



USAID
FROM THE AMERICAN PEOPLE

ETHIOPIA

TRANSFORM: MONITORING, EVALUATION, LEARNING AND ADAPTING (TRANSFORM: MELA) ACTIVITY

FINAL IMPACT EVALUATION OF THE TRANSFORM PROGRAM

Submission Date: August 19, 2022
Revised Version: September 2, 2022
Final Version: November 1, 2022

USAID/Ethiopia
Gebeyehu Abelti, COR

Submitted by:
The Mitchell Group, Inc. (TMG)
Contract Number: OAA-I-15-00028 Task Order No.: AID-663-TO-17-00001
Transform: MELA Activity, Addis Ababa, Ethiopia

Michael Midling, PhD, Chief of Party/Principal Investigator
Email: mmidling@ethiopiadmela.com
Phone: +251- 986356914

Principal Contact:
Abi Fasosin, Senior Advisor
Email: abif@the-mitchellgroup.com
Phone: +001-202-557-9412 (Direct)

ACTIVITY INFORMATION

Activity Title	TRANSFORM: MONITORING, EVALUATION, LEARNING AND ADAPTING (TRANSFORM: MELA): ACTIVITY
Contract/Task Order Numbers	Contract No.: OAA-I-15-00028 Task Order No.: AID-663-TO-17-00001
Name of Prime Implementing Partner	The Mitchell Group, Inc. (TMG) 1816 11th Street, NW, Washington, DC 20001 Tel: 202-745-1919
Activity Start Date	March 7, 2017
Activity End Date	August 31, 2022
Period of Evaluation Implementation	September 1, 2021 – July 31, 2022

TABLE OF CONTENTS

Abstract.....	1
Executive Summary.....	1
1. Evaluation Purpose and Audience.....	5
1.1 Evaluation Purpose	5
1.2 Audience and Intended Uses	5
1.3 Evaluation Questions.....	5
1.4 Program Assumptions at Inception	6
1.5 Unanticipated Disruptions.....	6
2. Program Context and Overview	7
2.1 Program Context.....	7
2.2 Program Overview	8
3. Evaluation Methods and Limitations	10
3.1 Evaluation Methods	10
Sampling	11
3.1.1 Woreda Selection Criteria.....	11
3.1.2 Selection of Households	13
3.1.3 Sampling for Key Informant Interviews.....	14
3.1.4 Sample Size Determination	14
3.2 Data Collection Methods	15
3.2.1 Data Sources	15
3.2.2 Data Collection Tools.....	15
3.2.3 Enumerator Training and Quality Assurance.....	15
3.3 Data Analysis	16
3.3.1 Rationale and Approach.....	16
3.3.2 Outcome Measures and Covariates	16
3.3.3 Quantitative Data Analysis - Estimation Strategies.....	17
3.3.4 Assessing Relative Contributions of the Transform Program	18
3.3.5 Qualitative Data Analyses	18
3.4 Limitations and Mitigation	19
3.5 Ethical Considerations.....	20
4. Findings.....	20
4.1 Family Planning Uptake.....	21

4.1.1	DID Results	21
4.1.2	Complementary Qualitative Findings.....	22
4.1.3	Multivariate Analyses: Determinants of FP Uptake	22
4.2	Maternal Health	24
4.2.1	DID Results	24
4.2.2	Complementary Qualitative Findings.....	26
4.2.3	Multivariate Analyses: Determinants of Maternal Health Behaviors	26
4.3	Child Health and Immunizations	27
4.3.1	DID Results	27
4.3.2	Complementary Qualitative Findings.....	30
4.3.3	Multivariate Analysis: Determinants of Child Health and Immunizations	30
4.4	Cross-cutting Themes	31
4.4.1	DID results	31
4.4.2	Complementary Qualitative Findings.....	33
5.	Conclusions.....	34
6.	Recommendations.....	35
	Annex IA. Transform PHC Activity Result Framework.....	38
	Annex IB. Transform HDR Activity Result Framework.....	39
	Annex 2. Adjusted Odds Ratios: Family Planning.....	40
	Annex 3. Adjusted Odds Ratios: Maternal Health Services Utilization	41
	Annex 4. Adjusted Odds Ratios: Child Health and Immunization	42

Acronyms and Abbreviations

ANC	Antenatal Care
CBHI	Community-based Health Insurance
COR	Contracting Officer's Representatives
DID	Difference-in-Differences
DRS	Developing Regional States
EPHI	Ethiopian Public Health Institute
EQs	Evaluation Questions
FP	Family Planning
GIS	Geographic Information Systems
GOE	Government of Ethiopia
HC	Health Center
HDR	Health in Developing Regions
HF	Health Facility
HH	Household
HMIS	Health Management Information System
HP	Health Post
HSTP	Health Sector Transformation Plan
IPs	Implementing Partners
KI	Key Informant
KII	Key Informant Interview
KPI	Key Performance Indicator
LAFP	Long-Acting Family Planning
LOE	Level of Effort
MCPR	Modern Contraceptive Prevalence Rate
M&E	Monitoring and Evaluation
MNCH/FP	Maternal, Newborn, Child Health and Family Planning
MELA	Monitoring, Evaluation, Learning, and Adapting
MOH	Ministry of Health
OECD/DAC	Organization for Economic Co-operation and Development/Development Assistance Committee
PCMD	Preventing Child and Maternal Deaths
PHC	Primary Health Care
PHCU	Primary Health Care Unit
PII	Personally Identifiable Information
PMP	Performance Management Plan
PPFP	Postpartum Family Planning
RH	Reproductive Health
RMNCH	Reproductive, Maternal, Newborn, and Child Health
RHB	Regional Health Bureau
SDGs	Sustainable Development Goals
SNNP	Southern Nations, Nationalities, and People
SOW	Scope of Work
TMG	The Mitchell Group, Inc.
ToC	Theory of change
USAID	United States Agency for International Development
USG	United States Government
WHO	World Health Organization
WorHO	Woreda (district) Health Office
ZHD	Zonal Health Department

Abstract

This report outlines the findings from the impact evaluation of the five-year Transform Program supported by USAID/Ethiopia. Transform’s objective was to address maternal and child morbidity and mortality in Ethiopia, focusing on maternal, newborn, and child health, and family planning (MNCH/FP). To determine program impact, the evaluation collected baseline and endline data from Transform intervention areas and from non-Transform intervention areas. The net relative difference over time in Transform intervention areas revealed the contribution of the Transform program activities, via Difference-in-Differences (DID) analysis. The surveys included over 6,500 respondents, all women aged 15-49.

The impact evaluation focused on indicators in four different categories: family planning, maternal health, child health and immunizations, and cross-cutting issues. In the area of family planning, all three indicators (modern contraceptive use, long-acting contraception use, and postpartum family planning use) showed significant positive differences in Transform intervention areas compared to the non-Transform intervention areas. In the area of maternal health, five of the seven indicators – including metrics such as early antenatal care visits and skilled birth attendance – showed significant positive differences. The results were mixed for child health and immunizations and the cross-cutting themes. In separate multivariate analyses, higher education levels, fewer births per woman, and having received MNCH/FP messages were consistently relevant predictors of desired outcomes. The report concludes with recommendations for the Ministry of Health and USAID.

Executive Summary

In 2017, USAID/Ethiopia launched the Transform Program, a five-year integrated program to address maternal and child morbidity and mortality in Ethiopia. The program focused on maternal, newborn, and child health, along with family planning (MNCH/FP). The Transform Program operated across ten regions in Ethiopia and comprised three programmatic activities: Primary Health Care (Transform PHC), Health in Developing Regions (Transform HDR), and Water, Sanitation, and Hygiene (Transform WASH). In addition, USAID/Ethiopia awarded a contract to The Mitchell Group, Inc. (TMG) to implement the Monitoring, Evaluation, Learning, and Adapting (Transform: MELA) Activity.

Transform: MELA provided high-quality monitoring and evaluation data. This program-level impact evaluation draws on baseline, midline, and endline household (HH) surveys conducted by Transform MELA in the areas in which Transform HDR and Transform PHC operated. Transform: MELA also carried out separate performance evaluations of these two Activities. This report presents findings based on comparisons from the baseline to endline data, and it shares conclusions and recommendations aimed at improving future MNCH/FP programming in Ethiopia.

Evaluation Question: The purpose of the impact evaluation was to answer the following question: To what extent did the Transform program improve MNCH/FP outcomes in Transform intervention areas compared to non-Transform intervention areas?

Methodological Approach: To answer the question, the same survey design and methodology was used at both baseline and endline. Transform: MELA collected data for both periods from Transform intervention areas as well as from non-Transform intervention areas. Surveys used three-stage clustered random sampling and included over 6,500 respondents, all of whom were women aged 15-49. The use of difference-in-differences (DID) analysis allowed evaluators to accurately measure change over time in intervention areas, net of any change over time in non-Transform intervention areas. Prior to the analyses, respondents in the Transform intervention and non-Transform intervention samples were matched using propensity score matching based on women's age, education level, and marital status to further improve the inferential leverage of the DID analyses. These methodological approaches, supplemented by document review and key informant interviews, allowed evaluators to measure the contribution of the Transform program activities while also accounting for factors that may have affected change across subject regions more generally.

Key Exclusion Criteria: The quasi-experimental nature of the surveys improved on simple cross-sectional approaches, but additional steps were necessary to account for potential sources of bias. The study excluded woredas facing security threats as well as those in major towns in both Transform intervention and non-Transform intervention areas. Comparison woredas were excluded if they share a border with a Transform intervention woreda to mitigate spillover effects. In the case of Transform PHC (TPHC), woredas in intervention and non-intervention areas were excluded if other partners intervened in the same woreda on more than two out of five thematic areas (maternal health, newborn health, child health, immunization, and family planning).

Limitations: Despite the study's design approach to mitigate threats to the comparison of baseline and endline data, challenges and limitations did arise. Notably, the COVID-19 global health pandemic emerged around the midpoint of the program activities, posing serious challenges to the health system in general and the progress on Transform indicators specifically. War in Ethiopia also broke out during the period of performance, and various smaller scale conflicts in different parts of the country further complicated the delivery and uptake of services. In addition, the Transform program operated in a context with multiple actors addressing similar objectives, so attribution was complicated. Spillover effects remained a challenge despite the study's design. Capacity strengthening support given at national, regional and zonal levels benefited not only the intervention areas but also extended to the non-intervention areas.

Findings: The impact evaluation focused on indicators in four different categories: family planning, maternal health, child health and immunizations, and cross-cutting issues.

Family Planning (FP)

- Modern contraceptive prevalence rates showed a net difference of 6.1 percentage points in favor of Transform intervention areas compared to the change over time in non-Transform intervention areas.
- The use of long-acting contraceptive methods also showed a significant positive difference, of 3.8 percentage points.

- The use of postpartum family planning increased by 8.8 percentage points, relative to the non-Transform intervention areas.
- In multivariate analyses, older and more educated women were generally more apt to take advantage of family planning options.

Maternal Health

- A statistically significant positive net difference occurred in Transform intervention areas, relative to non-Transform intervention areas, on the following metrics: early antenatal care visits, essential antenatal care, iron folic acid supplementation, at least four ANC visits, and skilled birth attendance.
- Differences in women who made at least one ANC visit and those receiving early PNC care were positive but not statistically significant.
- In multivariate analyses, education and fewer children were consistent predictors of better maternal health, as was receiving the program's MNCH message. Notably, women whose spouses accompanied them on ANC visits were nearly 32 times more likely to use skilled birth attendants.

Child Health and Immunizations

- The Transform intervention areas saw statistically significant positive differences on the following indicators over time, relative to the non-Transform intervention areas: full vaccinations, vitamin A supplementation, fever treatment, and exclusive breastfeeding.
- The program did not result in significant positive differences in early PNC for newborns, essential newborn care at health facilities, ARI treatment, diarrhea treatment, and early breastfeeding.
- In multivariate analyses, the mother's education level was again a strong predictor of improved uptake of child health services, as was receiving the MNCH message.

Cross-cutting Themes

- In Transform intervention areas, a statistically significant net difference of 2.5 percentage points was observed in access to improved sanitation facilities.
- The share of respondents using appropriate water treatment was the only indicator that declined significantly, by 2.6 percentage points.
- No statistical differences were detected regarding handwashing stations at the home, women's participation in their own health care decisions, spouses accompanying women on ANC visits or deliveries, and membership in the community-based health insurance program.

Recommendations: The report includes several recommendations that may inform future programming for both the Ministry of Health (MOH) and USAID/Ethiopia. They include the following:

Recommendations for MOH

- Revitalize community health programs.
- Institute mobile services policies and guidelines for developing regions.
- Strengthen and standardize mobile health service delivery to expand access to MNCH/FP services for pastoralist communities.

- Develop and implement health performance-based standards and budgeting for woredas.
- Particularly in developing regions, assign more female providers for childbirth services.
- Integrate FP counseling across all contacts in the continuum of care.
- Ensure availability of family planning services at health posts.
- Ensure sustainability of in-service training activities.
- Recruit and retain additional health facility personnel to ensure consistent provision of services, especially in difficult-to-reach areas.

Recommendations for USAID

- Design future activities to be performance-based and continue to include flexible target setting, monitoring methodology, and assumptions in order to adapt to changing circumstances. Revise targets as appropriate during the course of the program.
- Develop an operational research agenda to identify causes of service bottlenecks and explore the underlying factors contributing to the decline in performance of selected KPIs.
- Establish institutionalized mechanisms for field-level sharing of information between primary health interventions and interventions targeting pastoralist regions and communities.
- Strengthen the support to narrow the gap between effective coverage and contact coverage.

I. Evaluation Purpose and Audience

I.1 Evaluation Purpose

The purpose of the impact evaluation was to measure the contributions of the Transform program in improving MNCH/FP outcomes in Ethiopia. It aids USAID in determining the effect of the Transform Program on specific outcomes of interest related to FP, maternal health, child health and newborn health, and other cross cutting dimensions. In doing so, it tests USAID's development hypotheses by comparing changes in outcomes of interest to what would have happened in the absence of Transform interventions, which the evaluation treats as the counterfactual. This impact evaluation used comparison groups, referred to as 'non-Transform intervention areas' – where intervention activities were not implemented – and treatment groups, referred to as 'Transform intervention areas,' where Transform interventions were implemented. The comparison between the outcomes of interest in the Transform intervention and non-Transform intervention areas created the basis for determining the impact of the Transform Program interventions over the period of performance. The impact evaluation thus helps demonstrate attribution to the Transform Program interventions by showing the change from what would have occurred in its absence.

I.2 Audience and Intended Uses

The principal audience of this impact evaluation report is the USAID/Ethiopia mission, whose resources were invested to effect positive change in MNCH/FP outcomes. In addition, the report is intended to serve the Health Office, the MOH authorities, Regional Health Bureaus (RHBs) and other country-level partners, and healthcare providers. For these audiences, the report also provides recommendations that can inform new policies and MNCH/FP programs in addition to the findings.

I.3 Evaluation Questions

The program evaluations for the Transform: PHC and Transform: HDR activities addressed the following questions:

1. How effective were the Transform PHC Activity approaches in contributing to improving MNCH/FP outcomes? What were the drivers of the observed changes? What constraints affected the achievements?
2. How and to what extent did Transform PHC Activity facilitate local ownership, sustainability, and coherence?
3. How did Transform PHC consider gender dynamics in activity design and implementation?
4. How and to what extent has the Transform IP been able to adapt their interventions in response to learnings and new evidence?

The impact evaluation that represents the focus of this report focused on a fifth evidence-based evaluation question (EQ): To what extent did the Transform program improve MNCH/FP outcomes in Transform intervention areas compared to non-Transform intervention areas?

That EQ was further served by two specific sub-questions:

1. To what extent did the Transform program ensure improvement in MNCH/FP outcomes?

2. Are there any unintended results (positive or negative) specific to Transform program interventions?

1.4 Program Assumptions at Inception

The Transform Program was conceptualized following USAID/Ethiopia's implementation of a portfolio of activities focused on Ending Preventable Child and Maternal Deaths in 2015. At the inception of the program in 2017, Transform implementers relied on four core assumptions:

GOE capacity and cooperation: The Transform Program relied on the assumption that the GOE would act as a reliable partner with shared goals to improve maternal and child wellbeing. That assumption was well-founded, given the important progress the GOE had made in the prior decade to reduce both child and maternal mortality. Importantly, Transform program implementers needed open communication channels with counterparts in the MOH, and they needed to rely on GOE representatives to facilitate access to and operations in remote areas. Critically, the Transform Program assumed the presence of highly capable leadership committed to continued improvements.

Local partner capacity: A major component of the Transform Program was to utilize local organizations to complete the required tasks, services, and deliverables, while simultaneously building their M&E capacity and capability to work with USAID. The implementation team thus assumed that the local partners who contributed to the Transform Program would not be limited to data collection or data entry but would also engage in programmatic activities and high-level analyses. The roles and responsibilities of the local organizations were also expected to increase throughout the period of performance commensurate with the level of capacity attained to achieve results.

Cause and effect logic: Transform Program activities and expected outcomes were based on the logic that people behave rationally in response to the conditions they face, that systems tend to be organized in logical ways, and that investments in inputs should lead to predictable outputs. Thus, the theory of change linked activity plans to higher level program objectives with clear steps. Nevertheless, the Transform Program also assumed a complex operational environment in which people and circumstances can be unpredictable because of unanticipated environmental and contextual changes. For example, the implementation team assumed that other donors may operate in Transform intervention areas and thus complicate the ability to contribute effects to the Transform activities.

Political stability: At the inception of the program, Ethiopia was on the verge of a political transition that would end conflict with Eritrea and usher in new hope for long-term political stability. USAID/OTI launched democracy-strengthening activities in Ethiopia, and progress for women, youths, and human rights appeared imminent. The Transform Program thus assumed a salubrious political environment for making progress on important child and maternal health initiatives.

1.5 Unanticipated Disruptions

COVID-19. The Covid-19 global health pandemic emerged in the middle of the Transform Activity period. The initial spread of COVID-19 was not as severe as many had feared: in Ethiopia, 1,054 COVID cases per million were reported, and only 16 people per million had died from the condition

as of 2020.⁷ However, many KIs noted that by disrupting the health system, the pandemic risked creating devastating consequences on the public health system.⁸ A reduction in the utilization of essential health services could have long-term effects by increasing preventable morbidity and mortality.⁹ Some of the COVID-19 prevention measures taken in Ethiopia showed signs of adversely affecting the utilization of essential health services. Prior to the pandemic, health care utilization and health outcomes were already relatively poor in Ethiopia compared to other countries. For example, only 48% of women gave birth in health facilities at that time, with a neonatal mortality rate of 30 per 1,000 live births.¹⁰ Restrictions of movement, the conversion of selected health facilities into COVID-19 treatment centers and the redeployment of health workers to COVID-19 care likely further reduced access to and provision of essential health services, thereby adversely affecting population health.¹¹ Moreover, due to limited awareness and a lack of proper preparedness of health providers, some community members likely feared exposure to the virus while visiting health facilities and thus avoided or delayed needed care.

Conflict. Health systems in fragile and conflict-affected states often struggle to provide basic health services. Apart from direct casualties, war and civil strife disrupt health care delivery and thereby increase morbidity and mortality.¹³ Ethiopia was impacted by war, communal violence, and instability over the last two years of the Transform Activity. Most of the regions covered by the Activity experienced some form of disruption due to conflict and instability. Conflict and violence across the intervention areas disrupted the movement of ambulances; health workers were displaced; health facilities were looted; and in the worst instances, the infrastructure of health facilities was destroyed, causing Transform activities to cease operations on many occasions.

Climate and disease setbacks. Recurrent droughts, flooding, and disease outbreaks created disruptions in various parts of the country during the period of performance. For example, drought in the Somali region was pronounced, and recurrent flooding occurred in Afar and Somali. Multiple vaccine-preventable diseases and other disease outbreaks occurred in several regions.

2. Program Context and Overview

2.1 Program Context

The Government of Ethiopia (GOE) has made important progress in improving access to basic health services for Ethiopians. While there have been significant improvements, the population as a whole still has limited access to clean water and sanitation and quality health services, and it continues to be characterized by low levels of literacy and persistent food insecurity. These factors, among others, contribute to a high incidence of communicable diseases including TB, HIV and AIDS, malaria, neglected tropical diseases, and respiratory infections, as well as nutritional deficiencies, maternal, neonatal, and child mortality. Particularly relevant to the Transform Program are shortcomings in child health, maternal health, and FP.

Child Health: Ethiopia has made important progress since 2005; under-five mortality has decreased by 52 percent from 123 to 59 deaths per 1,000 live births (MDHS 2019). However, little progress has been made in reducing neonatal death rates, which now account for 33 deaths per

1000 live births but account for approximately half or more of under-five child mortality cases. Malnutrition rates also continue to be very high, with 37 percent of children under age five stunted and 7 percent wasted (MDHS 2019).

Maternal Health: The GOE states that it has significantly reduced the Maternal Mortality Ratio (MMR) and reached MDG 5. According to DHS 2016 data, Ethiopia still has one of the highest rates of maternal mortality in the world at 412 per 100,000 live births. Access to and utilization of quality MNCH services, including skilled birth attendance, remains limited, especially in rural areas. Only 50 percent of births are delivered by a skilled provider. Although this figure has improved over the past decade (from 11% in 2011 and 28% in 2016), based on current population estimates, an estimated 20,000 women still die from childbirth-related causes every year. Additionally, it is estimated that over 37,000 women currently suffer from obstetric fistula. Seventy-four percent of women who gave birth in the five years preceding the survey received antenatal care (ANC) from a skilled provider for their most recent birth, compared to 62 percent in 2016 (MDHS 2019). Thus, progress is evident in recent years, though important shortcomings persist.

Family Planning: The proportion of married women using modern FP methods increased steadily over the last two decades. According to the 2019 MDHS, the number of married women using modern FP methods increased from 6 percent in 2000 to 41 percent in 2019.

2.2 Program Overview

In 2017, USAID/Ethiopia launched the Transform Program, a five-year (2017–2021) integrated program to address maternal and child morbidity and mortality in Ethiopia. The program focused on maternal, newborn, and child health, along with FP (MNCH/FP), and it includes a water, sanitation, and hygiene (WASH) component. The Transform Program operated across ten regions in Ethiopia and comprises three programmatic activities: Primary Health Care (Transform: PHC), Health in Developing Regions (Transform: HDR), and Water, Sanitation, and Hygiene (Transform: WASH).¹

In addition to the three core program activities, USAID/Ethiopia awarded a contract to The Mitchell Group, Inc. (TMG) to implement the Monitoring, Evaluation, Learning, and Adapting (Transform: MELA) Activity. The role of Transform: MELA was to provide and synthesize high-quality Monitoring and Evaluation (M&E) data for the Activity implementing partners, USAID/Ethiopia and the MOH of Ethiopia and to guide them in learning and adaptive health system management. Transform: MELA also conducted final performance evaluations of the Transform: PHC and Transform: HDR Activities, the Transform Program Impact Evaluation, and numerous selected case studies to document important lessons learned that highlight the contribution of the Transform program. The evaluation data collection took place from May to July 2022.

This report details the findings from the Transform Program Impact Evaluation. It outlines the objective and audience and the methods for measuring the contributions of the Transform Program

¹ Beginning from September 2021 the Transform program was implemented in ten regions. This evaluation however will not include a separate analysis for the newly formed South-West Ethiopia region.

in improving MNCH/FP outcomes. It presents findings based on comparisons from the baseline to endline data collection activities, and it shares conclusions and recommendations aimed at improving future MNCH/FP in Ethiopia.

The two programmatic activities whose impact is evaluated in this report targeted complementary aspects of maternal and child wellbeing:

The Transform: PHC Activity was implemented by Pathfinder International (prime), John Snow International, Encompass, Abt. Associate, Ethiopian Midwife Association, and Malaria Consortium. Transform: PHC aimed to contribute to preventing child and maternal deaths (PCMD) by supporting the implementation of the Health Sector Transformation Plan (HSTP)-Government of Ethiopia (GOE) at different levels of the health system. Transform: PHC focused on four high-level intermediate results (IRs):

- 1) Improved management and performance of health systems;
- 2) Increased sustainable quality of service delivery across the primary health care unit's (PHCU) continuum of care;
- 3) Improved household (HH) and community health practices and health-seeking behavior;
- 4) Enhanced program learning, and impact policy and programming related to preventing child and maternal deaths (PCMD).

MNCH/FP, reproductive health (RH), and malaria were the primary intervention areas of Transform: PHC. The activity operated in five major regions in the country– Amhara, Oromia, Southern Nations, Nationalities, and People (SNNP), Sidama, and Tigray – and targeted a total of 434 woredas over the five-year lifespan of the program. However, due to the recent conflict, Tigray region was dropped from the impact evaluation.

The Transform: HDR Activity aimed to increase utilization of quality, high-impact MNCH/FP services in the developing regional states (DRS) of Ethiopia – Somali, Afar, Gambella, and Benishangul-Gumuz. To achieve its ultimate goal, Transform: HDR focused on four intermediate results:

1. Increased access to integrated, quality, high-impact MNCH/FP services at health facilities (HFs) and community settings;
2. Strengthened health system that provides quality MNCH/FP services;
3. Increased demand for high-impact MNCH/FP services; and
4. Improved use of information for evidence-based decision making and program learning.

Transform: HDR included, among other services, the provision of medical equipment, building capacities of the health workforce, supporting woreda health offices (WorHO) with the organization of comprehensive mobile outreach services, strengthening the provision of one-stop service to victims of gender-based violence (GBV), use of performance improvement interventions and strengthening implementation of health management information systems (HMIS).

The two programs under evaluation continued through 2021. Adjustments were made in implementation protocols to account for several emerging factors including the Covid-19 health

pandemic and conflict and security risks that intensified during the period of program implementation.

The Transform: MELA Activity was designed as a key component of the Transform Program to provide Monitoring, Evaluation, Learning, and Adapting support and to be responsible for implementing rigorous review, monitoring, and evaluation (M&E) of the Transform Program over the life of the three Transform Activities (PHC, HDR, and WASH). Specifically, Transform: MELA was responsible for evaluating the effectiveness of the programmatic activities in achieving their objectives. In 2018, Transform: MELA completed an extensive baseline survey that included information from over 6,595 households (HHs), 119 health facilities (HFs), and key informant interviews (KIIs) across the eight regions, with a 4:1 ratio of responses from Transform intervention² and non-Transform intervention areas. The baseline survey provided measurements according to key program indicators and established a foundation against which progress could be measured at subsequent points in the life of the program. The Transform program impact evaluation explored changes in program outcomes using a comparison of results from the baseline.

3. Evaluation Methods and Limitations

3.1 Evaluation Methods

The impact evaluation of the Transform program employed a quasi-experimental, pre-post non-equivalent comparison group evaluation design. The design of the impact evaluation was prospective in the sense that it was developed concurrent with the initial Transform Program design and was built into program implementation. The baseline evaluation was conducted in 2017, before the program implementation commenced, to establish measures for the outcomes of interest. The Transform intervention and non-Transform intervention areas were then identified, with non-Transform intervention areas acting as the comparison group. This design facilitated the measurements that can reveal Transform program contributions.

Data was collected at the baseline stage (pre activities) and the endline stage (post activities) from both Transform intervention and non-Transform intervention woredas; they were non-equivalent in that the Transform intervention and non-Transform intervention woredas were not selected by pure randomization. The change over time in intervention woredas – net of any change over time in non-Transform intervention woredas – revealed the contribution of the Transform program activities while accounting for factors that may have effected change across the region more generally (in intervention and non-Transform intervention woredas), notwithstanding the limitations addressed below. A mixed-methods approach, with both quantitative and qualitative elements, was used to gather and analyze data from a wide range of sources relevant to answering all the impact EQs.

² A Transform intervention woreda is defined as a woreda in which the USAID Transform Program implemented through Transform PHC and HDR Activities.

Sampling

3.1.1 Woreda Selection Criteria

Transform: PHC intervention woredas were classified as high, moderate, and low performing based on the results of the Greatest Impact Assessment (GIA),³ which used a select set of 12 MNCH/FP performance indicators from the 2016 Health Management Information System (HMIS) prior to the Transform Activities interventions. The purpose of the GIA I was to categorize woredas based on their RMNCH performance data and prioritize woredas in greatest need and with the highest potential for future evidence-based interventions. The baseline survey covered woredas across the three classifications.⁴ For comparability, the final evaluation drew upon a similar proportion of woredas in each of the three categories.

For Transform: HDR, the classification of woredas as high, moderate, and low performing was not used, due to the fact that a GIA was not conducted in the Afar, Benishangul-Gumuz, Gambella and Somali regions. Overall, woredas were selected by considering the following inclusion and exclusion criteria.

Inclusion Criteria for Intervention Woredas

- a) Kebeles in Transform intervention woredas were included in the sampling frame for the impact evaluation if there were no security risks at the time of sample selection. Across Transform intervention regions,⁵ kebeles in 97 woredas⁶ that have security threats were excluded from the sampling frame.
- b) To ensure greater comparability in a context in which other financial and technical partners not associated with the Transform program also intervene, criteria were imposed to account for external (non-Transform) programming. For Transform: PHC targeted regions, woredas were eligible if non-Transform partners intervened in the same woreda on a maximum of two out of five possible thematic areas (maternal health, child health, neonatal health, immunization, and FP).⁷ For Transform: HDR targeted regions, woredas were eligible irrespective of the number of thematic areas that other partners addressed, due to a different set of circumstances in the HDR regions.
- c) The same woreda classification criteria and proportions were used for Transform PHC intervention woredas.

³ Source: USAID GIA analysis summary (2016). Transform: MELA used the woreda performance categorization in the GIA analysis to ensure that there will not be sample selection bias for the endline evaluation between Transform program intervention and non-Transform intervention woredas.

⁴ Performance categorization of Transform PHC intervention surveyed woredas at baseline: High=14.3%, Moderate=50.8%, and Low=34.9%.

⁵ Tigray region is excluded from the evaluation.

⁶ Number of intervention woredas by region excluded because of security concerns: Amhara-6, Oromia-52, SNNP-21, Somali-1, BenishangulGumuz-8, Afar-5).

⁷ Transform: MELA mapped the program activities of approximately 12 external partners to make these determinations.

Inclusion Criteria for Non-Transform intervention Woredas

- a) Non-Transform intervention woredas (comparison woredas) that do not share a border with Transform intervention woredas were included to minimize contamination.
- b) The same woreda classification criteria and proportions for Transform intervention woredas were used for non-Transform intervention woredas in Transform: PHC regions. For non-Transform interventions, proportions similar to the baseline⁸ were used to rule out measurement errors in the final evaluation.

Exclusion Criteria for Intervention and Non-Transform Intervention Woredas

- a) Woredas that faced security threats were excluded from the sampling frame.⁹
- b) Major non-intervention town administrations were excluded from the sampling frame.¹⁰
- c) Woredas with no available data on non-Transform partner interventions were excluded. In addition, PHC targeted regions in which non-Transform partners intervened on more than two of the five thematic areas noted above were excluded.

Woredas in the sampling frame: After applying the inclusion and exclusion criteria, 136 Transform intervention and 84 non-Transform intervention woredas remained in the sampling frame for Transform: PHC targeted regions. From this sampling frame, a total number of 80 intervention and 23 non-Transform intervention woredas were selected in the PHC regions. Although no non-Transform partner data were available in Sidama, woredas in the region were also considered in the sampling to make the result comparable with the baseline evaluation and due to the fact that the region was not formed at the baseline.

Similar to the Transform: PHC regions, 45 Transform intervention and 104 non-Transform intervention woredas remained in the sampling frame for Transform: HDR targeted regions. From that sampling frame, 36 intervention and 16 non-Transform intervention woredas were selected for inclusion.

Replacement Protocols

During the course of data collection for the impact evaluation, some woredas were not accessible, due in most circumstances to security-related issues, flooding, and inaccessibility. In these instances, the evaluators sampled with replacement from other Transform intervention and non-Transform intervention woredas, respectively, depending on the performance classification of replaced woredas. In total, two woredas were replaced during the actual data collection.

⁸ Performance categorization of non-Transform PHC intervention surveyed woredas at baseline: High=13.4%, Moderate=53.3%, and Low= 33.3%.

⁹ The number of non-Transform intervention woredas by region excluded due to security concerns: Amhara-3, Oromia-53, SNNP-10, Somali-0, Gambella-0, BenishangulGumuz-8, Afar-5.

¹⁰ Exceptions included Gambella, Jigjiga and Dubti.

3.1.2 Selection of Households

The sampling technique for the endline household (HH) survey replicated the sampling technique used at baseline. This approach allowed for a reasonable comparison of data between the baseline and the final evaluations.

The HH survey employed a three-stage clustered sampling technique. The first stage used a random selection of kebeles (wards) in the Transform intervention and non-Transform intervention areas as primary sampling units, drawn from the list of kebeles in the respective regions stratified by administrative zones and performance before 2017. Thus, woredas in which the randomly selected kebeles were located were automatically included in the evaluation. The second stage of the sampling entailed the simple random selection of gotts (neighborhoods) from the selected kebeles. Because of the difficulties and time-consuming exercise of developing a fresh list of households in a kebele, which could take more than two days to list households in the selected gotts, 2-3 gotts were selected randomly from each kebele, generating an average list of 150 households from those selected gotts.

The third stage of the sampling selected HHs from the randomly selected gotts. A list of households from each of the selected gott served as a frame to select 30 HHs per kebele. The key eligibility criterion for selecting the HHs was the availability of a woman aged 15-49 in the HH, regardless of marital status, who had residential status there for at least 6 months prior to the date of data collection. If more than one such eligible respondent resided in a selected household, the enumerator randomly selected one respondent among the eligible women. Figure I summarizes the sampling procedures used for both the baseline and endline data collection efforts, which together informed the impact evaluation.

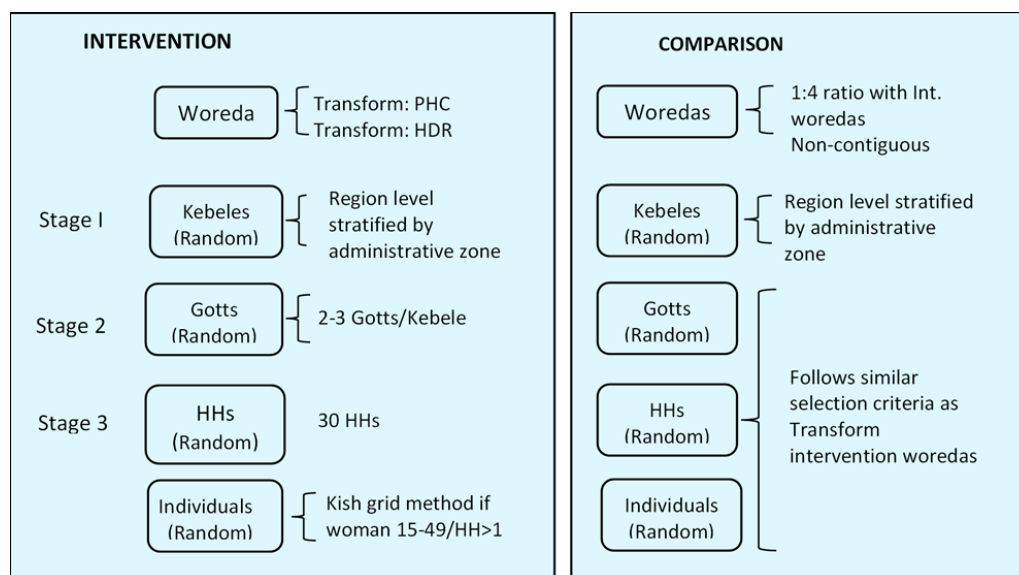


Figure I. Sampling Procedure for Transform Intervention and Non-Transform intervention Areas

3.1.3 Sampling for Key Informant Interviews

While the impact evaluation relied primarily upon pre- and post-activity comparisons of indicator outcomes, it was also useful to draw on information from KIIs (KIs) to add insight to the findings. Interview subjects for KIIs were enrolled using purposive sampling, which emphasized heterogeneity among those individuals to ensure a variety of perspectives. Key informants (KIs) were drawn from USAID, central and regional implementing partners (IPs), relevant directorates at MOH, RHBs, selected Zonal Health Department (ZHDs), selected WorHO, and selected HFs.

3.1.4 Sample Size Determination

Household Survey: the sample size was computed separately for the Transform: PHC and Transform: HDR Activities. Three key MNCH/FP indicators—modern contraceptive prevalence rate (MCPR), skilled birth attendance (SBA), and full vaccination coverage—were considered to calculate the sample size. A two independent population proportions sample size estimation technique was applied using G*Power 3.1.9.7 software. The sample size for each region was determined using power calculations with a 95% confidence interval, 80% statistical power, a Transform intervention to non-Transform intervention woreda allocation ratio of 4:1, and a non-response rate of 2% (the maximum non-response rate identified at baseline). Effect sizes for each of the indicators were estimated by taking the average changes observed in the Ethiopian Demographic and Health Survey (EDHS) between 2011 and 2016 and the Transform program life of project targets. Consequently, the optimum sample size was obtained using MCPR with an effect size of 7 percentage points, drawn from Transform program targeted regions. The calculated sample sizes were finally multiplied by the average regional design effects taken from EDHS 2016. The calculation yielded a total of 6,617 HHs with women aged 15-49 (5,294 from Transform intervention and 1323 from non-Transform intervention areas). Table 1 summarizes the sample sizes across intervention and non-Transform intervention areas.

Table 1: Sample Size for Transform Intervention and Non-Transform Intervention Areas in Transform: PHC and Transform: HDR

Category	P ₁ (Non-Transform intervention at end line)	P ₂ (Transform intervention at end line)	Transform intervention areas Estimated HH sample size	Non-Transform intervention woreda Estimated HH sample sizes	Design effect (2016 EDHS regional average)	Non-response (estimated from baseline evaluation)	Transform intervention woreda Total estimated # of HHs	Non-Transform intervention areas Total estimated # of HHs
Transform: PHC targeted regions	49	56	1567	392	1.72	0.02	2727	682
Transform: HDR targeted regions	28	35	1373	343	1.85	0.02	2568	641
Total			2940	735			5294	1323

Key Informant Interviews: To obtain qualitative information to substantiate the quantitative findings, a total of 54 KIs were conducted with respondents from MOH, RHB, ZHD, WorHO, Transform program IPs, and USAID/Ethiopia.

3.2 Data Collection Methods

3.2.1 Data Sources

The impact evaluation relies primarily on the baseline and endline HH surveys. Information from KIIs complement the data produced by those pre- and post-activity comparisons. Recommendations were further informed by document review, including material from IP performance reports, Mini-Demographic and Health Survey reports (MDHS), and Transform baseline and midterm evaluation reports.

3.2.2 Data Collection Tools

The Household survey was conducted using pre-tested, structured questionnaires. The structured questionnaires were uploaded to smartphones and the HH data were then collected electronically using Mobile phones/Tablets. KIIs were conducted using semi-structured interview guides. In addition, Gott lists, HH listing, and other forms were used to guide the sampling and the data collection process.

The **household survey at both the baseline and endline stages** included sections on:

- Household and women’s demographics, household living conditions
- Enrollment in community-based health insurance scheme
- Household decision-making practices
- Health service uptake including FP, ANC, delivery and postnatal care (PNC), and newborn health child immunization and child health services

3.2.3 Enumerator Training and Quality Assurance

Enumerators who administered the HH surveys at both the baseline and endline stages had at least a BA/BSc degree, were fluent in both the local language and English, had prior experience collecting similar data, and understood the culture and traditions of the communities in which they conducted surveys. Survey coordinators, supervisors, and qualitative data collectors had at least an MA/MSc in the health or social science fields and previous experience undertaking similar activities. Given that the HH respondents were women of reproductive age, and to ensure that they could speak comfortably with enumerators, the field teams included only female enumerators in all locations.

A three-day training-of-trainers (ToT) was provided to field coordinators and supervisor. The trained coordinators and supervisors in turn provided a five-day training to enumerators with close supervision and support from Transform: MELA technical staffs. The training covered topics such as research ethics, rights of human subjects during research, sampling procedures, informed consent, data collection tools, interviewing techniques, data handling, data security and quality, and gender considerations during data collection. The structure of the training included a review of survey instruments and role-play. Once the data collectors completed the training on the paper-based data collection tools, they received training on how to collect data using the electronic data collection template and how to upload the collected data onto the server.

The survey instruments were pre-tested by data collectors in the field in all languages at the end of the training and before the commencement of the actual data collection, to confirm question comprehension, to understand the sampling procedures, and to test operability of the electronic data collection template.

Before traveling to each of the selected woredas, supervisors and coordinators communicated with regional officials and local leaders about the evaluation. To ensure the quality of collected data, supervisors conducted spot-checking and reinterviewing, especially during the initial phase (first three days) of the data collection for the baseline and endline surveys. Supervisors also reviewed a sample of completed questionnaires daily before uploading them onto the server. All KIs were audio-recorded, and interview notes were translated to English. Five percent of the transcriptions were checked against the audio file for accuracy.

3.3 Data Analysis

3.3.1 Rationale and Approach

Transform: MELA tracked changes in key outcomes through performance monitoring over the course of the program activities, but comparing data from performance indicators against baseline values demonstrates only whether a change occurred; it does not establish what caused the observed change. Due to the potential effects of confounding factors such as natural events, external donor interventions, or changes in government policy, a simple comparison of outcomes in Transform intervention and non-Transform intervention areas would not allow USAID/Ethiopia to claim that their interventions caused the observed changes or results. In some cases, the Transform intervention areas receiving USAID/Ethiopia's assistance through Transform Program may have improved while outcomes in non-intervention areas may roughly stay the same or declined. In other situations, the Transform intervention areas may have already been improving, and the intervention helped to accelerate that positive change. Furthermore, the intended outcomes in the Transform intervention areas may appear to remain the same or deteriorate, but non-Transform intervention areas may have fared even worse. The impact evaluation aimed to identify the effects of the intervention of interest in all these cases, where both the Transform intervention and non-Transform intervention areas may have changed, but at different rates. By identifying the effects of the Transform Program based on differential changes over time in Transform intervention versus non-Transform intervention areas, and deciphering the reasons behind these changes, the impact evaluation aimed to help USAID/Ethiopia and key stakeholders learn which approaches and activities are most effective. This is critical for determining future development programming and resource allocation.

3.3.2 Outcome Measures and Covariates

Outcome measures: The impact evaluation was based on the results frameworks for both the Transform: PHC and HDR Activities. The results frameworks identified key performance indicators, which contributed to the design of qualitative and quantitative data collection tools. A result-indicator matrix links the key performance indicators (KPIs) to measurable results areas identified in the results frameworks of the respective Transform Activities. KPIs considered in this analysis include FP indicators (modern contraceptive prevalence rate (MCPR), long-acting contraceptive

methods use, and postpartum family planning (PPFP) use); maternal health indicators (ANC, essential ANC, SBA, early PNC, and iron-folic acid supplementation—IFS); child health (early PNC, essential newborn care (ENC), vitamin A supplementation, full vaccination, early breastfeeding, exclusive breastfeeding, fever and diarrhea treatments); and cross-cutting issues (access to basic sanitation, handwashing station with water and soap/ash, appropriate water treatment techniques, women's participation in decisions regarding their own health care, mothers accompanied by their spouse during ANC visits and delivery, and community-based health insurance—CBHI). These KPIs were used in the measurement of achievement of intermediate results, as included in the results frameworks of the two Transform Activities (Annex 1a and Annex 1b).

Covariates: The following covariates – variables that could also correlate with and potentially predict the use of MNCH/FP services – were included in the analysis: women's age, women's educational level, women's marital status, children ever born, exposure to FP or reproductive health and MNCH messages in the last few months prior to the survey, and enrolment in CBHI. In addition, women's empowerment indicators, such as the extent to which women participate in their own health care decisions, spouses accompanying the mother during ANC visits, and spouses accompanying during delivery were included as covariates. Subsequently, covariates including women's age, women's educational level, and women's marital status were used to match the Transform intervention and non-Transform intervention areas at the baseline and endline for propensity score matching analyses.

3.3.3 Quantitative Data Analysis - Estimation Strategies

The principal means of estimating the contributions of the Transform program on MNCH/FP outcomes was the estimation of effects versus counterfactuals using a Difference-in-Differences (DID) analysis. Similar data collection procedures were employed at the baseline and endline stages of the program. Transform: MELA collected the endline data from Transform intervention and non-Transform intervention areas to understand whether any systematic differences exist across those woreda types, as was done at baseline before the Transform interventions. The DID analysis – depicted in Figure 2 – then generated the difference in outcomes in Transform intervention areas over time (from baseline to endline) while accounting for changes in the non-Transform intervention areas over the same period. The latter represented the counterfactual since any changes in those areas were assumed not to have been a result of program interventions. The difference in those baseline-endline differences was the quantity of interest for each indicator.

To reduce selection bias, the DID was estimated by restricting the analysis only to the matched sample (using propensity score matching) at the baseline. The matched sample relied on the control covariates – including women's age, education level, and women's marital status – to match individuals in the Transform intervention and non-Transform intervention areas. In other words, individuals in the Transform intervention category were not compared to all other individuals in the non-Transform intervention areas; they were only compared to individuals with similar demographic characteristics. The evaluation team also excluded residents of large towns, given the differences in typical livelihood patterns in those locations. Once matched subjects were identified, the difference in proportions for each outcome was calculated between the Transform intervention and non-Transform intervention areas, and then the DID for key outcomes was calculated between the

baseline and the final evaluation.¹¹ A p-value of less than 0.05 was used to declare statistical significance in the interpretation of results.

3.3.4 Assessing Relative Contributions of the Transform Program

It is important to note that the Transform program was implemented in a complex and dynamic situation in which a range of contextual factors may have contributed to—or conversely, impeded—outcomes. As an important example, activities undertaken by other development partners in non-Transform intervention areas could have mitigated the perceived impact of the program, despite real improvements. Likewise, effective activities by other partners in Transform intervention areas could have created the appearance of greater benefits from Transform than should be attributed to the program. Additionally, factors such as the recent COVID-19 pandemic and political or security risks, which were outside the control of the Transform program, could have affected Transform intervention and non-Transform intervention areas in ways that disguise the true impact of the Transform activities. For these reasons, not all outcomes—either positive or negative—in Transform intervention areas can be directly and solely attributable to those interventions.

Although the evaluation could not completely eliminate effects due to other development partners, efforts were made to minimize those effects in the impact evaluation design using partner mapping information that includes details on development actors at the various levels (region, zone, woreda) and the type of MNCH/FP interventions they provide. By restricting the analyses to those woredas that received external interventions in a maximum of two out of five thematic areas, particularly in Transform: PHC targeted regions, the DID estimations minimized the potential bias from this important source.

Data collected from KIIs also added important qualitative nuance to the evaluation in terms of the comparative contributions of the Transform program. These data, particularly responses to questions that specifically ask KIIs to comment on the relative contributions of the Transform program and other MNCH/FP development partners, served as a valuable resource for contextualizing the Transform program's contributions to MNCH/FP outcomes.

3.3.5 Qualitative Data Analyses

Data Management: The qualitative data analysis team first evaluated the data collected through interview reports for completeness and accuracy, verifying against recorded audio versions. The data was then labeled using the KI type (IP, USAID, MOH, Region, Zone and Woreda stakeholder) and the activity type (Transform: PHC, Transform: HDR, and Non-Transform Woredas).

Coding: A codebook was developed based on evaluation questions and the interview guides. In addition, portions of the interview transcripts were reviewed, and the codebook was further refined as specific themes emerged. Next, the codes were uploaded to the qualitative analysis software

¹¹ Propensity score matching was used during the final impact evaluation to ensure comparability of study subjects between the baseline and endline by merging data from the two HH surveys. After merging these two datasets, propensity score matching was conducted. The matching was conducted on the merged dataset to ensure comparability of characteristics of the sampled individuals for the DID analysis.

(ATLAS.ti.7.1) by the analyst. The interview transcripts were then imported to the data analysis software and the analysis team conducted thematic coding using the final version of the codebook.

Analysis: Thematic analysis was conducted using ATLAS.ti.7.1, a qualitative data analysis software. Themes and analysis were driven largely by the substantive EQs associated with the Transform program. In addition, the evaluation team used the Results Frameworks of the IPs to understand and assess whether the intervention theory of change held up in the implementation of activities, and whether contextual factors affected outcomes.

Text excerpts irrelevant to a given code were greyed out, while remaining excerpts were annotated with comments and illustrative quotations were highlighted. Codes and subcodes produced code frequencies that were disaggregated by region, activity, and KI type; they were then transformed into themes and subthemes.

3.4 Limitations and Mitigation

Three principal challenges arose that limited the ideal implementation of data collection protocols for the baseline-endline survey data that formed the foundation of the impact evaluation.

- 1) **Spillover:** Singling out the net effect of the Transform Program in terms of measurable health outcomes was complicated by the fact that interventions were assigned in an environment that was not closed, a risk inherent in any quasi-experimental design. An important consequence is that spillover likely occurred because the intervention was meant to support existing health systems and not just individuals living in Transform intervention areas. Moreover, innovations from one support program were often borrowed throughout the system, thus spilling over into non-intervention areas. The evaluation did not specifically measure the impact of spillover; however, this limitation was mitigated to some degree by the exclusion criterion that precluded the choice of contiguous Transform intervention and non-Transform intervention areas.
- 2) **Sampling and Rollout Challenges:** The design of the Transform program was complex, with multiple core activities. Thus, obtaining a representative sample of individuals exposed (and not exposed) to interventions in a uniform manner proved to be challenging. The evaluation team mitigated this challenge by collecting complementary qualitative data to address some of the complexities that affected the quantitative comparison of baseline-endline data. In addition, some woredas were exposed to interventions from other partners in more than two thematic areas, surpassing the exclusion cutoff and potentially introducing bias in the sample. Two woredas were excluded from data collection for this reason.
- 3) **Potential Survey Bias:** Large-scale household surveys are subject to numerous potential sources of bias. For example, bias related to household role can emerge if heads of household (often a male) maintain oversight of the interview. In addition, social desirability bias can

emerge if survey respondents provide answers in an effort to please the enumerator (consciously or otherwise). While the evaluators could not measure the degree of bias associated with these threats, they were aware that such forms of bias commonly affect household-level surveys. To mitigate the threat, Transform: MELA relied on an intense enumerator training process to mitigate these potential sources of bias. In addition, the evaluators did not see grounds to suspect that potential survey bias may have differed in systematic ways across Transform intervention and non-Transform intervention areas.

3.5 Ethical Considerations

Data collectors and supervisors were trained in research ethics including informed consent, the privacy of participants, and confidentiality. Prior to data collection for both the baseline and endline surveys, Institutional Review Board approval for human subjects research was obtained from the Ethiopian Public Health Institute (EPHI).

With respect to the individuals and households included in the data collection efforts, enumerators obtained permission from the heads of households and individual participants gave their consent before the interviews took place. For women under the age of 18, additional parental permission and participant assent was obtained before data collection. After the completion of data collection, data was stored with no personally identifiable information (PII).

4. Findings

This section provides evidence on the contribution of the Transform Program to improving the uptake of FP methods and MNCH service utilization. The results are organized by thematic area: FP, maternal health, child health and immunization, and cross-cutting themes. They include the DID results from the comparison of data from Transform intervention and non-Transform intervention areas, over the period from the baseline in 2017 to the endline in 2022. In addition, qualitative information from KIIs complements the DID results. Finally, the findings include results from multivariate analyses that identify the covariates that correlate with the indicators of interest.

For the DID analysis, three covariates, women's age, women's educational level, and women's marital status, were used in the propensity score matching so that the distribution of measured baseline covariates is similar between subjects in the Transform intervention and non-Transform intervention subjects at endline. Otherwise stated, the matching strategy ensures that close matches exist across the Transform intervention and non-Transform intervention areas, which improves the comparative analysis.

4.1 Family Planning Uptake

4.1.1 DID Results

Three key indicators were used to measure FP uptake: MCPR, the use of long-acting contraceptive methods, and the use of PFP methods.

Modern Contraceptive Prevalence Rate

As Table 2 below shows, the Transform program contributed positively to the use of modern contraceptive methods. In Transform intervention areas, the rate rose from 17.3% at baseline to 20.6% at the endline, representing an increase of 3.3 percentage points. Meanwhile, MCPR in non-Transform intervention areas decreased from 20.1% at the baseline stage to 17.3% at the endline. The net difference of 6.1 percentage points is statistically significant at the 95% confidence level.

Table 2. DID Results of Family Planning Uptake

Indicators	Baseline			Endline			DID (95% CI)
	Intervention (%)	Non-intervention (%)	Diff	Intervention (%)	Non-intervention (%)	Diff	
Family planning							
MCPR	17.3	20.1	-2.8*	20.6	17.3	3.3*	6.1 (3.4; 8.8)*
LA users	3.9	6.3	-2.3*	7.5	6.0	1.5*	3.8 (2.3; 5.5)*
Postpartum FP use	15.0	20.4	-5.4*	15.0	11.6	3.4*	8.8 (3.7; 13.9)*

* p<0.05; Diff – Difference in proportions; DID – Difference-in-Differences; CI – Confidence Interval

Long-acting FP Methods Use

The results also show a positive contribution from the Transform Program in terms of long-acting family planning (LAFP) use. In Transform intervention areas, the rate of uptake increased from 3.9% to 7.5% from baseline to endline. In non-Transform intervention areas, the use of long-acting contraception decreased slightly from 6.3% to 6.0% over the same time period. Overall, the net difference from baseline to endline in Transform Intervention areas versus non-Transform intervention areas was 3.8 percentage points, representing a statistically significant difference (Table 2).

Postpartum FP Use

Finally, the Transform Program also contributed positively to the use of PFP methods. In intervention areas, the use of PFP remained consistent from baseline to endline, at 15.0%. However, in the non-Transform intervention areas, the rate of PFP use declined from 20.4% to 11.6% from baseline to endline. The net result is thus an 8.8 percentage point advantage over time to the group of respondents exposed to the Transform intervention, a finding that is again statistically significant at the 95% confidence level (Table 2). To underscore this logic, in DID analyses, a static outcome or even a decline in intervention areas can represent a net positive effect if the decline in the non-intervention area is greater. The methodological design treats the non-

intervention results as what would have occurred in the absence of the intervention, so a relatively better result can be viewed as a positive effect from the program.

4.1.2 Complementary Qualitative Findings

Key informant interviews (KIIs) with health workers at the regional, zonal, and woreda levels indicated that FP utilization improved largely due to the availability of FP options in health posts. As a result, mothers were able to make informed decisions on FP methods using a rights-based approach, whereas previously they were pushed to choose LAFP methods over short-term methods without much thought given to their own needs and circumstances, resulting in weaker uptake. In addition, KIIs indicated that PFP use increased due to program activities including service integration and reconfiguration. Regarding FP as well as other MNCH activities, KIs stressed the value of performance-based outcomes in generating positive change. For example, one KI indicated that drawing performance data from twinned woredas allowed for a clear comparison of results and a reasonable resource distribution strategy.

Some KIs indicated that FP uptake still remains suboptimal in Transform intervention areas because of a shortage of some FP options. In addition, the program fell short of its ultimate FP targets, according to KIIs, due to persistent misconceptions and the politicization of FP issues. For example, a KI in Amhara indicated that FP options skewed toward implants due to a lack of continuous supply of other methods, and cultural issues were also reported to have limited FP use among some women.

Overall, however, KIs felt that the Transform Program had an important effect on FP uptake, noting that the program implemented activities at a very local level. This was viewed as a positive approach relative to other program interventions.

An important theme that arose in conversations with KIs as well as in the midterm data (not part of the impact evaluation) was that progress was overwhelmingly positive until the outbreak of conflict and worsening covid-related setbacks. Those contextual factors can explain why outcomes declined in both Transform intervention and non-Transform intervention areas, even as the Transform areas outperformed the comparison areas. The overall declines mask progress that was then undermined by unforeseeable circumstances.

4.1.3 Multivariate Analyses: Determinants of FP Uptake

In multivariate analyses, several factors correlated significantly with FP uptake.

Modern Contraceptive Prevalence Rate

Regarding MCPR, odds ratios were calculated using the following reference categories: women aged 40+, secondary education, 7+ children, and having not received RH/FP messages.

First, age cohorts mattered: women aged 35-39 were 79% more likely to use modern contraception compared to the reference category of women aged 40+.¹² Meanwhile, women aged 15-19 and 20-24 were just 37% and 55% as likely, respectively, to use modern contraception compared to those aged 40+. Education also contributed to MCPR. Those with no education were only 20% as likely to use modern contraception compared to the reference category of secondary school educated women, while those with a primary education were just 59% as likely. Furthermore, having previously given birth (children ever born) correlated with MCPR: women who had given birth 1-2 times were 202% more likely to use modern contraception than were women in the reference category of having given birth 7+ times. Similarly, women who had given birth 3-4 times and 5-6 times were 113% and 58% more likely, respectively, than those who had given birth 7+ times to use modern contraception. In addition, having received the RH/FP message made a difference in women's uptake of FP: those who received the message were 54% more likely to report using modern contraception compared to those who did not receive the message. Finally, women who participated in their own health care decisions were 22% more likely to use modern contraception (See Annex 2, which displays the odds ratios for those significant predictors).

Long-acting Family Planning Use

Several factors also correlated with the use of LAFP methods. Using the same reference categories, age again correlated with using LAFP methods: women aged 15-19 were only 44% as likely to use LAFP methods as were women aged 40+, whereas those aged 35-39 were 69% more likely than those aged 40+ to use LAFP methods. Regarding education, those with no education and primary level education were only 29% and 74% as likely, respectively, to use LAFP methods compared to the reference category of secondary educated women. Women with fewer children were more likely to use LAFP methods: those with 1-2 children and those with 3-4 children were 2.34 times and 94% more likely to do so than those with 7+ children, respectively. Women who received the RH/FP message and those who participated in decisions about their own health care were also 41% and 30% more likely to use LAFP methods, respectively (Annex 2).

Postpartum Family Planning Use

Regarding the use of PFPF methods, numerous factors again correlated with usage. Women in the age category of 15-19 and 20-24 were 18% and 32% as likely to use PFPF methods as were women aged 40+. Education was also an important determinant, as those with no education were only 14% as likely to adopt PFPF methods and those with primary education were only 42% as likely to do so, relative to the reference category of secondary educated women. Receiving the RH/FP message was indeed critical, as those who received the message were 56% as likely to use PFPF methods versus those who did not. Neither the number of children born to a woman nor women's participation in decisions about their own health care had an impact on their decisions regarding PFPF use (Annex 2).

¹² An odds ratio of 1 indicates equal likelihood. Thus, an outcome of 1.79 is interpreted as 79% more likely. An odds ratio of 2 would be 100% more likely. An odds ratio of 3 would be 200% more likely.

4.2 Maternal Health

4.2.1 DID Results

The evaluation team used seven indicators to evaluate Transform Program contributions to maternal health: 1) early ANC visits, during the first three months of pregnancy; 2) making at least one ANC visit; 3) receiving essential components of ANC; 4) receiving iron and folic acid supplements (IFA) for at least three months during pregnancy; 5) making four or more ANC visits; 6) having the birth attended by a SBA; and 7) receiving early PNC, in the two days after birth.

Early ANC

As Table 3 shows, the share of women doing early ANC visits increased significantly among women in the Transform intervention areas from baseline to endline, relative to the non-Transform intervention areas. In Transform intervention areas, the rate rose from 26.9% at baseline to 36.2% at the endline, representing an increase of 9.3 percentage points. Meanwhile, early ANC in non-Transform intervention areas decreased from 33.0% at the baseline stage to 31.3% at the endline. The net difference of 11.0 percentage points, shown in Table 3, is statistically significant at the 95% confidence level.

Table 3. DID Results of Maternal Health Services Utilization

Indicators	Baseline			Endline			DID (95% CI)
	Intervention (%)	Non-intervention (%)	Diff	Intervention (%)	Non-intervention (%)	Diff	
Maternal health							
Early ANC	26.9	33.0	-6.1*	36.2	31.3	4.9*	11.0 (4.3; 17.7)*
At least one ANC	69.2	64.9	4.2	66.6	58.2	8.4*	4.2 (-2.9; 11.3)
Essential ANC	36.3	48.7	-12.4*	21.2	17.0	4.2	16.6 (9.9; 23.3)*
IFA 3-month	8.7	11.7	-3.0*	12.1	8.8	3.3*	6.3 (2.0; 10.6)*
ANC 4+	40.4	44.9	-4.5*	26.8	20.1	6.7*	11.2 (4.7; 17.7)*
SBA	45.4	52.7	-7.3*	39.9	34.5	2.5*	12.6 (5.5; 19.7)*
Early PNC	35.5	38.4	-2.9	30.1	28.0	2.1	5.0 (-1.7; 11.7)

* p<0.05; Diff – Difference in proportions; DID – Difference-in-Differences; CI – Confidence Interval

At Least One ANC

The DID analysis did not return a statistically significant difference in women making at least one ANC visit. The share of women in Transform intervention areas making at least one ANC visit declined from 69.2% to 66.6% from baseline to endline, while in non-Transform intervention areas the share dropped from 64.9% to 58.2%. While the net difference of 4.2 percentage points is not statistically significant, the trend nevertheless moved in a positive direction (Table 3).

Essential Components of ANC

The Transform Program also contributed positively to the adoption of essential components of ANC. In intervention areas, essential components of ANC dropped from 36.3% to 21.2%, a rather steep decline. Yet, the rate of essential components of ANC dropped even more precipitously in non-Transform intervention areas, from 48.7% to 17.0%, suggesting a steeper drop. The net result is thus a difference of 16.6 percentage points among the group of respondents exposed to the Transform intervention, a finding that is again statistically significant at the 95% confidence level (Table 3).

Iron and Folic Acid Supplementation (for at least 3 Months)

The Transform Program contributed positively to IFS. In Transform intervention areas, the rate increased from 8.7% to 12.1%, representing a 3.4 percentage point difference. Meanwhile, in non-Transform intervention areas, the rate dropped from 11.7% to 8.8%, a 2.9 percentage point decline. The net result was a difference of 6.3 percentage points among the group of respondents exposed to the Transform intervention, a finding that is again statistically significant at the 95% confidence level (Table 3).

ANC 4+

The results also show a significant, positive contribution from the Transform Program in terms of women attending at least four ANC visits. In Transform intervention areas, the rate of ANC 4+ actually declined from 40.4% at baseline to 26.8% at endline. However, the share of women attending at least four ANC visits dropped even more steeply in non-intervention areas, from 44.9 % to 20.1% over the same period. Overall, the net difference from baseline to endline in Transform intervention areas versus non-Transform intervention areas was 11.2 percentage points, representing a statistically significant difference (Table 3).

Skilled Birth Attendance

The Transform program contributed positively to using SBA. Among women in the Transform intervention areas, the use of SBA declined from 45.4% to 39.9%. Among women in the non-Transform intervention areas, however, the use of SBA dropped even more, from 52.7% to 34.5%. The net overall difference from baseline to endline in the Transform intervention areas compared to the non-Transform intervention areas was a positive 12.6 percentage points, which is statistically significant at the 95% confidence level (Table 3).

Early PNC

Finally, the use of early PNC, within two days after birth, did not differ significantly in Transform intervention versus non-Transform intervention areas. In the Transform intervention areas, the early PNC rate declined from 35.5% to 30.1%, although the drop was even steeper in non-Transform intervention areas (38.4% to 28.0%). While the overall, relative change was a positive 5.0 percentage points from baseline to endline, the results did not attain conventional levels of statistical significance.

4.2.2 Complementary Qualitative Findings

The DID results indicated that most indicators declined in the Transform intervention areas but to a lesser extent than they declined in the non-Transform intervention areas. These outcomes represent positive program contributions, even as they might suggest some frustration among health officials in Transform areas. Indeed, some KIs noted that performance can improve in terms of maternal health service quality.

Nevertheless, the KII results revealed that health actors saw real improvements in the Transform intervention areas. One KI in Amhara stressed that the trainings including the BEmOC made important contributions. Similarly, KIs reported that home deliveries declined in the Transform areas, while women showed more interest in ANC attendance, including early ANC. In fact, some KIs noted that maternal deaths at health facilities had declined (according to MDSR data) due to improved service quality and use. They indicated that the program's capacity building and quality improvement activities were responsible. Among the effective interventions that improved maternal health service access and quality, maternal waiting homes, compassionate care training, and BEmOC trainings were cited as explanations for relative improvements in delivery care quality. KIs also mentioned community level awareness activities and pregnant women conferences as catalysts for improved maternal health behaviors and outcomes.

KIs also provided input regarding the mobile delivery of services. Particularly in the heavily pastoralist communities in the Afar and Somali regions, mobile delivery of maternal health and other MNCH services are critical. KIs reported that mobile services were variable and uncertain, and that service providers did not have consistent pay or expectations across partner interventions.

4.2.3 Multivariate Analyses: Determinants of Maternal Health Behaviors

The evaluation team conducted multivariate analyses to examine the determinants of three key Maternal Health indicators: Early ANC, ANC 4+, and the use of SBA.

Early Antenatal Care

Regarding Early ANC, the results indicate that age has an effect among the younger mothers: those aged 15-19 were 42% as likely as those aged 40+ to take part in early ANC visit. The other age categories did not differ significantly from the reference category. Education also mattered importantly: those with no education were just 18% as likely to do early ANC visits, and those with a primary education were just 46% as likely to do so, relative to the reference category of secondary educated women. The fewer children a mother had, the more likely she was to take part in early ANC visits: those with 1-2 children were 122% more likely than those with 7+ children to do early ANC; those with 3-4 children were 78% more likely to do so, and those with 5-6 children were 58% more likely to do so. Having received MNCH message made women 39% more likely to take part in early ANC visits. Importantly, having one's spouse accompany her to ANC visits or delivery made women 284% more likely to take part in early ANC visits (Annex 3).

Four or More Antenatal Care Visits

Similar factors predicted the likelihood of engaging in four or more ANC visits. Younger women were less likely to engage in four or more ANC visits, with those aged 15-19 17% as likely, those aged 20-24 25% as likely, and those aged 25-29 51% as likely to do so compared to women aged 40+. Education was also a significant predictor of ANC 4+, as those with no education were only 23% as likely as those with a secondary education to do so, and those with a primary education were only 66% as likely to do so. Women with a smaller number of children were also more likely to have four or more ANC visits: women with 1-2 children were 89% as likely to engage in at least four ANC visits compared to the reference category of women with 7+ births. Women who received the MNCH message were 59% more likely to take part in at least four ANC visits. Having a spouse accompany the woman to ANC visits or delivery was also a significant predictor: those women were 262% more likely to do at least four ANC visits (Annex 3).

Skilled Birth Attendance

Regarding the use of SBA, women in the 15-19 year age category were 28% as likely as those in the 40+ category to use SBA; other age categories were not statistically significant. In addition, those with no education were only 31% as likely to use SBA, and those with a primary education were 59% as likely to do so, relative to those with a secondary education. Women with fewer children were much more likely to use SBAs: those with 1-2 children were 407% more likely to do so, those with 3-4 children were 150% more likely to do so, and those with 5-6 children were 93% more likely to do so, relative to those with 7+ children. Receiving the MNCH message also made a difference: women who did so were 54% more likely to use SBA. Finally, women whose spouse attended their ANC visits or delivery were a remarkable 3,196% more likely to use SBA (Annex 3).

4.3 Child Health and Immunizations

4.3.1 DID Results

Nine indicators were used to gauge the Transform program contributions to child health and immunizations. Those indicators are: 1) early PNC for newborns; 2) essential newborn care (ENC) at health facilities; 3) full vaccination; 4) vitamin A supplementation; 5) treatment of fevers; 6) ARI treatment; 7) diarrhea treatment; 8) early breastfeeding; and 9) exclusive breastfeeding during the first six months.

Early PNC for Newborns

The net difference in early PNC was positive, but the DID result was not statistically significant. In the Transform intervention areas, early PNC declined slightly from baseline to endline, going from 29.8% to 28.8%. In the non-Transform intervention areas, the early PNC rate declined from 31.9% to 25.5%, resulting in a net DID of 5.5 percentage points. Because the result would have been a decline of 6.4 percentage points but instead was only a decline of 1.0 percentage points, this is indicative of a positive Transform program contribution. However, the difference did not reach the conventional levels of statistical significance.

Table 4. DID Results of Child Health Services Utilization

Indicators	Baseline			Endline			DID (95% CI)
	Intervention (%)	Non-intervention (%)	Diff	Intervention (%)	Non-intervention (%)	Diff	
Child Health, Immunization, and Nutrition							
Early PNC newborn	29.8	31.9	-2.2	28.8	25.5	3.3	5.5 (-1.0; 12.0)
Essential newborn care (ENC) at facility	7.0	4.4	2.6	17.9	10.4	7.4*	4.9 (-1.4; 11.2)
Full vaccination	26.9	30.6	-3.7	18.7	15.1	3.6	7.3 (1.0; 13.6)*
Vitamin A supplementation	39.8	33.5	6.3*	55.0	40.7	14.3*	8.0 (3.1; 12.9)*
Fever treatment	44.1	53.1	-9.0*	22.9	19.9	3.0	12.0 (2.8; 21.2)*
ARI treatment	23.9	18.6	5.3	43.2	48.3	-5.0	-10.3 (-26.1; 5.3)
Diarrhea treatment	28.9	28.9	0.0	30.4	30.6	-0.2	-0.1 (-9.9; 9.7)
Early BF	63.5	56.3	7.3*	74.5	68.3	6.2*	-1.1 (-9.1; 6.9)
Exclusive BF	62.5	74.0	-11.5*	67.7	64.0	3.7	15.2 (4.0; 26.4)*

* p<0.05; Diff – Difference in proportions; DID – Difference-in-Differences; CI – Confidence Interval

Essential Newborn Care (ENC)

ENC increased over time in both the Transform intervention and non-Transform intervention areas. In Transform intervention areas, the ENC rate rose from 7.0% to 17.9%, whereas in the non-Transform intervention areas the ENC increased from 4.4% to 10.4%. Overall, the net difference was a 4.9 percentage points advantage in the Transform intervention areas, but again, the difference was not statistically significant at the 95% confidence level (Table 4).

Full Vaccination

In the Transform intervention areas, the rate of full vaccination declined from 26.9% at the baseline to 18.7% at the endline. However, the decline over time in non-Transform intervention areas was notably steeper: the full vaccination rate dropped in those areas from 30.6% to 15.1%. Thus, the overall relative difference in full vaccination rates was 7.3 percentage points in favor of those in the Transform intervention areas, a difference that is statistically significant at the 95% confidence level (Table 4).

Vitamin A Supplementation

The intake of vitamin A supplements also increased among children both in the Transform intervention and non-Transform intervention areas. Among the children in the Transform intervention areas, vitamin A intake increased from 39.8% at the baseline to 55.0% at the endline.

Among the non-Transform intervention areas, intake of vitamin A rose from 33.5% to 40.7%. The larger increase in the Transform intervention areas resulted in a net 8.0 percentage point difference due to the contribution of the Transform program, a difference which is statistically significant at the 95% confidence level (Table 4).

Fever Treatment

Fever treatments worked in the opposite direction, declining in both intervention and non-intervention areas. Among the Transform intervention areas, treatment of fevers among children decreased from 44.1% to 22.9% over time. Among children in the non-Transform intervention areas, the decline was sharper, from 53.1% to 19.9%, resulting in a net DID of 12.0 percentage points. Thus, despite the decline over time in the Transform intervention areas, the Transform program can be viewed as contributing positively to fever treatment, given the sharper declines elsewhere. The difference is statistically significant (Table 4).

ARI treatment

ARI treatment represented one of the child health indicators for which the relative difference over time was negative. In the Transform intervention areas, the ARI treatment rate rose from 23.9% to 43.2%, an increase of 19.3 percentage points. Yet, in the non-Transform intervention areas, ARI treatment rose from 18.6% to 48.3%. The net DID was thus -10.3 percentage points, although the difference was not statistically significant at conventional significance levels (Table 4).

Diarrhea Treatment

Change in diarrhea treatment over time was almost identical in Transform intervention and non-Transform intervention areas: in the Transform intervention areas, treatment increased from 28.9% to 30.4%, whereas in the non-Transform intervention areas, diarrhea treatment rose from 28.9% to 30.6%. The net DID was thus -0.2 percentage points, although the difference is substantively trivial and statistically insignificant. The Transform program thus had no discernible effect on diarrhea treatment (Table 4).

Early Breastfeeding

The rate of early breastfeeding increased similarly in Transform intervention versus non-Transform intervention areas. In the former, early breastfeeding rose from 63.5% to 74.5%, a difference of 11 percentage points. However, the rate also increased in non-Transform intervention areas, rising from 56.3% to 68.3%. The DID was thus -1.1 percentage points, a difference that again was substantively very small and not statistically significant (Table 4).

Exclusive Breastfeeding

Finally, the Transform program contributed to a significant difference in the rate of exclusive breastfeeding of children younger than 6 months old. In the Transform intervention areas, the rate of exclusive breastfeeding increased from 62.5% to 67.7% from the baseline to the endline. Meanwhile, in the non-Transform intervention areas, the rate of exclusive breastfeeding dropped from 74.0% to 64.0%, a 10-percentage point decline. The DID of 15.2 percentage points is statistically significant and substantively quite notable (Table 4).

4.3.2 Complementary Qualitative Findings

The DID results for child health and immunizations were somewhat mixed: of the 9 indicators, 4 showed positive, statistically significant contributions among the Transform intervention areas; 2 showed positive but insignificant differences; and 3 showed negative but insignificant differences. Feedback collected during KIIs reinforced the ambiguous nature of the program effects on child health and immunizations.

One KI at the RHB attributed program shortcomings to water and sanitation issues. That person said, “child health would have been much better if the program had included intervention on water availability in health facilities as water is a critical problem in health facilities in the target region.” Ensuring quality provision of services also proved to be challenging in some instances; according to one KI in Amhara, “In terms of access to Newborn care, we have reached roughly 38 woredas, but quality remains a problem, and neonatal health indicators are inadequately improved [as a result].”

On the other hand, some KIs did note important contributions from the Transform program. For example, numerous KIs highlighted the vaccination training, material support, and outreach support as program components that made a positive difference in their experience. Mentoring was also viewed as an effective component of the program. Speaking of the Transform: HDR intervention activities, one KI in Gambella said, “Transform HDR was effective and improved health care services from the grass root level. The training, coaching, and mentoring...contributed to quality service provision. It was particularly helpful for prenatal care and child health follow-up.”

4.3.3 Multivariate Analysis: Determinants of Child Health and Immunizations

The evaluation team conducted multivariate analyses to examine the determinants of three key indicators of child health and immunization: full vaccination, fever treatment within 24 hours of onset, and deworming. The covariates included the mother’s age, education, number of child births, and whether they received the MNCH message within the last few months prior to the survey.

Full Vaccination

Regarding full vaccination, the mother’s age was not a systematically important determinant of children getting fully vaccinated, although those aged 35-39 were 152% more likely to fully vaccinate their children than were mothers aged 40+. Education, unlike age, had a notable impact: mothers with no education were only 19% as likely as those with a secondary education to fully vaccinate their children; those with a primary education were 55% as likely to do so. Neither the number of child births nor having received the MNCH message had a significant effect (see the statistically significant effects in Annex 4).

Fever Treatment

With respect to fever treatments for children, neither the mother's age nor education had a significant effect. Those who had 1-2 children were 95% more likely than the reference category of those with 7+ births to obtain fever treatments for their children. Finally, receiving the MNCH message made women 77% more likely to obtain fever treatments for their children (statistically significant factors are illustrated in Annex 4).

Deworming

Multiple factors were predictive of the likelihood of obtaining deworming medications for children. In terms of age, mothers aged 15-19 and 20-24 were 58% and 73% as likely, respectively, to obtain deworming medication for their children, relative to women aged 40+. Meanwhile, those aged 35-39 were 47% more likely than those aged 40+ to do so. Education also had a significant effect on deworming: those with no education were 29% as likely to deworm their children as those with secondary education, although those with primary school education did not differ significantly from those with secondary school education.

Having fewer children also correlated with a greater likelihood of obtaining deworming for one's children: those with 1-2 children were 50% more likely to do so, relative to those with 7+ children; those with 3-4 children were 41% more likely to do so, and those with 5-6 children were 29% more likely to do so. Finally, receiving the MNCH message made mothers 41% more likely to have their children dewormed (see Annex 4, which shows the significant predictors of deworming).

4.4 Cross-cutting Themes

The cross-cutting themes included three sanitation related indicators: access to improved sanitation facilities, having a hand washing station with water and soap/ash at home, and using appropriate water treatment techniques. In addition, the cross-cutting themes included three gauges of gender-related indicators: the woman participating in her own health care decisions; having a spouse accompany the mother to ANC visits; and having the spouse accompany the mother to delivery. Finally, the cross-cutting themes included an indicator for whether the family enrolled in the community-based health insurance program (CBHI).

4.4.1 DID results

Access to Improved Sanitation

Access to improved sanitation increased modestly in Transform intervention areas, from 9.9% at the baseline stage to 10.3% at the endline. Conversely, it declined in the non-Transform intervention areas, from 14.7% at the baseline to 12.5% at the endline. The resultant difference of 2.5 percentage points is statistically significant at the 95% confidence level (Table 5).

Table 5. DID Results of Cross-cutting Themes

Indicators	Baseline			Endline			DID (95% CI)
	Intervention	Non-intervention (%)	Diff	Intervention (%)	Non-intervention (%)	Diff	
Cross-cutting							
Access to improved sanitation facility	9.9	14.7	-4.7*	10.3	12.5	-2.2*	2.5 (0.3; 4.7)*
Hand washing station with water and soap/ash	1.1	2	-0.8*	1.5	1.9	-0.4	0.5 (-0.3; 1.3)
Appropriate water treatment	11.8	10.5	1.4	7.6	8.8	-1.2	-2.6 (-4.6; -0.6)*
Women own health care decision	82.9	82.3	0.6	76.6	73.5	3.1*	2.5 (-0.6; 5.6)
Spouse Accompany – ANC Visits	53.4	55.5	-2.2	30.1	28	2.1	4.3 (-2.8; 11.4)
Spouse Accompany – Delivery	72.3	76	-3.7	82.7	79.8	2.9	6.5 (-1.3; 14.3)
CBHI	16.7	16.2	0.5	73.8	73.9	-0.1	-0.6 (-3.9; 2.7)

* p<0.05; Diff – Difference in proportions; DID – Difference-in-Differences; CI – Confidence Interval

Hand Washing Station with Water and Soap/Ash

Contributions of the Transform program to the presence of household hand washing stations with water and soap/ash were negligible. In the Transform intervention areas, the proportion of households with hand washing stations with water and soap/ash increased very modestly from 1.1% to 1.5% during the period of performance. In the non-Transform intervention areas, the rate remained largely static, decreasing slightly from 2.0% to 1.9%. Overall, the DID of 0.5 percentage points was positive but not substantively meaningful or statistically significant (Table 5).

Use of Appropriate Water Treatment Techniques

Households' use of appropriate water treatment techniques declined in both the Transform intervention and non-Transform intervention areas. From the baseline to endline period, use of appropriate water treatment techniques decreased from 11.8% to 7.6% among the Transform intervention areas, whereas among the non-Transform intervention areas, the rate decreased more modestly, from 10.5% to 8.8%. The net DID was thus -2.6 percentage points, which represents a statistically significant difference in the negative direction (Table 5).

Women Participation in Decisions Regarding their Own Health

In the Transform intervention areas, 82.9% of women stated that they participated in their own health care decisions at baseline. By the endline, that figure had decreased to 76.6%. Similarly, the share of women who stated that they participated in their own health care decisions in the non-

Transform intervention areas declined from 82.3% at the baseline to 73.5% at the endline. Overall, the Transform program contributed to a modest difference of 2.5 percentage points, but that difference is not statistically significant (Table 5).

Women Accompanied by their Spouses during ANC Visits

The rate at which spouses accompanied women to ANC visits declined consistently across Transform intervention and non-Transform intervention areas. In the former, the rate at which spouses accompanied women to ANC visits dropped from 53.4% to 30.1% during the period from baseline to endline, whereas in the latter, the rate declined from 55.5% to 28.0%. There was thus a net difference of 4.3 percentage points in the Transform intervention areas compared with the non-Transform intervention areas, although the difference is not statistically significant (Table 5).

Women Accompanied by their Spouses during Delivery

Conversely, the rates at which spouses accompanied women to the delivery of their children increased across the board both in Transform intervention and non-Transform intervention areas. In the Transform intervention areas, 72.3% of women were accompanied by a spouse to the delivery at the baseline, and that rate rose to 82.7% by the endline. In the non-Transform intervention areas, 76.0% of women were accompanied to the delivery, a rate which increased to 79.8% by the endline. The resultant DID was a net difference of 6.5 percentage points in the Transform intervention areas, although the difference did not attain conventional levels of statistical significance (Table 5).

CBHI

Finally, enrollment in CBHI sharply increased over the lifetime of the Transform program, but the increase cannot be attributed to Transform interventions. In the Transform intervention areas, the uptake increased from 16.7% to 73.8%. However, even in the non-intervention areas, CBHI uptake increased from 16.2% to 73.9%. Overall, the DID results yielded a -0.6-percentage point difference, which is neither substantively nor statistically meaningful. In short, the Transform program had no discernable effect on CBHI enrollment (Table 5).

4.4.2 Complementary Qualitative Findings

One important theme that emerged from the qualitative evidence and that speaks to the cross-cutting theme of women's decision-making power is that the availability of services can invigorate the autonomy of women. KIs noted that in places where options are made available to women, they tend to recognize and embrace the opportunity to gain information and make informed decisions. This can encourage them to be more involved in their own health care decisions and to bring information to their spouses. In developing regions, particularly Somali and Afar, mobile health systems and outreach also emerged as an important innovation, and mobile health has been recently codified through MOH policy.

Others noted that improvement in the cross-cutting areas of sanitation and gender dynamics can be complicated by patterns of dependency and cultural legacies. Many FP/MCH behaviors can be adopted without extensive cultural change, but gender dynamics in particular requires a shift in thinking and social standing for both men and women, according to a KI from the HDR Activity.

The evaluation team did not conduct multivariate analysis for the cross-cutting indicators as the variables were primarily covariates rather than regressands.

5. Conclusions

The impact evaluation focused on two key Evaluation Questions: to what extent did the interventions improve MNCH/FP outcomes, and were there any unintended consequences from the interventions?

The Difference-In-Differences analyses revealed unequivocally that the Transform interventions contributed positively to maternal and child health in Ethiopia, particularly in the thematic areas of FP and maternal health. Most of the indicator values declined both in the intervention and non-intervention areas; the positive results are due to relatively larger declines in non-Transform intervention areas compared to the Transform intervention areas. Because the results in Transform intervention areas were better than what would have been observed under the counterfactual condition of no Transform intervention, the DID findings are interpreted as a positive contribution from Transform, offsetting deleterious contextual factors to some degree.

Additional data (not part of the impact evaluation) indicated that the declines occurred after the midpoint of the Transform Program, which coincided with the outbreak of conflict and worsening covid-related consequences.

In the area of FP, all three indicators showed significant positive differences. In the area of maternal health, five of the seven indicators showed significant positive differences, and the remaining two were positive though not statistically significant. The results were somewhat mixed for child health and immunizations: four outcomes were positive and significant, two were positive but not significant, and three were negative though not significant. Only the interventions in the cross-cutting themes of sanitation, women's empowerment, and spousal relations did not bear fruit: one of the seven indicators (improved sanitation facilities) showed positive, statistically significant differences, while one showed significant difference in the negative direction (appropriate water treatment) and four yielded ambiguous outcomes. The positive effects in FP and maternal health were typically on the order of 5 to 15 percentage points, representing important contributions to public health and well-being in Ethiopia.

In addition to the effects of the interventions as measured using DID analysis, the multivariate analyses also shed light on key sociodemographic determinants of MNCH/FP outcomes. Across nearly all indicators, three factors stood out as consistently relevant: education levels, the number of children born to the woman, and receiving the MNCH/FP program messages. Women with more education showed significantly greater uptake of positive MNCH/FP behaviors, as did those with fewer children.

There were no obvious unintended consequences from program interventions. However, a perhaps underappreciated consequence arose from program efforts to encourage spouses to accompany their wives to ANC visits and deliveries. As the report noted above, the rates of spouses

accompanying their wives did not improve significantly in Transform intervention areas compared to non-Transform intervention areas. However, where spouses did accompany their wives, important secondary outcomes emerged; in particular, the use of SBA skyrocketed, and women became much more likely to engage in several ANC visits. These results, coupled with the positive effects of education, suggest that information, awareness, and learning present critical opportunities for continued progress on MNCH/FP issues in Ethiopia, and that both women and men have important roles to play.

6. Recommendations

A set of recommendations follow from the key findings and conclusions of the impact evaluation. These recommendations may be considered by MOH and USAID/Ethiopia to improve future MNCH/FP programming in Ethiopia.

Recommendations for MOH

1. **Revitalize community health programs.** Priority should be given to rehabilitating and revitalizing community health programs, especially in war-affected woredas. Health posts should be functional and provide services on all weekdays, and various community structures and platforms should be utilized to mobilize the community and ensure whole-of-society engagement. Key informants reported that community mobilization and engagement platforms have weakened, and this has resulted in declines in community-based performance. Community health programs established in the past, such as the women's development army, should be evaluated to identify where the gaps in the implementation of such programs lie.
2. **Institute mobile services policy and guidelines for developing regions.** The lack of standardized federal mobile services policies and guidelines has led to variation across partners which poses a challenge to institutionalizing mobile health approaches. Some partners provide certain services and not others; some provide incentives, and some do not; and per diems vary in ways that creates disincentives for health workers. Mobile health program standardization, particularly in Afar and Somali regions, is required to ensure the institutionalization and sustainability of access to key MNCH services to hard-to-reach pastoralist populations.
3. **Strengthen and standardize mobile health service delivery to expand access to MNCH/FP services for pastoralist communities:** Future Ethiopian federal and donor budget support to pastoralist regions should prioritize mobile health approaches to reach underserved communities. This is particularly important for services that require repeated visits (e.g., immunizations and antenatal care) and are designed to reduce discontinuation rates.
4. **Develop and implement health performance-based standards and budgeting for woredas.** Performance-based budgeting will better allow woredas to have resources to achieve their performance targets. The use of comparative data and budgeting from paired woredas can be tied explicitly to performance.

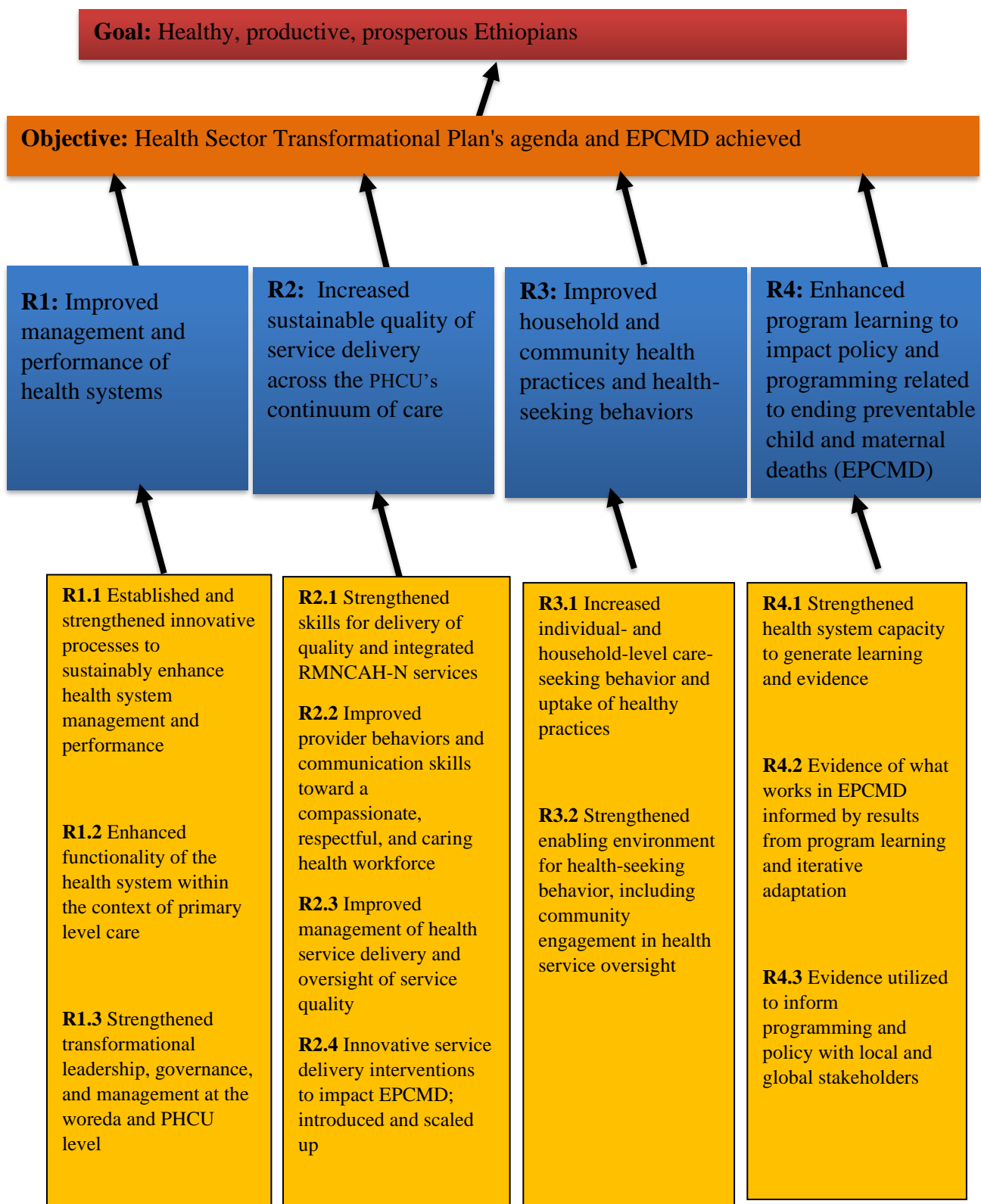
5. **Particularly in developing regions, assign more female providers for childbirth services.** The lack of female providers was reported to be a barrier to improving MNCH intervention coverage particularly in developing regions. Regional health bureaus should give priority to female applicants for enrollment and encourage donors to provide scholarships in midwifery training programs.
6. **Integrate FP counseling across all contacts in the continuum of care.** Contraceptive use is low in developing regions. Counseling for postpartum family planning was also rare and has not been integrated into programming. There is a need for standard integrated FP guidelines and service delivery checklists to ensure that every woman and man that comes to a facility gets counseled. Therefore, integrating family planning counseling during all contacts in the continuum of care (ANC, childbirth, postnatal care, immunization) should be demonstrated and achieved in the follow-on project.
7. **Ensure availability of family planning services at health posts.** Family planning is one of the packages of the health extension program in Ethiopia. MOH/RHB-supported activities programs should ensure full availability of family planning services in all health posts and work with partners to ensure sustainable “last mile” distribution of FP commodities.
8. **Ensure sustainability of in-service training activities.** In-service training was one of the vital investments of Transform PHC. Medicine and health sciences are dynamic and continuous medical education are always vital. However, the current in-service training lacks sustainability. To ensure sustainability, the in-service training should be linked to licensure and re-licensure programs, and future funding should consider other approaches, such as on-site training and technology-assisted approaches to minimize costs and ensure sustainability.
9. **Recruit and retain additional health facility personnel to ensure consistent provision of services, especially in difficult-to-reach areas.** Staff turnover can create gaps in service, and the evaluation finding suggest that distance is a factor in affecting activity performance in difficult to reach areas. MOH should therefore develop personnel retention plans that motivate qualified health staff to remain in their posts.
10. **Strengthen the support to narrow the gap between effective coverage and contact coverage.** For some MNCH/FP interventions, clients were unable to receive the recommended services despite their visit to health care facilities. For example, even though 26.8% of pregnant women in Transform Intervention areas had four or more ANC contacts at endline, only 12.1% received the full course of IFA, an implication that the quality of ANC services was below the recommended standard. Future programs should ensure that facilities have all the tools (skills of providers, motivation of providers to practice good quality services, availability of equipment and supplies) to provide the recommended services.

Recommendations for USAID

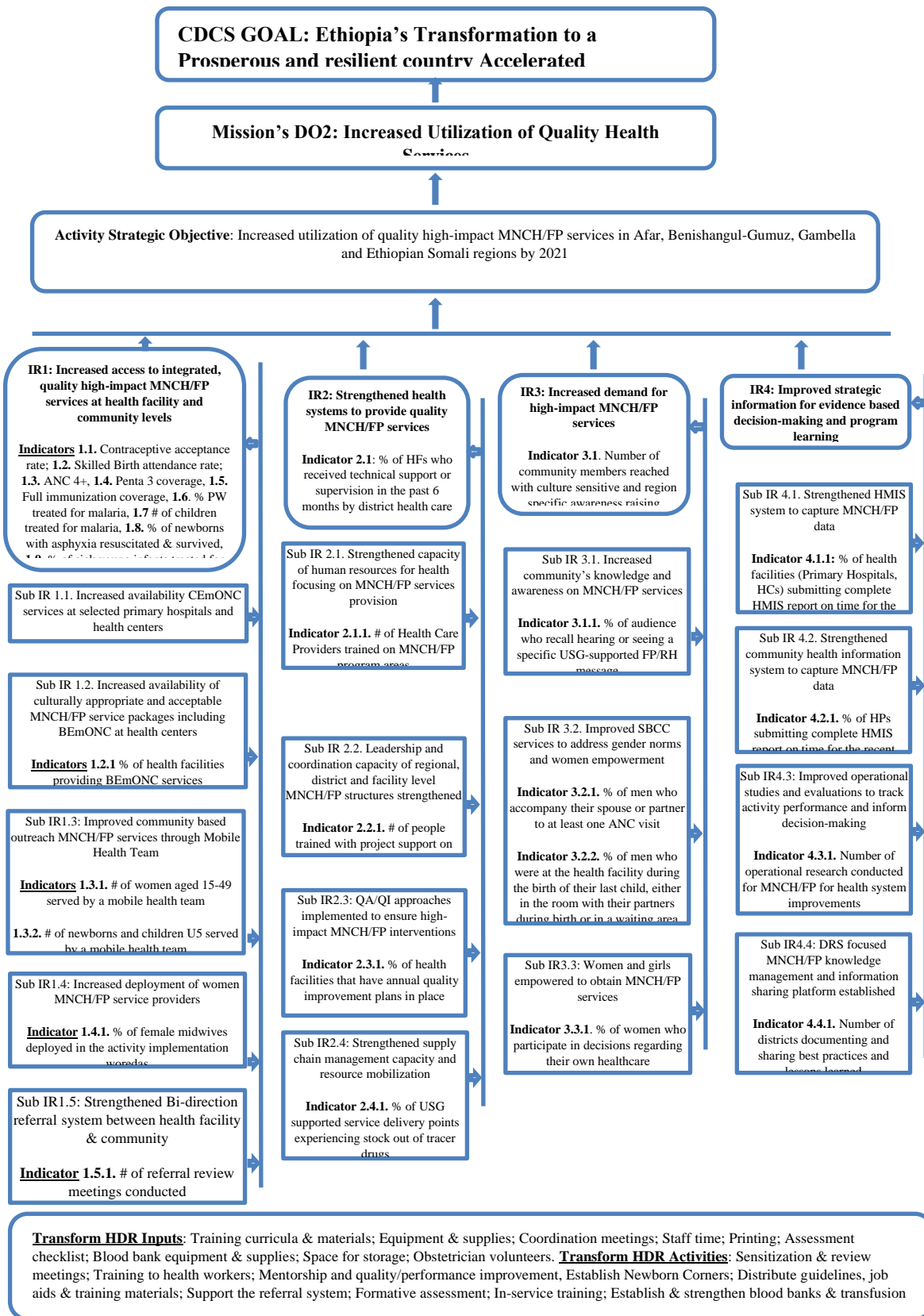
1. **Design future activities to be performance-based and continue to include flexible target setting, monitoring methodology, and assumptions in order to adapt to changing circumstances.** Targets against performance should be monitored continuously by a third party for course correction and adjustment of targets. Many Transform life of Activity targets were not fully achieved, but evidence from performance evaluations indicated that progress was positive on many metrics until the midterm, which coincided with the outbreak of conflict and increasingly worse Covid-related setbacks. Flexible targetting could account for such contextual factors and allow for timely adjustments. This would have improved the interpretation of target achievement.
2. **Develop an operational research agenda to identify causes of service bottlenecks and explore the underlying factors contributing to the decline in performance of selected KPIs.** Determining the root causes for the decline in performance for selected KPI requires further research. Key areas of interest are 1) reasons that women are unable to receive all services through the continuum of care and why they fail to return back after their first ANC contact; 2) approaches and strategies that contribute to improving shortages of supply within community-based health services, particularly routine immunization services; 3) the contribution and reimbursement rate of CBHI on health service uptake and quality; and 4) the most effective strategies to ensure health system resilience in conflict settings. The outcome of such research can inform future design and programming and contribute to the design of more effective and efficient operating procedures to improve operational productivity and coverage of supported health facilities and woredas. In addition, future research should include meta-analyses on the scope of setbacks due to conflict, disease, and other unforeseeable disruptions, so that such setbacks can be accounted for.
3. **Establish institutionalized mechanisms for field-level sharing of information between primary health interventions and interventions targeting pastoralist regions and communities.** USAID cooperative agreements with successor primary health interventions in agrarian regions should include specific clauses for formal mentoring and capacity development for developing region mechanisms. As one example, USAID primary health care implementing partners possess a wealth of experience in operations and implementation research, SBCC, and technical support for training and quality improvement. Future programming should formally leverage this capacity to ensure that lessons learned over the past decades in agrarian regions can be successfully adapted to developing regions.

Annexes

Annex IA. Transform PHC Activity Result Framework



Annex IB. Transform HDR Activity Result Framework



Annex 2. Adjusted Odds Ratios: Family Planning

95% CI from the logistic regression model for the determinants of family planning use among currently married women: Transform Program, Ethiopia, Endline survey, 2022.

Indicators	MCPR (n=5659; event=1240)		LAFP use (n=5659; event=443)		PPFP use (n=1611; event=230)	
	OR	95% CI	OR	95% CI	OR	95% CI
Age						
15 - 19	0.37	0.25 – 0.55	0.44	0.24 – 0.82	0.18	0.06 – 0.54
20 – 24	0.55	0.41 – 0.74	0.64	0.41 – 1.00	0.32	0.13 – 0.79
25 – 29	0.79	0.60 – 1.03	0.79	0.52 – 1.20	0.46	0.19 – 1.08
30 – 34	1.01	0.78 – 1.31	1.24	0.84 – 1.84	0.60	0.26 – 1.38
35 – 39	1.79	1.40 – 2.29	1.69	1.16 – 2.48	0.72	0.31 – 1.69
40+ (Ref)	1.0		1.0		1.0	
Education						
No education	0.20	0.16 – 0.24	0.29	0.22 – 0.40	0.14	0.09 – 0.22
Primary	0.59	0.48 – 0.73	0.74	0.55 – 0.98	0.42	0.28 – 0.62
Secondary + (Ref)	1.0		1.0		1.0	
Child ever born						
1 - 2	3.02	2.31 – 3.95	2.34	1.57 – 3.51	3.36	0.62 – 6.96
3 – 4	2.13	1.67 – 2.72	1.94	1.34 – 2.81	1.91	0.98 – 3.72
5 – 6	1.58	1.24 – 2.02	1.36	0.93 – 1.98	1.39	0.72 – 2.69
7+ (Ref)	1.0		1.0		1.0	
RH/FP message						
Yes	1.54	1.33 – 1.79	1.41	1.14 – 1.75	1.56	1.13 – 2.16
No (Ref)	1.0		1.0		1.0	
Women's participation in health care decisions						
Yes	1.22	1.04 – 1.44	1.30	1.01 – 1.66	1.25	0.87 – 1.78
No (Ref)	1.0		1.0		1.0	

Non-inclusion of one in the 95% CI indicates statistical significance: Ref – Reference category.

Annex 3. Adjusted Odds Ratios: Maternal Health Services Utilization

95% CI from the logistic regression model for the determinants of maternal health services utilization: Transform Program, Ethiopia, Endline survey, 2022.

Indicators	Early ANC (n=1611; event=566)		ANC 4+ (n=1611; event=407)		SBA (n=1611; event=623)	
	OR	95% CI	OR	95% CI	OR	95% CI
Age						
15 - 19	0.42	0.19 – 0.96	0.17	0.07 – 0.41	0.28	0.10 – 0.77
20 – 24	0.59	0.29 – 1.18	0.25	0.12 – 0.51	0.45	0.19 – 1.06
25 – 29	0.83	0.44 – 1.60	0.51	0.26 – 0.98	0.52	0.23 – 1.14
30 – 34	1.06	0.57 – 1.98	0.57	0.30 – 1.07	0.74	0.35 – 1.60
35 – 39	0.88	0.46 – 1.70	0.55	0.28 – 1.06	1.22	0.56 – 2.67
40+ (Ref)	1.0		1.0		1.0	
Education						
No education	0.18	0.12 – 0.27	0.23	0.15 – 0.34	0.31	0.19 – 0.52
Primary	0.46	0.31 – 0.68	0.66	0.45 – 0.96	0.59	0.35 – 0.98
Secondary + (Ref)	1.0		1.0		1.0	
Child ever born						
1 - 2	2.22	1.31 – 3.76	1.89	1.07 – 3.35	5.07	2.61 – 9.87
3 – 4	1.78	1.11 – 2.81	1.60	0.97 – 2.65	2.50	1.41 – 4.42
5 – 6	1.58	1.01 – 1.83	1.49	0.92 – 2.41	1.93	1.11 – 3.32
7+ (Ref)	1.0		1.0		1.0	
MCH message						
Yes	1.39	1.06 – 1.83	1.59	1.20 – 2.12	1.54	1.09 – 2.18
No (Ref)	1.0		1.0		1.0	
Spouse accompany during ANC visit or delivery						
Yes	3.84	3.00 – 4.94	3.62	2.79 – 4.71	32.96	23.62 – 46.00
No (Ref)	1.0		1.0		1.0	
Women's participation in health care decisions						
Yes	1.22	0.93 – 1.60	1.07	0.79 – 1.43	0.92	0.66 – 1.29
No (Ref)	1.0		1.0		1.0	

Non-inclusion of one in the 95% CI indicates statistical significance: Ref – Reference category.

Annex 4. Adjusted Odds Ratios: Child Health and Immunization

95% CI from the logistic regression model for the determinants of child health and immunization: Transform Program, Ethiopia, Endline survey, 2022.

Indicators	Full vaccination (n=1241; event=223)		Fever treatment (n=943; event=207)		Deworming (n=4829; event=1756)	
	OR	95% CI	OR	95% CI	OR	95% CI
Age						
15 - 19	0.52	0.17 – 1.60	1.61	0.62 – 4.21	0.58	0.34 – 0.96
20 – 24	1.25	0.51 – 3.02	1.57	0.77 – 3.24	0.73	0.55 – 0.98
25 – 29	1.68	0.73 – 3.88	1.38	0.72 – 2.65	0.96	0.75 – 1.24
30 – 34	1.46	0.63 – 3.37	1.36	0.73 – 2.55	1.03	0.81 – 1.31
35 – 39	2.52	1.11 – 5.72	0.96	0.49 – 1.87	1.47	1.16 – 1.87
40+ (Ref)	1.0		1.0		1.0	
Education						
No education	0.19	0.12 – 0.31	1.48	0.84 – 2.60	0.29	0.23 – 0.37
Primary	0.55	0.35 – 0.86	1.06	0.59 – 1.93	0.96	0.75 – 1.23
Secondary + (Ref)	1.0		1.0		1.0	
Child ever born						
1 - 2	1.55	0.83 – 2.90	1.95	1.04 – 3.67	1.50	1.18 – 1.90
3 – 4	1.31	0.73 – 2.32	1.20	0.67 – 2.11	1.41	1.15 – 1.72
5 – 6	1.22	0.69 – 2.17	1.57	0.92 – 2.69	1.29	1.06 – 1.57
7+ (Ref)	1.0		1.0		1.0	
MCH message						
Yes	1.38	0.98 – 1.94	1.77	1.23 – 2.54	1.41	1.21 – 1.64
No (Ref)	1.0		1.0		1.0	

Non-inclusion of one in the 95% CI indicates statistical significance: Ref – Reference category.