages...if you say to someone, 'Ok look, this will take two hours, three hours...' they say no. But if you say, 'We can do it in parts, at different times,' then they agree."
(General Supervisor)*



Team members working by headlamp at night.



A child stands on the scale so anthropometrists can record her weight.

"There's still a lot of machismo, sometimes men wouldn't allow their wives to answer the women's questionnaire. So we explained to them that no, it's really the woman we need to speak with, not the man, because the questions have to do with women's problems." (General Supervisor)

"One success story happened in places where the entire community refused to participate...we'd ask why and they'd say, "Because you're going to vaccinate the women and that will make it so they can't have children." I went with [two general supervisors] to a community like this, we gave a detailed explanation and showed them the letters we had from the town mayor and the local health facilitator... that was one success: we recouped

this community where previously every sector had refused to participate." (Fieldwork Coordinator)

"The level of refusals in this study was minimal, and that was a huge achievement...the interviewers were 100% committed, they gave it their all and were really well trained, which helped avoid the mistakes and confusion that can lead to refusals." (Community and Services Coordinator)

"The knowledge that study personnel had [was vital]...from the supervisors down to the smallest position on the team everyone was prepared at any moment to explain the study." (General Supervisor)

"I think it was crucial...that everyone knew and had the tools necessary to respond to people's concerns at any given moment, when people would start saying 'no, I'm not doing this study,' to be able to respond: 'look, it's not what you think.'" (General Supervisor)

The act of implementing the empowerment questionnaire was itself perceived by interviewers and interviewees as promoting women's empowerment. Despite notable problems with implementation (see the Challenges and Obstacles section of this report) interviewers felt that the empowerment survey represented a unique and important opportunity for promoting agency among women. Since women and men in a household are explicitly given equal importance as empowerment module participants, and because the questionnaire covers topics and activities normally dominated by men in the community, the survey was perceived by team members as unusually inclusive in both its concept and approach. Individuals interviewed for this report expressed satisfaction that the subject of women's decision-making power and control over economic assets was being introduced in Guatemala, and were interested to see how the integrated program might make a difference. Overall, despite significant challenges with operationalization, team members singled out the empowerment questionnaire as one of the most innovative study components.

"I was really glad to see that women were participating [in the survey]. Because other surveys I've worked on weren't like that. It's hard sometimes to convince the men that women should participate. I like that for this survey we say to them, "Look, I need to ask you some questions, and I also need to ask her some questions. So...they're equal. They participate as equals." (General Supervisor)



Fieldwork staff review completed questionnaires.

"This is a ground-breaking survey...it was different than other studies because it included women's empowerment, which hasn't been measured here before...other surveys have asked about violence, but things like how much influence women have, how much decision-making power; that's different and important." (Fieldwork Coordinator)



A man participates in an interview for the survey.

The baseline study benefited from extensive, sustained cooperation between stakeholders. Three separate Offices at USAID Guatemala contributed resources and staff to a core team charged with oversight for the study; Health and Education, Economic Development, and the Program Office. The two years that passed from inception to study completion brought a number of shifts in the oversight team's personnel;

these changes were accommodated without major difficulty because institutional memory and project-related expertise were shared among members of the group rather than situated with a single office or one individual. In accordance with guidance provided in the agency-wide Evaluation Policy, the Program Office assumed leadership of the project, engaging the relevant technical Offices in study planning and decision-making and liaising with the technical leadership team from MEASURE Evaluation.

The technical leadership team was also characterized by unusually cooperative effort. It included key staff from two MEASURE Evaluation partners, UNC and Tulane, and affiliated consultants in Guatemala and Mexico. The presence of project staff from two universities under the MEASURE

Evaluation umbrella broadened the pool of available expertise and expanded the resources and mechanisms available to the researchers for accomplishing key tasks. Questionnaire translation into English for the US-based IRB application, for example, was conducted by a language professional based in Guatemala City who has worked for Tulane on a number of prior occasions. She was therefore already approved to work as a consultant for the University, significantly expediting the contracting process. The presence of multiple MEASURE staff on the technical team also allowed for both joint and individual visits to Guatemala, providing for an extended presence in-country at key junctures.

Operational support from the Ministry of Health and Social Assistance (MSPAS) in Guatemala was also instrumental to the study's successful implementation. This support came in two primary forms: 1) physical resources including donations of ferrous sulfate for study participants with anemia and office space to supplement the space available at INCAP; and 2) official endorsement of the study including permission to use the MSPAS name and logo on study documents. MSPAS was also the local agency through which institutional ethics review and approval were sought and received. Official support for the study's ethical and implementation protocols from the country's highest authority on health service and research represented an invaluable resource. Without ample political will for the project and coordination between MSPAS, USAID, and MEASURE Evaluation partners, study milestones would certainly have been achieved much later, if at all.

The National Statistics Office (INE) also provided essential services; INE cartographers worked with the research team to provide updated maps of requested sectors for use in identifying RVC participants' locations and developing the sample in every domain. Although communication problems sometimes led to delays in cartography-associated tasks—community team members told of arriving to work in a sector only



The principal investigator addresses the local team at the project's closing meeting and celebration.

to find that they were missing some of the required maps—ongoing support from INE was absolutely crucial to the study's successful implementation. Finally, a team of statisticians and data management experts at Westat, working under the auspices of the FEEDBACK project in support of FTF programs worldwide, calculated the WEAI, sub-index and component values for the baseline dataset in Guatemala. Their competence with these calculations, and assistance with interpreting the results, greatly expedited the submission of country-level FTF results to the Bureau of Food Security at USAID in Washington.

"The Mission had...a clear vision of what they wanted to do...I think from very early on the group realized that they had an important project...[WHIP] involved a lot of resources, it was going to serve a large population, and it was also very important to the Mission's portfolio in terms of resources and visibility." (Principal Investigator)

"We also had support from USAID Washington...they kept in communication with USAID Guatemala, and with us, to provide as much support as they could." (Principal Investigator)

"I think that being able to use information from INE, in particular the cartography materials...really helped to define the area of influence for the RVC program." (Sampling Statistician)

"Something that helped a lot was having immediate financial support from the donor. When there's no delay in resources it's easy to assign funds to the technical units so they can pay for salaries, services, vehicles, materials purchases, and everything else they need..." (Finance Manager)

"One really good thing was that INCAP was very, very fast in the financial aspect—we'd draw up a request for checks or purchases or whatever, and they'd work on it right away. I think this helped a lot." (Project Manager)

"I didn't expect to see so much interest in working together. It was a really nice surprise. [This included] the government [of Guatemala], our Office, the implementers who come from different sectors...everyone shared an interest in collaborating." (M&E Advisor, USAID)



A team organizes questionnaire forms.

The community and facility survey teams performed multiple critical functions, and their diplomacy was one of the study's most important assets. Nine people working in three groups of three traveled to study communities ahead of the household teams, working first to identify local authorities and opinion leaders. In meetings with these individuals (and sometimes entire communities), the team explained the study, negotiated entry permission for themselves and the household survey workers, enumerated health facilities whose catchment area included the community, and completed community and facility surveys. The team also visited markets and food stores to complete weights and cost equivalency sheets used to itemize household expenditures on the expenses and consumption questionnaire. Last but not least, the community team was responsible for verifying the cartographic information that would be used by the household interviewers, including confirming the locations where



A fieldwork staff member on the move.



A corkboard in the project office shows tasks by status and priority.

RVC program participants resided.

The community teams were also the first to discover important logistical problems such as other studies being conducted in study areas at the same time as the WHIP Evaluation baseline. This advance knowledge allowed household teams

to be preemptively re-routed to avoid overlap and minimize associated delays. The community teams' efforts frequently enabled them to provide early notice of other problems in study sectors including social unrest, and facilitated planning to identify and hire translators in areas where this was a necessary service. Members of the community team acted as diplomats and strategists, not only collecting primary data directly but also playing a highly significant role in facilitating the work of their peers at the household level.

"One new thing that we did was...divide up into two groups with one that went first. And I think it was really successful, because when the (household) group has to do all the work required to get permission from the municipality they use up a lot of time just sitting and waiting, and it's a waste of resources." (Community and Services team member)

"There were some sectors in domains 1 and 2 where we had to newly re-locate the [RVC program] associates. So the supervisor had to know every inch of the map." (Community and Services Coordinator)

"We also conducted a health facility survey. We looked at coverage, and took an inventory of medication stocks, to see what was lacking. All of this had to be explained to the community too, because often the health posts are really connected to the community leaders, in fact it's usually the same people in charge." (Community and Services Coordinator)

"We went to speak with the medical staff person in charge, who would say, "In this community the catchment area for this health post is x." We always had some advance knowledge about it, since we'd already spoken with the community leaders and asked them where health services were available. So we were corroborating the information...and it was like having two filters." (Community and Services Coordinator)

"Upon arriving in a community we went directly to the Department office to announce ourselves to the municipal mayor, the area health directors, even the police. That way they knew exactly what we were planning to do, and when. We provided our license plate numbers as well...Guatemala is a complicated place and misunderstandings are common..." (Community and Services Coordinator)

"Something that helped us out a lot, a whole lot, was having the Community and Services team traveling ahead of us. That really opened the doors to communities... when we got there all we had to do was introduce ourselves and get to work." (General Supervisor)



Fieldworkers crossing rough terrain to reach a study community.

Findings: Challenges and Obstacles



A team shares a meal of beans and tamales during fieldwork.

Fieldworkers would have benefited from additional preparation for the physical and psychological challenges associated with extended commissions in remote areas. Fieldwork conditions were often very harsh, reflecting the program's exclusive focus on people living in one of the most inaccessible and under-resourced parts of the country. Even workers with extensive fieldwork experience noted that the circumstances under which the WHIP Evaluation baseline was conducted were unusually extreme. Apart from the normal difficulties associated with travel in remote areas, the teams' mobility was complicated by external factors including heavy seasonal rains, and police and drug traffickers in some areas imposed further restrictions on access to study communities.

Workers were occasionally discouraged by the unfamiliar and hard-to-administer survey modules, unusually long working hours and extended interview durations. The volume of work was described by some study team members as overwhelming; even experienced supervisors spoke of fearing that they would not meet established goals on time. The study also had a steep learning curve since major aspects, such as the empowerment questionnaire, were being implemented at scale



Teams carrying supplies move through areas only accessible on foot.

in Guatemala for the first time. The expenses and consumption questionnaire had been implemented in Guatemala before but never as part of such a large questionnaire set. Many interviewers had no prior experience with the consumption questionnaire, which involves considerable repetition and requires keen precision.

Because of the long duration of interviewing within a household, and the need to speak with multiple respondents, interviewers often spaced their visits out over several days. Practically, this meant that the members of a household survey team sometimes spent as many as three days working in a single household. The strategy offered clear advantages but also heightened the intensity of the relationship between interviewer and interviewee. Because the survey covered a number of sensitive or potentially upsetting topics, among them infant and child mortality, food insecurity and poverty, this closeness may also have made interviewing more difficult in some cases. Several interviewers reported that the questions regarding household hunger were among the most emotionally taxing to administer.

Supervisors also indicated that in some study communities, procuring basic accommodation and food for their teams was neither simple nor assured. While the groups did their best to come prepared—traveling with air mattresses to use for sleeping in schools or vacant buildings, for example—these set-ups frequently presented problems. Staying in locations without electricity, for example, meant that editing and other work normally performed well into the evening hours had to be done using candlelight or headlamps. Accommodations sometimes had dirt floors, doors that didn't lock, or lacked running water. When the teams stayed in municipal offices or other buildings used for another purpose during the day, all of their belongings had to be packed up and removed from the premises every morning. Moving between communities or households in remote areas that could only



Travelling on foot.

be reached on foot was physically exhausting, particularly in the highland weather extremes. Teams spoke of losing their way in the dark, falling on muddy slopes, and navigating rudimentary footpaths that gave way to unmarked jungle or mountainside.

Upon arrival in a community, supervisors generally looked to hire individuals to prepare meals for the team over their 2-3 day stay. But food was scarce in some study areas, complicating the community's ability to accommodate the new arrivals. Even in locations where food was readily available, it often consisted of only a few staples, resulting in a monotonous diet that affected morale. Team members also suffered gastrointestinal illnesses that were presumably food and water-borne. In qualitative interviews, team members indicated understanding that these hardships were simply inevitable in the field. Nevertheless, many also suggested that having more detailed information about the conditions they might encounter could have better prepared them mentally to live and work in study communities for extended periods, and suggested that fieldworkers be specifically briefed on what to expect during training for future data collection rounds.

"On the one hand, I think we were all very well trained to be able to perform the roles we were assigned, but...it's a really exhausting job. Extremely exhausting. Sometimes you have a quota that seems maybe a bit too high for the group's capacity, you know? It can wear you out, this kind of work, and then the quality might suffer a little bit, too." (Community and Services Supervisor)

"Most people deal with [the stress] pretty well for the first month, but after three, four, or five months the exhaustion starts to creep in..." (Community and Services Supervisor)

"The longest we had to walk to reach a community was 6 hours...we did it in two stages and rented mules to help carry the equipment." (General Supervisor)

"Sometimes we had problems getting food—just because these are agricultural communities doesn't mean they have all kinds of food available – tortillas, coffee and beans are all most communities have...for 27 days straight, the same diet everywhere you go." (General Supervisor)

"Sometimes we stayed in schools, community centers, meeting rooms, houses that were under construction...not places that are going to have a bed, a bathroom, electricity..." (General Supervisor)

"I had an incredible experience getting to one community where we were told the walk would last an hour, and it turned out to be more than three hours up into the mountains without even a footpath. The fog was unbelievable; you couldn't see anything....and then it started to rain. (Community and Services Coordinator)



Workers pose in front of a community development committee office.



Editors working together.

*“Another time...we walked for more than 12 hours just to get to the community...and that meant walking twelve hours back.”
(Community and Services Coordinator)*

“...there weren’t roads, just little paths made from tree trunks...how people came to live there we don’t know. How they ever leave to get things, we also have no idea. It’s incredible to be in some of these places and know that there are people living there.” (General Supervisor)

“There were sectors where updating the maps was impossible because the army prohibited it. In some places the police weren’t allowing anyone in, and in others the drug traffickers weren’t.” (Technical Director)

“We traveled with our questionnaires and incentive items, but the household teams really had the heaviest burden because they carried all the forms—a huge quantity of forms—plus the anthropometry equipment, plus more incentive items...and on top of that, all of their clothing and personal supplies because they stayed much longer in the communities. They deserve a lot of credit; they had a very hard job.” (Community and Services Coordinator)

“There were places where we didn’t have electricity, water... personal hygiene is something that can...well, it can even affect you psychologically...I think the thing to do is offer a small workshop. That’s what I would do, to let people know, so they don’t get to the field and get scared off. I mean, people are accustomed to [difficult conditions], but not to the extent that it was like during the WHIP survey.” (Community and Services Coordinator)

The survey’s burden on respondents was also extremely high, provoking response fatigue and related concerns.

Interviewers spent as many as six to eight hours working in each participating household, and individual interviews usually ran between 30 minutes and two hours long. In pilot testing, respondents in one community noted that participation in the study took so long that it effectively substituted for a day’s work. Key informants interviewed for this report echoed similar concerns about the burden placed on respondents, and the real potential for the study itself to further encumber already-struggling families. The expenses and empowerment survey in particular prompted respondent fatigue, and interviewers had to work hard to keep interviewees from ending the interview early. This questionnaire, central to poverty estimation and the method of choice for FTF population-based surveys, took at least three to four times as long to administer as any other study module. Although they were highly successful at enlisting respondents’ cooperation, study staff believed that the length of the questionnaires may have contributed to item nonresponse. In particular, interviewers felt that consumption was being systematically



A study participant holds her child as she completes an interview.



A dirt road winds through a study community.



An anthropometrist tests a woman for anemia.

underreported by fatigued respondents hoping to shorten the interview process.

By calling specifically for male respondents, the empowerment questionnaire introduced additional challenges. Men in households selected for the study were routinely away from the home for long hours during the day and often unavailable for interview on weekends because they were accustomed to devoting this time to recreation and social activities. Interviewers tried adopting non-standard interview schedules and making repeated return visits, but nonetheless the cooperation rate among men eligible for the empowerment questionnaire was somewhat reduced:

approximately 92%, compared to women's 97% in sectors where interviewing took place. Markedly lower rates of item response for specific questions in this instrument are explored in detail later in this report.

"I think the biggest challenge was with the Expenses and Consumption Questionnaire, because it was so extensive and took the most time to implement in the field. Getting people to answer such a lengthy questionnaire was one of the hardest things... figuring out how to incentivize it, so people wouldn't cut off the interview before they finished." (Editing Coordinator)

"[One challenge] was just how much information we were trying to collect, and at the same time be respectful of people's time..." (Mission Economist, USAID)

"On average, the household questionnaire took 30-45 minutes. The women's questionnaire took an hour, maybe two, depending on the number of children. Empowerment, 45 minutes to an hour...but the one that took the longest-sometimes between 5 and 6 hours-was the expenses one." (Technical Director)

"The first issue with interviewing men was locating them. Generally they don't get home until late, and they're tired... but there's also the problem of mistrust...in some places the mining companies had done the same thing, asked people about the value of their land and how much they would sell it for, and then they came back with the money and took the properties. So now when you come around asking questions like that it scares people." (Technical Director)

"We started to wonder if [the study] was even a good idea, because we heard that interviews were taking six or eight hours [in each household]...our first thought was, listen, six or eight hours is two days of work, that's two shifts...if we take two days of work away from the poorest people are we helping them or making the situation worse?" (M&E Advisor, USAID)

"The other thing [to ask] is, what best practices are there for evaluations, to compensate people for the time they're devoting to participating in the survey...if someone showed up on your doorstep...would you spend five hours talking to them?" (M&E Advisor, USAID)



A baby is tested for anemia.

Despite precautions the teams faced personal risk in some communities, and negotiating permission to work in these areas was a time consuming process. The study sample included 22 sectors located in the three municipalities that make up the ‘Ixil Triangle,’ the center of political violence during the early 1980s. Residents of the triangle remain largely closed off to outsiders. In addition, communities in another study area were engaged in conflict with government and private agencies over proposed hydroelectric and mining projects. During the latter half of fieldwork, events in one location came to a head after the death of a soldier that was blamed by residents on the government and vice versa. Suspicion coupled with confusion in the area led to instances where study teams hoping to begin work in communities were instead detained there for as much as an entire day. Throughout September, the study’s technical director kept colleagues in the US and at USAID Guatemala apprised of the situation:

9/3/2013: “... today was a very tense day and our movement in some sectors is limited. The team working to locate RVC program participants has not been able to advance according to schedule, and neither have the cartographers...”

9/30/2013: “I’m attaching a link to an article from today’s paper about the situation in [community]. The teams there have had serious problems. On Friday one team was detained from 8:00 a.m. until 7:00 p.m.; they were released unharmed...”

At the urging of USAID and the Principal Investigator, the teams left the community in question, opting to return there and to other places where gaining access had been especially problematic at a later date. Ultimately, the return visit also had to be cut short:

10/15/2013: “Colleagues, this is to inform you that we are halfway through the fourth fieldwork commission... in [community], we’ve tried a number of times to complete the sectors selected for the sample, but none of the local authorities are willing to guarantee the teams’ safety...”

In certain areas, the study’s affiliation with the government of Guatemala also occasioned suspicion or opposition among community leaders and other residents. Official letters of support for the research, study materials carrying the MSPAS logo, and other evidence of government involvement—instead of legitimizing the work, these served in some places to provoke automatic refusals among community leaders and residents. Of note, at baseline most interviewees likely had little to no knowledge of the integrated program or its component activities, which could have heightened suspicion about the study’s purpose and content. As program awareness increases across the WHIP focus area, people may become less apprehensive, particularly if the program is viewed favorably.



Editors at work.



A supervisor conducts an interview.



Interviewers talk with their supervisor.

The study's use of hemoglobin tests requiring a fingerstick also provoked strong reactions among respondents unfamiliar with the procedure and wary of deception. Residents in some areas recounted isolated incidents of aggression towards health or survey workers conducting blood tests under other auspices in the past, prompting fieldwork team members to proceed carefully. Further complicating matters, while a two-week course of treatment was originally planned to be distributed directly to women and children with anemia, restrictions on the purchase of pharmaceuticals with USG funds delayed acquisition of these supplies. Consequently, for a short time participants were instead provided with a referral card to take to the nearest health center for treatment, but many expressed little confidence that these facilities would have medication available. Ultimately MSPAS donated ferrous sulfate to the study, resolving the issue, and supervisors reported that offering the short course of treatment on the spot markedly increased participants' acceptance of the test. Situations like these underscored the need for both cultural

sensitivity and operational flexibility, especially in areas where the trust between the government and local residents is already fragile.

"People said one family in this community had taken their child to the health post, and he was vaccinated there and then he got sick. And to them this meant that at the health post they did something bad to the child. It caused major problems; the nurse in that instance was assaulted...and the community forced him to leave. Because of this kind of experience it's hard to convince people to let you [take the blood sample.]"
(Community and Services Supervisor)

"There were other surveys where they took a lot of blood from participants. So people would say to us, 'you're going to take a liter [of blood], aren't you?' Sometimes we met and worked with the health facilitators in a community, which helped a lot. One time we spent four hours explaining the Hemocue process to everyone, and demonstrating how it was done, and that helped a lot, too." (Fieldwork Coordinator)

"Well [the highlands] are very different than the city...in places where there had been political problems we had to make very clear that we weren't a part of that...that we weren't with any political party, or the private companies, nothing to do with that. And all of the letters and everything we had with us proved it." (General Supervisor)

"There was one place where we couldn't [convert the community's refusal]. We went back to try...but the people were really angry. I called my supervisor and said 'Look...they don't want us here and we're going to leave.' I could hear them yelling, 'get out, get out!'" (Fieldwork Coordinator)



Stunning sights along the way.



An anthropometrist offers a child a lollipop for her bravery.

“Even though we put a lot of effort into getting permission from the local authorities, from the Ministry of Health, to work in these communities, the fieldworkers faced a lot of security problems. There was, in some sectors, misunderstanding about their role.” (Principal Investigator)

“We entered [municipality] at a difficult time. [Residents] are highly resistant to the study, to everything, because they’re afraid everything people come to do is related to plans for installing a mine, a hydroelectric plant, something that will harm them. During the last part of the commission, we left just in time because that weekend everything devolved into total chaos. They

seized the police station and lit it on fire, erected barricades, burned tires, everything.” (General Supervisor)

“One lady told us, ‘Look, there’s a car following you but don’t worry about it because they’ve already decided that you’re just doing your job.’” (Team Supervisor and Driver)

“Our safety is our first priority, and also the security of [respondents], because people in the community might be open to participating in the study, but the leaders or other people who have power within the community will sometimes hold it against them. We had to be very, very vigilant. That’s why the Community and Services team was so important...” (General Supervisor)

“In [community] the teams completed their work without any problems, but residents advised them that the next day there was going to be a major demonstration and that if the teams left using the main road they’d be in danger. So they recommended that they leave at three o’clock in the morning and use the road that runs through Mexico instead. They did it, but it took an extra two days.” (Technical Director)

Empowerment concepts proved difficult for many respondents to understand, and some questionnaire content raised serious objections. Empowerment questions frequently proved complex to translate into Mayan languages, and the team felt they were conceptually difficult for many Spanish speakers as well. Whereas participants commonly had experience answering factual questions about their birth histories, caregiving practices and use of health services, the WEAI involves new and arguably more complex response tasks. Source questions for the index ask interviewees to conceptualize and quantify their relative decision-making power for the production, use, and sale of crops, among other topics. Interviewers noted that study participants seemed to have difficulty assigning relative influence to their own role. It was not uncommon for respondents to forget what the question was by the time the response options had been read to them, or to ask that response options be repeated a number of times. To prevent confusion, the team developed visual response aids to accompany the administration of selected WEAI component questions as needed.

The questions related to respondents’ assets (Section 2 on the WEAI questionnaire) were mentioned as problematic over and over again in qualitative interviews. Non-participation on this section was unusually high: as previously mentioned, 18% of cases in the WEAI subsample did not complete Section 2 and were subsequently dropped from the empowerment index analysis. Many study personnel reported that because of Guatemala’s history, including recent sociopolitical tension in study communities, asking people to enumerate



Thousands of questionnaires are arranged in numbered stacks at the project's central office.

the worth of their land and goods incited significant concern among respondents that their assets would be seized. Residents began to question the interviewers' motives and intent, jeopardizing survey work in general. Interviewers were eventually hesitant to encourage worried respondents to complete the section, since doing so had led others to stop participating in interviews entirely. Because respondents who skipped Section 2 had to be omitted from WEAI calculations, in preliminary work analysts at Westat investigated the potential for bias in the WEAI score from the missing data. Fortunately the demographic differences between respondents who did not respond to the assets section and those who completed it were largely unremarkable. Nonresponse was more common in urban areas than rural ones (33% versus 16%) and was somewhat concentrated in the Departments of San Marcos and Huehuetenango.

In some sectors where interviewing was halted entirely due to security concerns, the questions about assets were also at the root of the problem. In a message to the rest of the MEASURE Evaluation team in October 2013, the technical director summarized the issue: "We're told that two or three years ago a hydroelectric company conducted a survey here asking people about their property values and income, then forced them to sell their land. Since our survey includes similar questions the local authorities feel it could provoke a bad reaction." Ultimately, the decision to apply the WEAI module in only one out of every three eligible households was also the result of implementation difficulties. Conducting the module in a subsample of households instead of every one was critical to allowing the team to complete fieldwork on time and likely helped maintain high response rates. However, as the subsample was designed to include only households in program areas, the program's performance on WEAI outcomes can be assessed but its impact cannot.

"Another thing was the empowerment questionnaire...because of the kinds of questions included in it...about their possessions;

people here don't trust you when you ask that...Section 2 was where we saw the most problems with getting people to agree to give the information..." (Editing Coordinator)

"Commonly people would say they didn't own something, but then another question later on would relate to [the first thing], and they'd say yes...it was as if they didn't want us to know what they had, for the same reason I mentioned earlier: extortion..." (Editing Coordinator)

Differences in institutional norms for contracting and accounting led to administrative misunderstandings. The study was implemented through the combined efforts of a large number of organizations, and its management structure, with a technical director situated outside of the agency managing hiring and purchasing, was unusual. Occasionally, lack of communication across collaborating institutions led to misunderstandings. In one notable instance, several interviewers opted to leave their positions to work on another study also being implemented under subcontract with INCAP. The outgoing workers indicated that the other study was able to offer higher pay rates and per diems, although the WHIP baseline rates had been established at what the study's leaders understood to be the maximum permitted by the contracting agency.

Originally, the project's technical leadership had also sought to issue fieldwork employees contracts lasting 6 months



Workers explore a study community.



A broken-down bus blocks the roadway, preventing the teams' vehicles from passing.

or longer, to cover the entire period during which their contributions were expected to be needed. Term contracts such as these would have provided extended job security and perhaps guarded against the attrition. Instead, monthly contracts were in effect for most positions during a majority of the implementation period. These contracts had to be renewed at the start of each fieldwork commission, costing additional time and resources. The reasoning for this was related to institutional rules prohibiting the issuance of a contract without sufficient funds already in place to cover all anticipated payments. Because the administrative subcontractor received monthly payment advances to cover estimated expenses for the project, rather than a lump sum at the start of the contract, the organization was reluctant to consider all of the funds required to cover the longer-term contracts guaranteed.

Other examples of miscommunication that may have led to losses or delays were also in evidence. Cartography doesn't appear by name on the position list used to establish pay scales for the hiring organization, and people recruited for this position on the community and facility survey team, despite being in high demand, were offered a lower-than-expected pay rate. The requested rate would have required that cartographers hold a university degree, something that's uncommon in a position where experience has traditionally substituted for an advanced degree. Further, in-house assets for photocopying and INCAP's own fleet of vehicles were passed over for use in the study since lower prices were obtained elsewhere. This represented an unexpected loss for the subcontractor. Ancillary fieldwork costs, including translators' pay and stipends for guides and porters, were often higher than in other surveys, and in some areas near the border with Mexico essential supplies such as vehicle fuel were only available through informal channels. These expenses were difficult to document satisfactorily for reimbursement

purposes. Although the conditions teams would face in the field were difficult to precisely anticipate, earlier and more thorough communication between the technical and administrative leadership teams about anticipated needs and procedural requirements might have prevented some of these issues.

Lastly, differences between accounting systems used by the study's technical implementers and administrative subcontractor occasioned uncertainty and minor delays during early invoice preparation and review. One system reflects unliquidated expense obligations as a portion of total expenses, while the other reflects only liquidated expenses in this category. Because unliquidated obligation amounts were occasionally greater or less than the eventual expense value, the two systems reflected slightly disparate balances until the end of calendar year 2013, when all remaining obligations were liquidated. The overall result was the need to maintain two sets of expense reports in order to meet the formatting and content requirements of both systems, a time-consuming process that was sometimes confusing. Costs under the subcontract, however, were ultimately extremely well-managed. Total funds requested by INCAP were expended to within approximately 91%. The balance was returned to the project and used to cover other study costs.

"...the contracts were month-to-month, so people thought: well, next month they might not renew my contract, so if someone else offers me something that lasts a year I should take it." (Technical Director)

"We maintain an administrative system that obviously people weren't familiar with. So that implied a period of adaptation." (Finance Manager)



Fieldworkers make plans for the day ahead.

“Most projects are INCAP/donor...but in this case there were three parties. In that sense it was a little different than what we’re accustomed to, you know? But I think that...we learned to manage it.” (Finance Manager)

“This project ran on a lot of trust because we had so many expenses in the field, translators—the amounts we paid! And we have no way to certify all that, there’s just no way to do it...” (Finance Director)

Fieldwork was affected by a number of other studies being undertaken simultaneously in the same geographic area.

In July 2013, the teams began to encounter interviewers employed on other studies working in areas selected for the WHIP baseline. In many cases, not only the same communities but the very same households were the focus of these simultaneous research efforts. Ultimately, as many as five separate data collection initiatives were identified as coinciding with the WHIP baseline in communities included in the sample. One of these simultaneous efforts was a baseline study commissioned by AGEXPORT and meant to capture results specific to its partners under the RVC program.

One effect of the overlap between studies in the field was an increase in the demand for translators in many communities, which in turn led to unexpected increases in the asking price for these services.

The main cause for the overlap, again, seems to have been failure to communicate. The development of the WHIP sample and subsequently, the

fieldwork schedule, were lengthy and complicated processes. The potential for the study to cross paths with other data collection initiatives may either have seemed remote or simply escaped the notice of individuals who were in a position to recognize it ahead of time. Under pressure to get the study into the field as soon as possible, the implementation team and USAID focused narrowly and out of necessity on operational details including finalizing the sample and questionnaires and preparing to travel. Once in the field, confusion prevailed when a number of community teams were turned away by local authorities under the mistaken impression that WHIP interviewers had already visited that location.

As soon as the situation started to become clear, USAID facilitated a multi-institutional effort to coordinate necessary changes to the teams’ routes and schedules. Despite this timely response, the ramifications of the changes were significant, since successful data collection hinged on the efforts of a series of interdependent groups traveling in sequence. Cartography work, community and services interviews, and the completion of consumer item weights and equivalency forms all had to be conducted in careful succession, and ideally just prior to the household teams’ arrival. As a result, re-routing one team actually meant re-routing several. Apart from the time and resource costs associated with route and schedule modifications, the situation may also have lowered response rates as frustrated community leaders and residents



Conducting household interviews.



Anthropometrists get assistance with steadying the height board.



A team poses for a picture.



Recording information on a questionnaire.

voiced exasperation at being asked to participate in multiple surveys covering similar topics and executed so closely together.

"Today we were informed by the Departmental Health Office that fieldwork by [other agency] just started in [three WHIP municipalities]. The problem is that the WHIP baseline is supposed to be happening now in the same places...we already agreed to change our route twice for [other studies in the area], but since these are the last sectors left, we don't have anywhere else to go." (Email from the Technical Director to the rest of the leadership team, October 2013)

"[The weather and terrain] affected us but we were prepared... it was the other studies in the area that really caused us problems...they were also doing anthropometry, doing hemoglobin tests, and when we got there the people didn't want to do it anymore." (Fieldwork Coordinator)

"We'd get to a community and the people there would say, 'Your teams already came here, we already answered these questions.' A supervisor called to tell me, and I said to her, 'We're the only ones that leave a nutrition card, so if our teams have been there before, the families should have the cards.' They said 'No, no one gave us any cards.' So it definitely wasn't our team..." (Fieldwork Coordinator)

"The local authorities were getting impatient because they were receiving all these interviewing teams with very little time between them...probably some coordination would have been useful to avoid that." (Principal Investigator)

"When you have too many surveys at one time, there's also increasing competition for resources. You're competing for good interviewers, and there were some instances of people leaving our teams to join other survey teams." (Principal Investigator)

"The simultaneous evaluations that were going on...certainly complicated the work...we had a meeting here with the health team, with the agriculture team, with Mission management and hashed through all of that." (Mission Economist, USAID)

"It caused a lot of problems, including with the local authorities who were asking, 'you're here to measure things again?' if someone had been there only the week before to take measurements or ask questions. In the future it would be good to have a shared agenda for evaluations that are being planned." (M&E Advisor, USAID)

Program implementing partners labored to compile the detailed information necessary for the sample. The baseline posed a kind of catch-22: in order to plan fieldwork, the research team needed to match people enrolled in RVC programs to the census sectors where they lived. At a minimum, information was needed on approximately how many beneficiaries resided in individual sectors within the WHIP focus area. When the research team started requesting related information in July 2012, however, ANACAFE and AGEXPORT had only recently received notice that their proposals would be funded by USAID for the 2013-2017 program cycle, and work plan development was still underway. Program expansion and beneficiary enrollment in many areas was incomplete, but fieldwork delays were steadily reducing the



An interviewer speaks with study participants.



Mountaintop view in one highlands community.

researchers' ability to assess the situation near the program start date.

In addition, accustomed to working with producers' associations more than individual producers, and used to collecting data for traditional agricultural support program indicators such as crop yields, the RVC partners had little experience tracking the kinds of information being requested for the WHIP sample. The data on hand also didn't conform to standard census coding schemes. For most producers' associations signed up to work with the program, for example, every member had the same "home address" on file. Often this reflected one of several colloquial names for the area where the association's office was located, which did not conform to place names used in the census and other official sources.

Ultimately, both the RVC implementing partners and the study team spent prodigious amounts of time and resources generating sample source data that still required considerable revision in the field. Community teams arriving in sectors within domains one and two had to re-confirm the information they had gotten previously regarding RVC participants and their household locations. On occasion final confirmation came only when the household teams arrived to conduct interviews in a household. The initial back-and-forth between researchers and ANACAFE and AGEXPORT prompted fieldwork delays lasting several months, and the fieldwork teams identified a number of census sectors in the sample frame for RVC program enrollees where no enrollees actually resided.

Throughout the sample development process, the study's technical advisors recommended that RVC implementing partners adopt beneficiary-level tracking and use national standard coding systems. However, this would have required time and investments in monitoring and evaluation systems and personnel that had neither been anticipated by the

partners nor written into their operating agreements with USAID. Indeed, even the time that partners spent compiling information for the researchers was a largely unanticipated burden. Despite the challenges posed by the prospect of revamping project monitoring systems and procedures, such investments would inarguably facilitate operational efficiency during future data collection rounds for the WHIP evaluation and build implementing partners' capacity for evidence-informed decision making overall. Key informants interviewed for this report also suggested that it would be helpful if assistance required of implementing partners for externally-conducted evaluations could be specified in the partners' contracts. This would allow them time to prepare and enable early identification of any potential overlap between their own data needs and the evaluation being planned.

"Everybody overestimated the information available...knowing where the [RVC] interventions would be implemented took a long time, much more time than was originally planned for, which had some implications for the budget..." (Principal Investigator)

"The implementing agencies were still in their planning stages...and yet being asked to provide data [on beneficiaries]. So there was some tension between the evaluators and the program implementers, because we couldn't move forward without the data, and the programs didn't have it." (Principal Investigator)

"In the end, we visited 16 sectors [selected for domains 1 and 2] where we completed weights and equivalencies forms, community and health facility interviews—and there weren't actually any RVC participants living there." (Technical Director)

Findings: Lessons Learned



Teams converge in the same town during fieldwork, a rare opportunity to re-connect.

Don't underestimate the investment required. Rigorous program evaluations—especially impact evaluations—will always necessitate major investment. Stakeholders accustomed to output monitoring may be taken aback by the cost of efforts that call for representativeness, extensive primary data collection, and teams with specialized expertise in sampling statistics and methods for measuring complicated high-level outcomes like poverty. The baseline phase of an evaluation is likely to be the costliest; many documents including the study protocol generally do not need to be redeveloped for later stages of the research, questionnaires should undergo few if any changes, and for some longitudinal designs sampling is essentially complete after the first round. Work associated with the WHIP Evaluation baseline study cost more than three and a half million dollars, with a large share of this total dedicated to local operating costs. Considerable time is also required: data entry alone took the equivalent of 976 work days. Fully two years elapsed from initial planning talks between USAID Guatemala and MEASURE Evaluation and the completion of WHIP baseline fieldwork; 18 months passed before fieldwork even began.

New projects also require new ways of doing things: training content, team structures, paper and electronic forms, and functional systems of supervision all had to be developed for the WHIP baseline through trial and error. In Guatemala, training for fieldworkers lasted an unprecedented five weeks, and upon reflection some supervisors still felt that additional time would have been beneficial to allow for more individualized attention and a better assessment of skills mastery and fieldwork readiness. Several key informants interviewed for this report also suggested that providing individual vehicles for the supervisors who moved between household teams in the field would have greatly increased the level of support these supervisors were able to offer, including faster response time for problems requiring immediate attention.

Although costs are often closely related to a study's sample size, many other factors influence the amount of resources

needed to support successful implementation. In this case, for example, the extended duration of interviewing that took place within the average study household was a major contributor, driving up the time required to complete fieldwork and by extension, its cost. In addition, the WHIP baseline called for two groups of interview teams traveling separately: working with community/facility leaders and households, respectively. This approach was essential to the study's success but also increased the number of personnel required and effectively doubled fieldwork travel expenses. Costs were further amplified by the study's focus on a geographic region where access is reliably difficult. Incredibly, fieldwork travel for the WHIP baseline study totaled the equivalent of circumnavigating the globe 5.3 times. Study teams traversed 212,276 km (131,902 miles) as measured by the odometers in the 16 vehicles used for fieldwork, and used 46,704 liters (12,338 gallons) of fuel in the process.



Team members traveling in the Western Highlands

"The time needed to implement the survey was much longer than expected and the budget ended up being higher than anybody thought it would be." (Principal Investigator)

"Whoever is in charge of doing the follow up [will need] a good window of time and resources...preparation time is important." (Principal Investigator)

"The first problems were about coordination,...how we could identify the RVC participants and locate them on a map...this took a lot of time and frustrated everyone including USAID...it seemed like it should be easy but it was complicated, and took many months—more than expected. It started to cause distrust and pessimism." (M&E Advisor, USAID)



A team pauses to take a group picture.

“Some people [at USAID] were interested in one thing, others in another; some wanted it done right away...they thought it was easy to just show up, get the data, and present results the next month. This caused some conflict between the different Offices that were financing the evaluation because some people understood that it was a process that would take time to do correctly... but others saw it just like any other activity. There was a lot of money involved, so the feeling was like, ‘if it costs so much, then it should be done quickly and exactly as I say.’” (M&E Advisor, USAID)

Some components of a large evaluation will be afforded higher priority, and internal communication may have limited reach. Despite abundant resources, bigger studies may paradoxically require trade-offs to accommodate their size and complexity. The WHIP Evaluation baseline study was a unique multi-level effort that included interviews in households, health facilities, and with community leaders. While all three of these structures are important to the program and the study, it would have been frankly impossible, and counterproductive, to assign equivalent resources and attention to each. Because of the particular difficulties associated with sampling in multiple domains and interviewing people in their homes, the large household sample sizes, and the importance of the indicators sourced through the household survey, these efforts were in several ways prioritized over community and facility work.



The anthropometrists' most important tool.

More time was dedicated to the preparation and testing of the household, women's, empowerment and expenses questionnaires, for example, than the community and facility instruments, which taken together were considerably shorter and less complex. Household-level data entry was also completed first, in order to expedite data sharing and official reporting on priority indicators such as those required by FTE. The household interview teams were larger than their community counterparts, in order to meet the higher demands of using longer, harder-to-administer questionnaires with more respondents in a greater number of locations. Community teams were also more likely to encounter interviewees who could communicate in Spanish, and traveled fewer hours on foot since their work was somewhat more centralized than the work of the household interview teams.

The size of the WHIP Evaluation also limited communication. The study's administration was supported by a considerable number of personnel working in at least three countries, with varying levels of fluency in Spanish and English. The team in Guatemala was extensive by itself, and comprised an elaborate hierarchy. Many members of this team retained only a passing familiarity with MEASURE Evaluation as an entity and its leadership role in the baseline implementation process. Working from two locations just a short walk away from each other, the administrative and editing teams in Guatemala City nonetheless reported that communication between their offices was often minimal. In effect, the study was so large that it functioned as a set of teams instead of one. Problems resulting from this structure were infrequent but still point to a need to explain the study's management partnerships clearly and introduce its leaders and fieldworkers to one another early in the research process. In addition, communication between the project's two central offices should be specifically facilitated through regular meetings or other means.



The technical director speaks with a team supervisor.

"[One thing I think we could improve] is to have meetings [between the study's principal investigators and the fieldwork staff]...I know it's really difficult when things are up and running because everyone's so involved in the work, and it doesn't need to be all the time, but...just to have some kind of connection." (General Supervisor)

"Between the group that did data processing, and us over here [in project administration], sometimes the communication [was lacking]. I planned a work commission to leave on a certain day, but it turned out they weren't ready. We just didn't have good communication between us...I would like to change things so that we're all working in the same location." (Project Manager)

Integrated programming calls for new approaches, which is both difficult and rewarding.

Integrating two or more programs or major aspects of programming can introduce considerable complexity to a research design but also offers important strategic advantages. It makes little sense to invest significant resources into an evaluation that ignores important aspects of an integrated program's conceptual model or fails to facilitate assessment of the value added by the integrated approach. To do so would risk drawing the wrong conclusions about the level of program impact or the mechanisms by which key interventions are effective. As integrated programs become the service delivery



A member of the community and services survey team at the project's closing celebration.



A girl carries her younger sibling on her back in a crowded market.

model of choice in many settings, evaluation designs like the one applied in Guatemala are likely to be widely replicated. Lessons learned here may be especially applicable to programs that coordinate nutrition and agriculture interventions, but are far from restricted to this milieu.

"The hardest thing was the newness and complexity of the study, because the DHS uses only one questionnaire and this study used a lot of them. Figuring out what it was going to take to implement it, how long it would take—because we really had no idea—that was the complicated part." (Fieldwork Coordinator)

"I'd never worked on a project that was being implemented for the first time in Guatemala before. So it was full of challenges, because everything was new; the concerns were new, the problems were new...although on other projects we certainly encountered problems, you knew exactly how to deal with them: if this thing happens you have to respond in this way. But here it wasn't like that, and like I said before, it really caused me to grow professionally and personally too. Because I was called on to make decisions about the survey and ultimately they were good decisions that helped to get the job done right." (Community and Services Coordinator)



Supervisors review a community map.



Supervisors at work.

"This was the first time a WHIP survey was ever implemented, so there wasn't any template...you couldn't just say, 'okay, here's how everything is going to work.' Instead it was pure trial and error, but now if someone does another WHIP survey the advantage is that we already have some parameters..." (Community and Services Coordinator)

"I think one challenge was that it was a new project. Another was that the workload was really high, both at the central office and in the field. Another was—since we were covering the parts of the country with the highest poverty—the locations we visited were really inaccessible and it was

hard on the fieldworkers. But in the end I think it was good, because we went looking for the people who are most affected [by poverty], not like in other places where you might just scratch the surface and not get to the bottom of the country's problems." (Editing Coordinator)

"If I had to rate it in terms of complexity, I'd say the WHIP Evaluation is 5 to 6 times more complex than the DHS." (Technical Director)

"This culture of evidence-informed decision making is new in Guatemala...we're more accustomed to working from intuition....having information from the evaluation is going to help us completely change the paradigm of how development is done in Guatemala." (M&E Advisor, USAID)

Broad community support is frequently the lynchpin of effective implementation. Once the teams had permission to work inside a community, interview response rates among individuals were uniformly high. Community leaders' support was less assured and generally took longer to acquire, especially in areas with histories of conflict. The research teams learned think outside the box when trying to enlist local support for study activities, and their efforts at baseline will likely pay dividends in future data collection rounds. In Guatemala, religious leaders were often at least as important as mayors and other administrative decision makers in garnering a community's cooperation. Closer coordination with program implementing partners also has significant potential help to move approvals forward. This may be particularly true if the program has become a trusted presence in the community.

The community teams working on the baseline survey also found that attention to protocol was a highly valued aspect of the approval process, and that arranging even initial visits with community leaders in advance was essentially required. Local gatekeepers were often convinced of the legitimacy and importance of the study only after multiple meetings, and no single study element or characteristic proved decisive to the decision. In one instance, more than a month passed between the community team's discussions with leaders and the household team's arrival there. Despite having approved the work previously, community leaders were confused by the second team's presence and failed to realize that two teams were part of the same group. The misunderstanding was resolved, but not before fieldwork was delayed.



Members of the study's technical leadership team at the project's closing meeting and celebration.



Plastic pitchers stacked in the project's office await distribution to study households.

"One technique that we adopted was when we arrived in a place, to ask the local leaders, 'What kind of religion do the people here practice?' Then we'd go to the church and speak with the pastor, and at the next worship service the pastor would tell the congregants about us. We had a lot of success with that." (Fieldwork Coordinator)

"A lot of local leaders had doubts about our work. That was the first challenge, convincing them to open the doors to their communities. And some said, 'Ok, go ahead and work here because I see that you have written authorization from the mayor, from the health department,' but others didn't believe us...or weren't open to dialogue." (General Supervisor)



An anthropometrist explains a woman's body mass index and anemia test results.

"[Community leaders] like for you to show up, to wait, to make a presentation to them, show some things. They don't respond yes or no right away. So it's a process that takes time. But in the long run, we found that people were convinced." (Community and Services Coordinator)

"Before entering a community we ask permission, we set the date beforehand. If you just show up, well, people get the wrong idea, because of the dangers here in Guatemala. So we ask beforehand: who are the people recognized as leaders? And

then they come: auxiliary mayors, development committee members, church directors, midwives, teachers...whomever the community recognizes as leaders, that's who we interview." (Community and Services Coordinator)

The USAID Evaluation Policy and flagship programs like FTF are driving major research. Experience in Guatemala and elsewhere suggests that the Evaluation Policy is working to encourage Missions to identify their largest and/or most innovative programs and administer funds and technical support for rigorous evaluation of these initiatives. Impact evaluations are particularly on the rise. Further, the USAID staff members interviewed for this report were clear about not wanting the WHIP Evaluation results to 'just sit on a shelf.' The study was viewed by USAID and its research and implementing partners as a strategic, high-profile investment with significant potential to spark replication and lead directly to program improvement. Finding ways to encourage the data to be used widely and to disseminate research results to a broad audience could help add even more value to program evaluations, and might also attract additional investment.

"...we knew that these [reporting] requirements would be coming from the Bureau of Food Security, and at the same time there was a big push here at this Mission, which has a great capacity to comply with the USAID Evaluation Policy, so that was my lens: I wanted to make sure we were complying with that policy." (Mission Economist, USAID)

"In Guatemala very little statistical information has been available in the past about poverty and malnutrition at the community level, and this information, when it has existed, hasn't always used strong methodologies or been adequately funded...so for us as Guatemalans working in development, it was a compelling offer...to be able to understand the impact we're having in the WHIP zone." (M&E Advisor, USAID)



The technical director speaks with fieldworkers getting ready to leave on the next commission.

Coordination across USAID offices and partners vested in a study is critical from the earliest stages. Coordination between study stakeholders can promote efficiency, avoid bottlenecks, and make sure everyone's interests are adequately represented throughout the research process. Some challenges faced during the implementation of the WHIP Evaluation baseline might have been avoided with earlier coordination between stakeholders. This includes the overlap between research initiatives in the field. RVC partners opted to implement similar data collection efforts around the same time as the WHIP baseline study in part because the evaluation did not meet their specific requirements for baseline data on the project. While the evaluation is necessarily focused on identifying the integrated program's aggregate outcomes and impact, it might have been possible to reach a compromise that would have provided sufficient project-level data to meet partners' basic information needs.

Unless a program targets an entire geographic area, sampling for an evaluation will likely always rely on beneficiary lists and other information provided by implementing partners, as the WHIP Evaluation baseline did. Partners' M&E capacity, monitoring priorities, existing data systems, implementation timeline, and political will to provide this information are all influential factors. Stronger early coordination between USAID, INE, the

research implementers and program partners might also have facilitated better early-stage sampling. Interviewees agreed that it was unrealistic to expect RVC implementing partners to have compiled complete information on their beneficiaries very early in the project funding cycle. However, both RVC partners had been implementing similar programming with USAID support for some time, and short term agreements were issued to bridge the gap between predecessor projects and those under the integrated program. Earlier and more intensive coordination, as well as targeted support, might have readied these partners to provide the information needed to source the baseline sample sooner and more completely.

The WHIP Evaluation is unusual in receiving technical and financial support from three separate Offices at the Mission in Guatemala. This unified support obviously offers tremendous advantages, among them an expanded network for advocacy, extensive technical resources and a larger pool of potential funding. Coordinated inter-office involvement, however, is time-intensive and requires significant central effort to sustain. The WHIP Evaluation baseline was no exception, and qualitative interview participants expressed some regret that more exhaustive efforts to involve all interested parties at USAID on equal footing weren't advanced sooner in the research planning process.

"It hasn't been easy, first because an information culture doesn't really exist in our country, so we tend to feel possessive of the information, and jealous about it. So if, as an NGO I have some interesting data, I'm not going to want to share it with anyone, even if it might be something helpful for other people to use...I'm not going to want my friend to know anything that might give him an advantage when he's writing a proposal to USAID." (M&E Advisor, USAID)



The project manager holds children's thermometers, ready for distribution to study households.



Interviewers at work.

"I think [the research team] went the extra mile, because when you run into the kind of problems that we did, the easiest thing would be to look the other way...but the research team...kept insisting that the evaluation needed everyone's support. Every time a serious problem arose they called everyone together to find the best solution. If you don't have a team as dedicated as that, it would be easy to fail." (M&E Advisor, USAID)

Respondent-centered research approaches can reinforce USAID's commitment to transparency and benefit everyone involved. The USAID Evaluation Policy acknowledges that "high expectations exist for respectful relationships among donors, partner governments, and beneficiaries... many stakeholders are demanding greater transparency in decision-making and disclosure of information." Establishing forums to share results, even preliminary ones, with implementing partners and their beneficiaries would send a powerful message that the sponsors' interest in communities' wellbeing is genuine and motivated by a desire to implement the best possible programming. Even in places where people were initially skeptical of the research, those who chose to participate frequently expressed sincere appreciation for the opportunity to have their experiences heard and recorded.

Leaders and residents held a common view that while many researchers come to study their communities, few if any obvious changes usually result. Because respondents do not receive direct benefits from their participation in the research, and may not perceive the links between the program and the evaluation even as time goes on, identifying ways to share results from the study widely would be a valuable effort.

Communicating baseline results could strengthen participants' intentions to participate in later data collection rounds or other studies, increasing information quality at the same time as it underscores USAID's commitment to its policy ideals. Adding qualitative approaches for learning from program participants as a supplement to the quantitative surveys is another potential avenue for expanding community engagement with the research.

"I don't know why, but the people in the communities that were the hardest to reach always seemed the most touched by our presence. I guess they could see the strong desire we had to get there, and if nothing else we'd been able to witness how they lived, how they were faring." (General Supervisor)

"It was really an incredible experience, it touched me...there were communities where people weren't sure about us and we were able to change their minds, and then they welcomed us with open arms, ultimately happy they hadn't been overlooked..." (Community and Services Coordinator)

"The last interview I did was the best one in the whole study. It was like tying everything up with a bow...we showed up to a household that had previously refused. When we knocked on the door I thought they might say 'we don't want anything to do with you,' but the lady agreed, and asked, 'How many of you are there? We want to invite you to have lunch with us.' It was such a meaningful way to end the project. What stuck with me was how appreciative some of the people we met were...I treasure those memories." (General Supervisor)

Looking Ahead



A driver waves from his vehicle as teams head out to work.

The release of the USAID Evaluation Policy in 2011 signaled an ambitious commitment to support research that will foster more informed program decision making around the world. Many program evaluations have already been conducted, and USAID Missions and their government hosts are continuing to include major program evaluations in their plans and identify research partners qualified to implement them. Moving forward, identifying best practices for effective collaboration between program and research implementers – especially for identifying and reaching study participants and promoting the use and dissemination of results – is likely to be at the forefront of evaluation practice for USAID and partners. Expanding the use of statistical matching techniques for comparison groups also holds substantial promise to increase the generalizability of evaluation research in scenarios where randomization is not ethical or feasible. In addition, the growth of integrated programming calls for expanded use of innovative research designs that not only identify overall program impact but also measure the added value of new cross-sectoral approaches.

The WHIP Evaluation baseline in Guatemala typifies the kinds of challenges faced in the early stages of a large, rigorous longitudinal program study: trying to identify interests that stakeholders share and leverage sustained joint investment, finding effective strategies to harness capacity for high-quality design and implementation, looking for ways to keep

studies cost efficient while also safeguarding the quality of the research products. Equally importantly, the WHIP baseline provides valuable lessons about existing strengths: the long histories of experience with population-based surveys that already position many countries for excellence in evaluation implementation, enormous international political will to strengthen the evidence base for health and development programming, and increasing assurance that significant investments in program evaluation are unambiguously backed by USAID policy.

Undoubtedly, many lessons are still to come. As WHIP interventions continue to work in the program's areas of influence, USAID Guatemala will have to make decisions about how to design and implement the evaluation's midline and final phases. Survey measures of program exposure, for example, may need to be developed as a supplement to 'intent to treat' analyses. Even before then, baseline results will need to be disseminated and building awareness and capacity through workshops or other interactive forums could promote data use. Evaluation efforts everywhere should focus on ensuring that funders are informed about their options for program studies and knowledgeable about the costs and benefits of choosing a particular evaluation model, program implementers are prepared for their important role as collaborators even when the research is externally run, and evaluation teams can anticipate and resolve common planning and implementation challenges.



Teams organize their supplies before leaving a town in San Marcos for study communities.

"We don't want...to go to great lengths to obtain this information and have it end up going straight to the library shelf, just another decoration...my message to other USAID Offices who aren't doing [evaluation] and are thinking of doing it is: take on the challenge. It's difficult, costly - but...we're already happy because this is an important evaluation that's going to help us with decision making." (M&E Advisor, USAID)

Appendix: WHIP Evaluation Performance Indicators



The region is home to mountainous terrain and striking views.

- 1 Under five mortality rate (5-year period of reference)
- 2 Under five mortality rate (10-year period of reference)
- 3 Total fertility rate
- 4 % of people living in poverty (using relevant poverty line consistent with national standard)
- 5 % of people living in extreme poverty (less than \$1.25 per day, 2005 USD-PPP)
- 6 Average per-capita daily consumption
- 7 % of children 0-59 months with chronic malnutrition (stunting)
- 8 % of children 0-59 months with acute malnutrition (wasting)
- 9 % of children 0-59 months who are underweight (for age)
- 10 % of children 6-59 months who are anemic
- 11 % of pregnant women who are anemic
- 12 % of lactating mothers who are anemic
- 13 % of reproductive-age women who are anemic
- 14 % of reproductive-age women who are underweight
- 15 % of reproductive-age women who are overweight or obese
- 16 % of children 0-59 months who are overweight or obese
- 17 % of households with moderate or severe food insecurity (using hunger scale)
- 18 % of children 6-23 months with minimum acceptable diet
- 19 Women's dietary diversity: average number of food groups consumed by reproductive-age women
- 20 % of newborns who received postnatal care within two days of birth
- 21 % of children under age five with acute respiratory infection in the two weeks prior to the survey
- 22 % of children under five with diarrhea in the two weeks prior to the survey
- 23 % of reproductive-age women who recognize at least three signs of risk in the newborn
- 24 % of reproductive-age women who recognize at least three signs of risk during pregnancy
- 25 % of reproductive-age women who recognize at least three signs of risk during childbirth
- 26 % of reproductive-age women who recognize at least three signs of maternal postnatal risk
- 27 % of pregnant women who received prenatal care from a qualified provider (doctor or nurse)
- 28 % of births with at least four antenatal visits
- 29 % of expectant mothers in the last 5 years that took at least two preparatory actions in case of emergency (secured transport, saved money, identified blood donors, identified place of delivery)
- 30 % of births attended by a doctor, nurse or qualified midwife
- 31 % of births attended by a doctor or nurse
- 32 % of children under 5 with diarrhea who received ORT
- 33 % of children under 5 with ARI treated at a community health center
- 34 % of women ages 18-24 whose first birth was before age 18
- 35 % of women 15-49 who used a modern contraceptive method the last time they had sex
- 36 % of women 15-49 with unmet need for family planning
- 37 % of children exclusively breastfed for 6 months
- 38 % of children breastfed within one hour of birth
- 39 % of households with a functioning handwashing station
- 40 % of households with a source of safe water for drinking
- 41 % of households that consider malnutrition to be a problem affecting them
- 42 % of households that consider malnutrition to be a moderate or serious problem in their community
- 43 % of children ages 12-59 months who received three doses of pentavalent vaccine
- 44 Women's Empowerment in Agriculture Index (WEAI) score

Appendix

- 45 Gender Parity Sub-Index (GPI) score
- 46 Five Domains of Empowerment (5DE) sub-index score
- 47 % of households that produce vegetables for personal consumption
- 48 % of children under 5 who received antiparasitic treatment within the past 6 months
- 49 % of children under 5 with a health card
- 50 % of children under five who received iron supplements in the past 6 months
- 51 % of children under five who received vitamin A supplements in the past 6 months
- 52 % of women who received family planning counseling as part of antenatal care
- 53 % of married or in union women currently using any method of family planning, by method used
- 54 % of women who received nutrition counseling as part of antenatal care
- 55 % of women who received hygiene counseling as part of antenatal care
- 56 % of households who have heard of USAID or recognize the USAID logo
- 57 % of households who report that USAID provides services or assistance in their community, among HH who have heard of USAID or recognize the USAID logo
- 58 % of households who list USAID when asked to name up to three international organizations providing services or assistance in their community
- 59 % distribution of facilities by level of services offered
- 60 % of facilities with micronutrient supplements available
- 61 % of facilities offering basic MCH services
- 62 % of facilities with basic medications available (by medication type)
- 63 % of facilities with vaccines available (by vaccine type)
- 64 % of facilities with contraception available (by contraceptive type)
- 65 % of facilities with medical or paramedical personnel
- 66 % of facilities with basic infrastructure (electricity, refrigeration, sterilizing equipment)
- 67 % of facilities with improved walls, floors and roof
- 68 % of facilities with personnel by type: general practitioner, obstetrician, nurse, nurse assistant, health promoter, student
- 69 % of facilities with a pharmacy on site
- 70 % of facilities with essential supplies including scales, height board, sphygmomanometer
- 71 % of facilities that have had a stock-out of any essential medications or supplies in the past 6 months
- 72 % communities with a health center, by level of service
- 73 % of communities with a bank, microfinance institutions, or saving and loan association
- 74 % of communities with a paved or gravel access road
- 75 % of communities with public transport
- 76 % of communities with public transport available every day
- 77 % of communities with piped water
- 78 % of communities with piped water considered to be of good quality
- 79 % of communities where the majority of residents have piped water access at home
- 80 % of communities with a sewerage system
- 81 % of communities with garbage disposal services (and the subset with services covering the majority of households)
- 82 % of communities with schools, by level of instruction
- 83 % of communities with at least one school offering instruction in Spanish
- 84 % of communities with at least one school offering instruction in a local (Mayan) language
- 85 % of communities where the majority of students attend school (according to community key informants)
- 86 % of communities whose nearest health center offers services 7 days a week
- 87 Average distance from communities to nearest health post
- 88 % of communities with a pharmacy or other source of pharmaceutical sales
- 89 % of communities with a formal market for the sale of commodities including food or produce
- 90 Average distance from communities to the nearest market selling consumer items