# Evaluation of Two Programs Supporting Global Family Planning Data Needs

## Assessing Achievements, Informing Future Directions

Melinda Moore, Laura J. Faherty, Shira H. Fischer, Kathryn E. Bouskill, Julie DaVanzo, Claude Messan Setodji, Bill Gelfeld, Emily Hoch, Luke J. Matthews, Sarah Weilant, Michele Abbott, Gabriela Armenta, Rouslan I. Karimov, Adeyemi Okunogbe, Uzaib Saya, Mahlet A. Woldetsadik

RAND CORPORATION

For more information on this publication, visit www.rand.org/t/RR2112

Published by the RAND Corporation, 2018, Santa Monica, Calif.

RAND® is a registered trademark.

This work is licensed under a Creative Commons Attribution 4.0 International License. All users of the publication are permitted to copy and redistribute the material in any medium or format and transform and build upon the material, including for any purpose (including commercial) without further permission or fees being required. For more information, see https://creativecommons.org/licenses/by/4.0/

*Cover image: Susan Elden/Department for International Development/Flickr*

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Support RAND
Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

# Preface

Family planning helps countries achieve manageable levels of population growth through voluntary choices about the number and timing of pregnancies, making family planning programs an important contributor to economic development. Regulating fertility through safe and effective contraception also has numerous health benefits for both the mother and infant (Seltzer, 2002). Family planning programs must both deliver commodities and services and collect data to track progress toward national goals. In 2012, countries agreed on an ambitious global goal of achieving, by 2020, 120 million new users of modern contraception in 69 of the world's poorest countries. This Family Planning 2020 (FP2020) goal, and the initial commitment by at least two dozen of those countries, triggered support by major donors for various programs to help countries advance toward the FP2020 goal.

Among such efforts, the Bill & Melinda Gates Foundation launched two key programs in early 2013 to help countries collect, analyze, and use data to monitor progress toward the FP2020 goal. The Performance Monitoring and Accountability 2020 (PMA2020) program, implemented by Johns Hopkins University's Gates Institute for Population and Reproductive Health, was to focus on supporting data collection in nine countries through annual, rapid-turnaround, national surveys typically led by university-based experts, using mobile phone technology and local "resident enumerators." The Track20 program, implemented by Avenir Health, was to work with governments in a larger number of countries to gather data from various sources, analyze and model the data to derive estimates for core family planning indicators, facilitate consensus around data to be reported globally, and promulgate the effective use of these data.

In early 2017, roughly the midpoint between the launch of these programs and the 2020 target date, the Gates Foundation sought to take stock of the progress of PMA2020 and Track20 in order to inform its future directions. It contracted with the RAND Corporation to undertake an objective external evaluation of the two programs. The study team conducted the evaluation from April through September 2017. The findings should be of interest to the Bill & Melinda Gates Foundation, the two programs, the governments of countries participating in one or both programs, associated donors and implementing partners, and the larger global family planning and development communities.

This research was performed as part of the Population Health program within RAND Health. RAND Health has built an international reputation for conducting objective, high-quality, empirical research to support and improve policies and organizations around the world. Its work focuses on a wide array of domestic and international policy areas, including quality of care, health promotion, financing, organization, public health preparedness, domestic and international health care reform, and military health policy. A profile of RAND Health, abstracts of its publications, and ordering information can be found at www.rand.org/health.

# Contents

**Preface** ..... iii
**Figures and Tables** ..... ix
**Summary** ..... xi
**Acknowledgments** ..... xxix
**Abbreviations** ..... xxxi

CHAPTER ONE
**Introduction** ..... 1

CHAPTER TWO
**Background** ..... 3
Overview ..... 3
PMA2020 ..... 3
Track20 ..... 5
Family Planning Data Sources ..... 8

CHAPTER THREE
**Methods** ..... 11
Overview ..... 11
Assessment Frameworks ..... 11
Stakeholder Interviews and Qualitative Analyses ..... 13
Statistical Analyses of PMA2020 ..... 18

CHAPTER FOUR
**Assessment Frameworks** ..... 23
Overview ..... 23
PMA2020 and Track20 Logic Models ..... 23
The RAND Data Maturity Framework ..... 26
The RAND Sustainability Framework ..... 30

CHAPTER FIVE
**Stakeholder Views on Family Planning Data Needs** ..... 33
Overview ..... 33
Desired Frequency of Family Planning Data ..... 33
Desired Geographic Granularity of Family Planning Data ..... 34
Summary ..... 35

CHAPTER SIX
**PMA2020 Goals, Accomplishments, and Challenges** ..... 37
Overview ..... 37
Goals of PMA2020 ..... 37
Accomplishments of PMA2020 ..... 40
Challenges for PMA2020 ..... 41
Summary ..... 45

CHAPTER SEVEN
**Statistical Properties of PMA2020** ..... 49
Overview ..... 49
Summary of Country-Specific Analyses ..... 50
Synthesis of Findings Across the Five Countries ..... 52
Summary ..... 54

CHAPTER EIGHT
**PMA2020 Survey Design** ..... 61
Overview ..... 61
Stakeholder Perceptions of PMA2020's Survey Design ..... 61
Potential Changes to the PMA2020 Survey ..... 64
Summary ..... 77

CHAPTER NINE
**Track20 Goals, Accomplishments, and Challenges** ..... 79
Overview ..... 79
Goals of Track20 ..... 79
Accomplishments of Track20 ..... 81
M&E Officers ..... 82
Other Challenges Faced by Track20 ..... 86
Summary ..... 87

CHAPTER TEN
**Interactions Between PMA2020 and Track20** ..... 89
Overview ..... 89
Need for a Shared Agenda ..... 89
The Power of a Common Mission ..... 90
Highly Variable Interactions Between the Programs in Different Countries ..... 90
Summary ..... 91

CHAPTER ELEVEN
**Data Use** ..... 93
Overview ..... 93
The Importance of Data and Capacity to Use Them ..... 93
Examples of Using PMA2020 Data ..... 94
Facilitators of and Barriers to Using PMA2020 Data ..... 96
Examples of Using Track20 Estimates ..... 99

Facilitators of and Barriers to Using Track20 Estimates ..... 100
Summary ..... 102

CHAPTER TWELVE
**Data Maturity** ..... 105
Overview ..... 105
Qualitative Analyses of Interviews with In-Country Stakeholders ..... 105
Analyses of Data Maturity Ratings from the 15 Program Countries ..... 111
Summary ..... 125

CHAPTER THIRTEEN
**Sustainability** ..... 127
Overview ..... 127
Qualitative Analyses from In-Country Stakeholder Interviews ..... 128
Analyses of Sustainability Ratings from the 15 Program Countries ..... 134
Summary ..... 148

CHAPTER FOURTEEN
**Conclusions and Recommendations** ..... 149
Overview ..... 149
Conclusions Underpinning RAND's Recommendations ..... 149
Recommendations ..... 154
Summary ..... 166

APPENDIXES
**A. Details of PMA2020 Sampling Procedures** ..... 169
**B. Indicators Collected by PMA2020, FP2020, and DHS** ..... 175
**C. Comparison of Selected Family Planning Data Sources** ..... 181
**D. Additional Background on Logic Models** ..... 183
**E. Additional Background on Data Maturity Models** ..... 185
**F. Additional Background on Sustainability Enablers** ..... 187
**G. Contextual Information on the 15 Countries Evaluated** ..... 191
**H. Additional Detail on Statistical Methods and Country-Specific Analyses** ..... 205

**References** ..... 229

# Figures and Tables

## Figures


S.1. Fifteen Program Countries Included in RAND Evaluation ..... xiv
S.2. Data-Driven Accountability Cycle as a Foundation for Data Maturity and Sustainability ..... xxii
S.3. Logic Model for Potential Future Directions for PMA2020 and Track20 with Proposed DATA-FP ..... xxvii
4.1. Initial PMA2020 Logic Model ..... 24
4.2. Initial Track20 Logic Model ..... 25
6.1. Empiric PMA2020 Logic Model ..... 47
8.1. Mobile Cellular Subscriptions per 100 People in PMA2020 Countries, 1996–2015 ..... 72
9.1. Empiric Track20 Logic Model ..... 88
14.1. Data-Driven Accountability Cycle as a Foundation for Data Maturity and Sustainability ..... 153
14.2. Logic Model for Potential Future for Gates Foundation Family Planning Data Programs ..... 158
14.3. Logic Model for Potential Future Directions for PMA2020 and Track20 with Proposed DATA-FP ..... 167


## Tables


2.1. PMA2020 Survey History, by Country ..... 5
2.2. Summary of Track20 M&E Officers in the 15 Countries RAND Evaluated ..... 7
3.1. U.S.-Based Interviewees ..... 14
3.2. Countries Included in the RAND Evaluation ..... 14
3.3. Interviewees in PMA2020 and Track20 Program Countries ..... 16
3.4. Level at Which Data Are Representative and Margin of Error, by Country, Used in Gates Institute Sample Size Calculations ..... 21
4.1. RAND Data Maturity Framework Definitions ..... 26
4.2. RAND Data Maturity Framework for PMA2020 and Track20 ..... 27
4.3. RAND Sustainability Framework: Enabling Factors and Associated Measures/Indicators ..... 31
7.1. Countries Included in the Statistical Analyses ..... 50
7.2. Design Effect and Optimal Design for mCPR, by Round ..... 53
7.3. Between-Round Differences in mCPR for Different Time Intervals ..... 55

7.4. Indicator Estimates and Significance of Changes Across Rounds ..... 56
12.1. Consolidated Data Maturity Ratings of PMA2020 and Track20, by Stakeholder Group ..... 112
12.2. Ratings by All Stakeholders of Key Data Maturity Domains and Areas, by Country ..... 114
12.3. PMA2020 Personnel Ratings of Key Data Maturity Domains and Areas, by Country ..... 118
12.4. Track20 Personnel Ratings of Key Data Maturity Domains and Areas, by Country ... 120
12.5. Government/NGO Ratings of Key Data Maturity Domains and Areas, by Country ..... 122
12.6. Overall Data Maturity Ratings, by Respondent Group and Country ..... 125
13.1. Sustainability Ratings of PMA2020 and Track20, by Factor and Stakeholder Group ..... 135
13.2. Ratings from All Stakeholders of Sustainability-Enabling Factors, by Country ..... 137
13.3. PMA2020 Personnel Ratings of Sustainability-Enabling Factors, by Country ..... 141
13.4. Track20 Personnel Ratings of Sustainability-Enabling Factors, by Country ..... 143
13.5. Government/NGO Personnel Ratings of Sustainability-Enabling Factors, by Country ..... 145
13.6. Ratings of Sustainability-Enabling Factors, by Respondent Group and Country ..... 147
B.1. FP2020 Core Indicators Captured by PMA2020 ..... 176
B.2. Definition of PMA2020 Family Planning Indicators ..... 177
B.3. Key Indicators Measured by PMA2020 and Other Selected Surveys ..... 179
C.1. Comparison of Selected Family Planning Data Sources ..... 181
E.1. Review of Existing Data Maturity Models ..... 185
G.1. Demographic, Economic and Social Profile of the 15 Countries ..... 191
G.2. PMA2020, DHS, and MICS Surveys and DHIS2 Use in the 15 Countries ..... 192
H.1. Ghana PMA2020 Rounds and Sample Size ..... 207
H.2. Ghana Population Characteristics—DHS and PMA2020 ..... 208
H.3. Ghana: Design Effect and Optimal Design Gain—Round-to-Round Comparisons ... 211
H.4. Ethiopia PMA2020 Rounds and Sample Size ..... 212
H.5. Ethiopia Population Characteristics—DHS and PMA2020 ..... 213
H.6. Ethiopia: Design Effect and Optimal Design Gain—Round-to-Round Comparisons ..... 215
H.7. DRC—Kinshasa PMA2020 Rounds and Sample Size ..... 216
H.8. DRC—Kinshasa Population Characteristics—DHS and PMA2020 ..... 217
H.9. DRC—Kinshasa: Design Effect and Optimal Design Gain ..... 219
H.10. Nigeria PMA2020 Rounds and Sample Size ..... 220
H.11. Nigeria Population Characteristics—DHS and PMA2020 (Kaduna and Lagos only) ..... 222
H.12. Nigeria—Kaduna and Lagos: Design Effect and Optimal Design Gain ..... 223
H.13. Kenya PMA2020 Rounds and Sample Size ..... 224
H.14. Kenya Population Characteristics—DHS and PMA2020 ..... 225
H.15. Kenya: Design Effect and Optimal Design Gain ..... 227

# Summary

## Introduction and Background

Family planning is an essential strategy for reducing maternal and infant morbidity and mortality and enhancing social and economic development in developing countries. At a summit convened in London in July 2012, the global community set an ambitious goal of achieving, by 2020, 120 million additional users of modern contraception in 69 low-income countries. In contrast to traditional contraception methods, such as rhythm or withdrawal, modern contraception includes both reversible and nonreversible approaches, such as condoms, pills, intrauterine devices (IUDs), and sterilization. The modern contraceptive prevalence rate (mCPR) is the percentage of women of childbearing age who use some form of modern contraception. The numeric target for the Family Planning 2020 (FP2020) goal was set based on national populations, national mCPR rates at the time (2012), and projected increases in mCPR. The global representatives at the summit recognized that achieving this ambitious goal would require major efforts by many different stakeholders to increase the demand for and broaden population coverage of family planning commodities and services. It would also require timely, accurate, and useful data to track progress on a more-frequent basis than that provided through the Demographic and Health Surveys (DHSs) carried out in most of these countries every five years.

In 2013, the Bill & Melinda Gates Foundation launched two complementary programs to help monitor annual progress toward the FP2020 goal. The Performance Monitoring and Accountability 2020 (PMA2020) program was designed to *generate data* through (at a minimum) annual, rapid-turnaround, nationally representative surveys of households and service delivery points in nine countries, using mobile phone technology for data collection. The Track20 program was designed to support global standardization and reporting of family planning indicators and, in 22 countries, *draw upon data from various sources (including data from PMA2020 surveys) to produce estimates of those indicators* through Bayesian modeling.

In early 2017, at roughly the midpoint between the launch of PMA2020 and Track20 and the FP2020 target date, the Gates Foundation contracted with the RAND Corporation to assess its investments in these programs and identify potential future directions. It specified that the evaluation should address four questions:

1. How well are the PMA2020 and Track20 programs functioning now?
2. What changes to the design of PMA2020 surveys might improve their use?
3. How do stakeholders view PMA2020 and Track20 (including their perceptions of the level of data maturity in program countries)?
4. Are the two programs sustainable?

This report describes the process and results of RAND's evaluation.

### PMA2020

Stemming directly from the 2012 London Summit on Family Planning, PMA2020 was designed to support nationally representative surveys on which to base annual progress reports of core FP2020 indicators. PMA2020 is implemented by the Gates Institute for Population and Reproductive Health at Johns Hopkins University (hereafter Gates Institute), in collaboration with in-country partners.

As part of the evaluation, RAND researchers assessed PMA2020 achievements against the program's objectives, including both the original objectives (for April 2013 through March 2018) and the Gates Institute's modified set of objectives from 2015:

Original objectives:

1. Expand country-level monitoring capacity (also Track20).
2. Integrate a rapid data collection system using mobile devices.
3. Improve data monitoring to strengthen service delivery (also Track20).
4. Promote use of data to respond to family planning needs at the community level.

Modified objectives:

1. Generate survey data.
2. Build a sustainable business model.
3. Progress toward survey sustainability and financing.
4. Integrate PMA2020 into countries' monitoring and evaluation (M&E) architecture.

PMA2020 was launched in Ghana in 2013 and, most recently (in 2017), in Côte d'Ivoire. Surveys are typically fielded twice per year for two years and annually thereafter. Most of the surveys produce national-level estimates for the desired FP2020 core indicators; in some countries, surveys are carried out only in selected states/provinces. Using mobile technology, the surveys collect demographic information about households, women's reproductive history and use of contraception, and supply of reproductive commodities and services. A unique feature of PMA2020 is that the surveys are carried out by local data collectors (resident enumerators), who are women over the age of 21 from the surveyed area or nearby, enabling efficient data collection and management. PMA2020 consists of surveys of households and service delivery points (i.e., facilities that provide family planning services) in the same areas.

As of July 2017, at least one round of PMA2020 surveys had been carried out or was under way in 11 countries.

### Track20

Track20 was designed as a global resource to monitor progress toward FP2020 goals by standardizing and reporting on national family planning indicators. It was also intended to help countries develop their capacity to collate and analyze data, facilitate consensus around estimates for key FP2020 indicators to be used for monitoring global progress toward the FP2020 goal, and promote the use of family planning data by country-level decisionmakers. While the initial Track20 proposal specified working in 22 countries, as of summer 2017, the program

was active in 37 of the 69 poorest countries targeted by FP2020. Track20 is implemented by Avenir Health (hereafter Avenir).

A core feature of Track20 is the M&E officer model, with one or more M&E officers in each program country. These personnel ideally have sophisticated quantitative skills, considerable experience related to data analysis and use, and strong interpersonal skills. In roughly 20 countries, the Track20 program pays part or all of officers' salaries. Avenir supports and helps strengthen the technical capacity of M&E officers, supports (at least) annual national data consensus meetings among key stakeholders in each country, and partners with the United Nations Population Division to expand statistical modeling of family planning indicator estimates using the Family Planning Estimation Tool (FPET). FPET is a statistical model that estimates the likelihood of a given result using prior observed values and incorporates all available survey data and service statistics that are of sufficiently high quality. Track20 also examines country-level family planning expenditures using data from multiple sources, and Avenir supports M&E officers to improve strategic planning around family planning through its FP Goals model. The model incorporates demographic data, program information, and evidence of intervention effectiveness to help decisionmakers set goals and prioritize family planning investments.

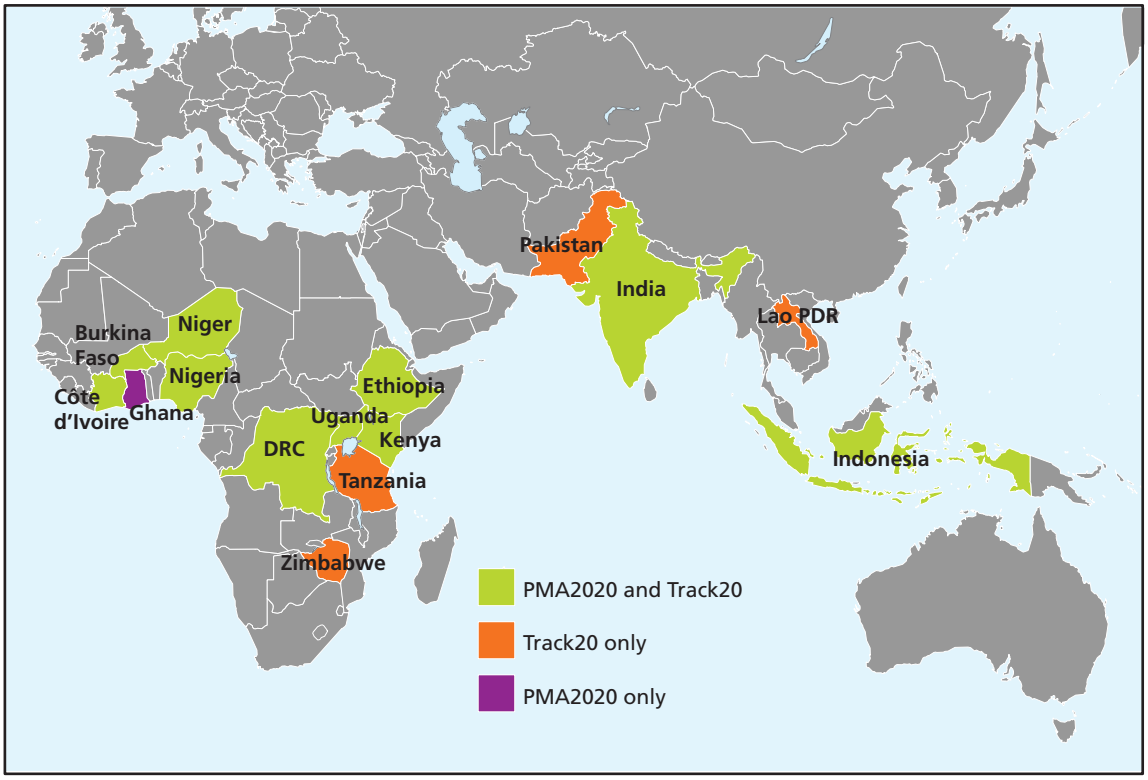
## Methods

To evaluate the performance of PMA2020 and Track20, we adopted a mixed-methods approach. After reviewing key documents, we conducted discussions with the two grantee organizations and interviewed stakeholders based in the United States and in program countries. We analyzed the statistical properties of the PMA2020 survey as requested by the Gates Foundation (focusing on five countries with multiple rounds of available data). We asked stakeholders to provide numeric ratings of their country's performance on various elements of the data maturity and sustainability assessment frameworks we developed for the evaluation.

Together with colleagues at the Gates Foundation and the two grantee institutions, we finalized the 15 countries to be included in the evaluation (Figure S.1). Eleven are in Africa, and four are in Asia; ten have both PMA2020 and Track20 programming, four have Track20 only, and one has PMA2020 only. The countries vary greatly in population size, but all are classified as low income or lower middle income by the World Bank, and most rank quite low on the Human Development Index.

We conducted **nearly 200 semistructured interviews in these countries** with three broad stakeholder groups: PMA2020 staff; Track20 staff; and a group that included government officials and representatives from bilateral organizations, multilateral organizations, and nongovernmental organizations (NGOs) with family planning programming. The semistructured interview protocols included questions about the two programs and about family planning data needs and data use. To complement the interviews with in-country stakeholders, we also conducted (in person or by phone) **nearly 40 interviews with U.S.-based stakeholders** representing various perspectives: staff from the Gates Foundation and the two grantee organizations; members of the PMA2020 External Consultative Group, including individuals based in such organizations as the U.S. Agency for International Development, FP2020, the United Nations Population Fund, and academic institutions; other family planning experts; and statistical experts.

**Figure S.1**
**Fifteen Program Countries Included in RAND Evaluation**




NOTES: DRC = the Democratic Republic of the Congo; Lao PDR = Lao People's Democratic Republic.
RAND *RR2112-S.1*

As part of our evaluation of PMA2020 and Track20, we developed three frameworks: a **logic model for each program**, a **data maturity framework**, and a **sustainability framework**. We used the logic models to guide the development of the stakeholder interview protocols and subsequently modified them to reflect observations from U.S. and country stakeholders and then to reflect our recommendations for future program directions. At the request of the Gates Foundation, we developed and applied a framework to assess data maturity associated with PMA2020 and Track20 in program countries. The framework includes several specific areas within three main domains: *organizational readiness*, *data systems*, and *data use.* We developed a sustainability framework for PMA2020 and Track20 that organizes various sustainability-enabling factors into four broad domains—*financial*, *technical*, *operational/programmatic*, and *data culture.*

## Findings

### Stakeholder Views on Family Planning Data Needs

To provide a baseline understanding of the potential contributions of PMA2020 and Track20, we asked in-country decisionmakers and PMA2020 and Track20 program staff about their data needs for decisionmaking, including how often they thought data should be collected and their preferences for national and/or subnational estimates. Our purpose was to explore

the extent to which PMA2020, other family planning data sources, and Track20 are producing the kind of data that decisionmakers find useful and whether or how they use these data.

We found that, overall, **the desired frequency of data collection varied by type of data**. Stakeholders called for **more-frequent collection of service statistics** (that is, data on health facility–based commodity supplies and provision of family planning services)—at least every six months, but ideally quarterly or even monthly—and less-frequent collection of survey data. **Annual PMA2020 surveys were sufficient** for their purposes. They also found service statistics to be particularly useful but of variable quality. Finally, data users expressed a clear need for subnational family planning data in addition to national estimates, in order to inform local decisionmaking.

### PMA2020 Goals, Accomplishments, and Challenges

Having gathered information about family planning data needs, we sought stakeholder views of PMA2020's goals, accomplishments to date, and the challenges the program faces going forward. In general, stakeholders felt that **PMA2020 has laid the foundation for collecting high-quality data** on family planning. They valued the fact that the **PMA2020 is fielded (at least) annually** and that **data are quickly available** for monitoring family planning activities.

However, stakeholders also felt that **PMA2020 has not yet met some of its original or subsequently modified objectives**—in particular, promoting the use of data to enhance evidence-based decisionmaking, responding to family planning needs at the local level, ensuring sustainability, and integrating PMA2020 into countries' M&E architecture. They regarded these yet-to-be achieved goals as opportunities for future efforts, including better links to government decisionmakers; facilitating data use for planning, resource allocation, and program management (and, secondarily, for research purposes); and modifying the survey design to make it more useful to decisionmakers at all levels. Stakeholders viewed this evaluation as an opportunity to clarify PMA2020's vision and scope (that is, to reassess the survey's goals and potential uses), to acknowledge PMA2020's notable achievements to date, and to build on those achievements in pursuing goals for the future.

#### *The Design of PMA2020*

In addition to examining big-picture achievements and challenges for PMA2020, we also analyzed the survey's statistical properties, integrating quantitative analysis with qualitative perspectives of stakeholders. We explored the representativeness of the data (which has implications for reporting national estimates), determined the margin of error for key indicator estimates (which has implications for survey sample size), and examined intra-class correlation and design effect (which reflect how similar individuals within a given cluster are with respect to their characteristics and their responses, with implications for the ideal number of clusters and respondents within each one). We also examined differences in indicator values at different intervals (six, 12, 18, and 24 months) to explore how changing the frequency of data collection might affect the survey's ability to detect statistically meaningful changes in key indicators.

Our analyses of PMA2020 surveys in five countries with at least three survey rounds (Ghana, Ethiopia, the Democratic Republic of the Congo [DRC], Nigeria, and Kenya) offer insight into how well the survey is meeting its goals:

- While the PMA2020 survey samples may not be representative, as compared with the population surveyed by the "gold standard" DHS, **differences in demographic and**

**socioeconomic characteristics between the PMA2020 and the DHS populations may not necessarily impact estimates of contraceptive indicators.**

- Margins of error for major indicators, such as mCPR, were small by typical survey standards, indicating that **the sample sizes are large enough to produce estimates with the desired level of precision.**
- In some countries, participant characteristics and responses were clustered (relatively homogeneous) in sampled enumeration areas. In statistical terms, intra-class correlation and design effect were high. This clustering results in **substantially smaller effective sample sizes**. To achieve larger effective sample sizes, the program could implement a sampling strategy that samples more clusters with fewer respondents in each cluster.
- All clusters currently have the same number of respondents. Another way to potentially improve efficiency is through an **optimal design scheme** in which different numbers of respondents could be selected within each cluster, depending on the variance of key indicators of interest: more respondents from clusters that are heterogeneous and fewer respondents from clusters that are fairly homogenous. Our results showed that implementing such an optimal design will lead to improved precision and efficiency in some countries.
- Our statistical analysis supports stakeholder views on the desired frequency of data collection, demonstrating that, in most countries, **the household portion of the PMA2020 survey could be fielded every 12 months instead of every six months** (which is done for the first four rounds of data collection) without losing significant information about key indicators. However, stakeholders would like data from the service delivery points more frequently than annually.

To complement the quantitative findings about PMA2020's design, we elicited stakeholder views. Stakeholders in both the United States and PMA2020 countries expressed **concerns about resampling the same enumeration areas** and up to one-third of the same households in serial rounds of PMA2020 surveys. They also expressed skepticism about the survey marketing itself as providing nationally representative estimates when only sampling in a few states in certain countries. Nearly all stakeholders wanted more geographically granular information about key family planning indicators, specifically at the state or provincial level and even at the district level.

We describe several options for potential changes to PMA2020. Ultimately, PMA2020's design should be responsive to the data needs of its users and should fill a gap in the family planning data landscape in the countries in which it operates. One of the most important needs, for which PMA2020 is currently well positioned but would require a change in sampling, is to provide **annual (rather than semiannual) estimates at subnational levels**.

Other promising directions for PMA2020, depending on program priorities, include expanding the pilot of **panel data collection** and incorporating it more systematically, if successful; pursuing more **data collection via mobile phone rather than face to face** to reduce survey costs; implementing **rolling samples** as a way to provide continuous employment for data collectors; and **rotating modules** or administering them in a targeted fashion to particular subgroups of respondents as a way to broaden use of the PMA2020 platform and attract co-financing. In addition, **the service delivery point surveys could be used both to triangulate (i.e., validate) and to provide more context** to routine service statistics.

However, the benefits and drawbacks of changing one element of the survey design depend on decisions about other elements. The Gates Foundation asked for options that assume little change to basic resources, such as funding and the use of resident enumerators for data collection. Additional changes to PMA2020's survey design are possible if the program were considering larger-scale changes, such as moving away from the resident enumerator model.

### Track20 Goals, Accomplishments, and Challenges

Paralleling our evaluation of PMA2020, we solicited stakeholder views of Track20's accomplishments to date, including the work of the M&E officers, and the challenges Track20 faces. Stakeholder interviews reveal that the Track20 approach is **well respected**. The program has **achieved most of its objectives and is on track to reach its original goal**.

Track20 has developed a data-driven and methodologically consistent process to establish consensus around key indicators from a variety of sometimes-conflicting data sources and has built country capacity by embedding M&E officers within the existing structure of ministries of health (MOHs) or comparable institutions. Track20 has the potential to make significant, lasting impacts on family planning data collection, analysis, and use around the world. Its evolving emphasis on developing country-owned agendas through the FP Goals model and strengthening the quality of the data inputs into the FPET model (particularly service statistics) has yielded important successes. Avenir recognizes the need to continue to work with country partners to develop realistic, data-driven, and country-generated costed implementation plans for family planning (defined by FP2020 as multiyear strategic plans with clear action items and associated costs, intended to guide donor investments and government efforts to achieve their family planning goals).

Track20 has achieved a strong balance between standardizing, across varying country contexts, a methodologically sound system for producing consensus estimates (using FPET) from available family planning data sources, while allowing for substantial flexibility and encouraging country ownership. This decentralized model has allowed countries to determine what their specific needs are with respect to technical assistance, to decide how to finance their M&E officer(s), and to define their priorities for improving family planning in their country. Key challenges for Track20, representing opportunities for improvement, are that **M&E officers are often stretched very thin**, there is a need for continued **capacity-building** among decisionmakers to **facilitate data use**, and the quality of the FPET estimates is **impacted by the quality of the various data inputs**, including incomplete or inaccurate service statistics.

### Interactions Between PMA2020 and Track20

The PMA2020 and Track20 programs were originally intended to be "twinned." However, since their respective launches in 2013, **their activities were never really coordinated as intended**, to the point where there is very little interaction between them in some program countries. In interviews with the Gates Foundation, U.S.-based stakeholders, and in-country stakeholders, we sought to understand the extent to which PMA2020 and Track20 complement each other in achieving FP2020 goals and how the programs could work together more effectively.

Respondents had very few concrete suggestions for how the two programs could interact more effectively beyond involving the PMA2020 staff in the Track20-supported national data consensus meetings and ensuring that PMA2020 data are incorporated into Track20 indicator estimates. The strength of collaboration between the two programs varies by country, but

there is general acknowledgement that **more communication, shared activities, and contact between PMA2020 and Track20 would benefit each program**. Going forward, it would prove fruitful to intensify efforts to find common ground. As one respondent commented, "We are on the same team."

### Data Use

The ultimate goal of collecting and analyzing data is using them to inform decisionmaking and action. We asked all stakeholders how important they thought data were for these purposes and how prepared decisionmakers were in their respective countries to understand and use data. We also asked them to describe facilitators and barriers to using PMA2020 data and Track20 estimates and to give us examples of situations in which those resources had been used.

Stakeholder views about the importance of data use varied substantially by stakeholder group and by country. In general, Gates Foundation staff felt that the demand for data had come from the global level, making it more difficult to integrate data at the country level. However, **in-country stakeholders asserted that the data were invaluable for a variety of uses**, including, for example, program planning, development of costed implementation plans, and forecasting commodity needs. Data were also seen as essential to tracking progress against milestones and for demonstrating pressing needs to donors. In-country respondents were candid about the **inadequate capacity of some decisionmakers in their country to interpret data** so that they could use them when making decisions, or even so that they could follow discussion of family planning data with PMA2020 or Track20 staff or with other experts.

Some respondents saw a natural division between PMA2020 as data generators and Track20 as promoters of data use. The Gates Institute shares the view that its responsibilities stop short of advocating for or actively facilitating data use for in-country decisionmaking. However, the institute is convening workshops and hosting meetings to help in-country researchers conduct and publish data analyses, and PMA2020 and Track20 are piloting a collaboration to promote use of service delivery point data.

Respondents commented in general terms about the importance of family planning data but provided **few concrete examples of using either PMA2020 data or Track20 estimates for decisionmaking**. Some of the more specific examples included program planning, prioritizing family planning activities and investments using the FP Goals model, developing costed implementation plans, evaluating commodity stock-outs, improving commodity logistics, and soliciting funding from governments and donors. A clear role for Track20 estimates was that they served as a harmonized figure around which there was consensus. Stakeholders appreciated that Track20's FPET took into account disparate survey results and service statistics to produce more-realistic and high-quality estimates than any one single survey or other data source could produce (although there was little elaboration on how those harmonized consensus estimates were put to use).

When asked to describe facilitators of and barriers to using PMA2020 data, in-country respondents noted **challenges with meeting decisionmakers' data needs**. Compounding this barrier is the perception that decisionmakers do not value, understand, or have the **capacity to use the data**. On the other hand, **trust** in the rigorous nature of the data (with the exception of concern about extrapolating a national estimate from a limited geographic sample), **leadership buy-in**, having **champions** of the survey, and **connections with policymakers** were seen as

facilitators of PMA2020 data use; in contrast, weak connections with policymakers represented an important barrier. The most commonly cited barrier to use was general **lack of awareness** of the PMA2020 survey, highlighting the need for more-effective dissemination to increase the visibility of PMA2020 within countries. Similarly, having an in-country **champion** of the Track20 methodology helped to promote use of Track20 data, while stakeholders noted a **need to build awareness** of and **obtain buy-in** around the FPET model. Decisionmakers emphasized their desire for **subnational estimates** from Track20.

Multiple respondents thought that receptivity to using data for decisionmaking had increased—what they referred to as an improved **data culture**, *particularly as a result of Track20's efforts*. However, not all respondents shared that view. Ultimately, they thought that, in some countries, use of PMA2020 data and Track20 estimates was hindered by a lack of demand for data and a lack of appreciation of why data are needed.

### Data Maturity

Data maturity refers to the extent to which high-quality data are collected, well managed, well governed, rigorously analyzed, shared, communicated, and, ultimately, used. Data maturity models are tools used to evaluate and manage continuous improvement. The Gates Foundation asked RAND to develop a framework to assess data maturity for PMA2020 and Track20. Our framework encompasses multiple factors that we organized into **three broad domains***:* **organizational readiness** (including staffing, leadership and staff buy-in, and infrastructure), **data systems** (including data collection, data management, data analytics, technology, data governance, and institutionalization), and **data use**.

Our data maturity framework is tailored to PMA2020 and Track20. Consistent with industry data maturity models, ours uses a 10-point scale for rating maturity levels within three stages: beginning (scores 1–3), developing (scores 4–7), and advanced (scores 8–10). We developed a comprehensive assessment tool consisting of all elements in our data maturity framework and asked interviewees in program countries to rate their country's performance on each element. We analyzed the data maturity of PMA2020 and Track20, drawing from both qualitative data (from our stakeholder interviews) and quantitative data (from the stakeholder ratings). PMA2020 and Track20, while increasingly well established and highly informed by best practices, fell into the developing maturity level in most countries. The level of data maturity in countries can stem from both **country-specific factors** (e.g., political support for family planning) and **program-specific factors** (e.g., technical capabilities of staff). Stakeholders conveyed to us that even after data generation has been optimized, **further improvements can be made to data management, data analysis, organizational readiness** (in terms of knowing how to use the data for decisionmaking), **institutionalization, ownership, and data use**.

Quantitative analysis of the data maturity ratings points to three groups of countries: a more advanced group (Burkina Faso, Côte d'Ivoire, Ethiopia, Ghana, and India), a middle-of-the-road group (Indonesia, Kenya, and Nigeria), and a group that remains in the nascent stages of data maturity (DRC, Lao PDR, Niger, Pakistan, Tanzania, Uganda, and Zimbabwe, with DRC as an outlier well below the entire group of countries included in this evaluation). However, given that data maturity is highly country-specific, our framework is perhaps more appropriate for monitoring progress within countries over time than for comparisons across countries.

Stakeholders were **enthusiastic about striving to improve data maturity elements** that are lagging in their countries. The factors impacting data maturity in each country and for

each program are highly variable, and the programs can directly influence some factors (e.g., organizational readiness, data system processes) more than others (e.g., governance).

### Sustainability

Sustainability is a nearly universal priority in the development community but is often difficult to define precisely and measure. Here the RAND team has defined the sustainability of a data system as its endurance over time, which is directly related to the degree of its institutionalization within the routine functioning of the government. Our assessment of sustainability of PMA2020 and Track20 programs included qualitative analysis of our stakeholder interviews and quantitative analysis of stakeholders' ratings of various factors within the sustainability framework we developed for this evaluation. The qualitative interview data and quantitative data from the sustainability ratings provide complementary views on enablers and barriers to PMA2020 and Track20 sustainability in these countries.

Our framework includes **sustainability-enabling factors within four key domains**: **financial sustainability, technical sustainability, operational sustainability, and data culture**. We consider these factors as key to the sustainability of PMA2020 and Track20 and a means to measure progress over time.

*Financial sustainability* (framed as co-financing) is of critical concern to the Gates Foundation, grantee organizations, and program countries and is **particularly salient for PMA2020** because of the resource intensity of launching PMA2020 surveys in a new country and then ensuring that they run smoothly in subsequent rounds of data collection. There is currently very little domestic resourcing of these programs, with the notable exception of in-kind resourcing—e.g., paid government personnel who serve as Track20 M&E officers and university personnel who serve as PMA2020 principal investigators.

*Technical sustainability* is also a concern. Numeric ratings within this domain of sustainability highlighted the **need to ensure hardware and software maintenance**, and interview data consistently highlighted the **need for more well-trained M&E personnel**.

Stakeholders provided the most feedback on *operational sustainability* and *data culture*. The strongest sustainability-enabling factors for operational sustainability were **leadership buy-in** and **cultural acceptability**; the weakest were **satisfaction of policymakers' needs, accountability in using data to inform policy, engagement of local communities and civil society, and use of local expertise**. The strongest sustainability-enabling factor for data culture was **data impact on outcomes**; the weakest was **institutionalization of data use**, including inadequate numbers of people with sufficient technical capabilities. Our interviews signaled that some key stakeholders do not understand the data or know how to use them effectively. This complementary information from the qualitative and quantitative analyses suggests specific actions for PMA2020 and Track20 as they move into their next grant cycle.

Moreover, the data indicate countries that are doing well with regard to operational sustainability and data culture (e.g., India, Uganda) and those that are doing less well (e.g., DRC, Tanzania); the data also identify specific barriers in a given country, again suggesting areas for focused attention. One theme that emerged clearly from these analyses was the **need to plan for sustainability and take deliberate actions to help enable it**.

## Recommendations and Conclusions

We evaluated the functioning of the PMA2020 and Track20 programs, assessed the statistical properties of PMA2020 surveys, gathered stakeholder perceptions of the two programs, and assessed the data maturity of countries and the sustainability of the two programs. We synthesized qualitative input from the more than 260 stakeholders we interviewed in the United States and Canada and 15 program countries and quantitative analyses related to PMA2020 statistical properties and numeric ratings related to both data maturity and sustainability. Based on this synthesis, we offer recommendations, both overarching and specific to each program, for future actions.

**The achievements of the programs to date are significant.** Our recommendations, reflecting different perspectives and based on different methods, suggest **opportunities to further enhance their contributions**. Because we found that stakeholders largely perceive PMA2020 and Track20 as driven by and largely serving the global community more than the program countries themselves, our recommendations all reinforce the general **shifting of program focus** from the global level to the country level—**from a mostly "donor-driven agenda" toward an "owner-driven agenda"** (Moore et al., 2012). Deciding whether and how to pursue this shift will be a program-specific effort, as pursuing a strategic reorientation can require additional resources as well as a new mindset among implementing partners in order to change course.

### Overarching Recommendations

#### *1. Promote country-driven agendas.*

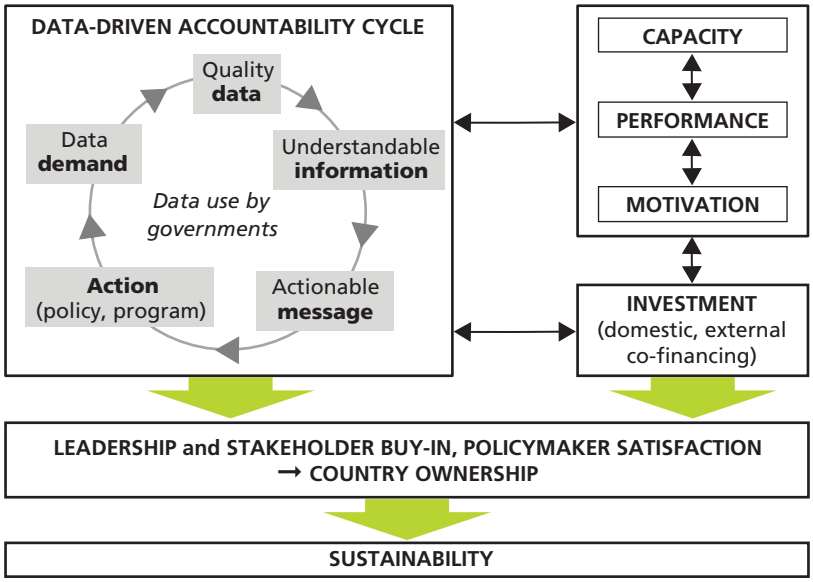
The preponderance of stakeholder feedback suggested that countries have unmet needs that could be addressed through strategic reorientation of the two programs (PMA2020 in particular) to tailor family planning data collection, analysis, and use to each country's needs. Orienting both PMA2020 and Track20 toward country-driven agendas will help to strengthen country ownership of the programs' processes, including data collection, management, analysis, and use. Country ownership entails active engagement with key stakeholders from national and local governments, NGOs, civil society, and multistakeholder working groups. Such groups are often overseen by government, which enhances government ownership of multistakeholder programming. Country ownership also entails planning for transition to full government ownership and institutionalization of data systems (e.g., through mutually agreed-upon expectations and exit strategy for donors).

#### *2. Intensify focus on data use.*

One of the clearest messages that emerged from stakeholder interviews was the need to strengthen data use for decisionmaking and action. Building on the achievements of the first few years of implementation, which focused on generating high-quality data and estimates, both programs should now focus more intensively on promoting data use. This challenge may be more salient for PMA2020, which has focused almost exclusively on data generation and not data use, than for Track20, which has promoted data use in its program efforts. Nonetheless, there are indications of room for improvement in data use at all levels across both programs.

Data must be actively and explicitly transformed into understandable information from which actionable messages are created to help decisionmakers know what appropriate actions they might take. Figure S.2 shows a data-driven accountability cycle, in which high-quality *data*

**Figure S.2**
**Data-Driven Accountability Cycle as a Foundation for Data Maturity and Sustainability**




SOURCE: Adapted from Figure 5 in Moore et al., 2012.

RAND *RR2112-S.2*

are translated into understandable *information*, which is, in turn, packaged into an actionable *message* that prompts policy and program *action*, driving and institutionalizing the *demand* for more data. Such a cycle can drive capacity, performance, and motivation to continue this cycle of accountability, attract investment, and contribute to mature and sustainable data systems that are country-owned. The RAND team recommends that all data presentations, including graphics, tables, or narrative reports, include an interpretive statement (translating data to information) and an actionable message (even if only a set of alternatives to consider) for data users, whether for advocacy or decisionmaking purposes.

***3. Plan for and measure data maturity.***

Stakeholder interviews and ratings indicated that most of the program countries included in our evaluation fall into the developing level of data maturity, suggesting room for improvement. But improvement does not happen automatically. Mature data systems require planning and explicit effort. The RAND team recommends that PMA2020 and Track20 systematically plan to advance each country's data system maturity and use (or adapt) the data maturity framework developed for this evaluation as a tool to enable countries to perform periodic self-assessments to track their progress.

***4. Plan for and measure sustainability.***

Stakeholder interviews and ratings also indicated gaps and opportunities to strengthen sustainability-enabling factors and thereby enhance the prospects for sustainability of PMA2020 and Track20 data systems. But this also requires planning and explicit effort. The RAND research team recommends being proactive in planning for sustainability of PMA2020 and Track20 and for the programs to implement activities accordingly. As noted by

the implementing organizations, it is more challenging to retroactively structure co-financing arrangements and other specific sustainability enablers than to develop them from the outset.

RAND's sustainability framework can serve as a tool for periodically taking stock of progress in each country and addressing bottlenecks along the pathway toward sustainable data systems. The RAND team supports stakeholder suggestions that the programs establish formal, written agreements (such as memoranda of understanding) with governments around ownership and sustainability. These agreements should articulate mutual expectations, the roles and responsibilities of key parties, and the ultimate aim for countries to take greater ownership of the systems and processes as one of the key factors that will enable their sustainability.

#### *5. Institutionalize data capacity development.*

Another clear message from our interviews was the need for larger numbers of qualified personnel to carry out family planning monitoring and data use, from national to subnational levels. As a final overarching recommendation and a unifying effort to institutionalize data capacity and use, RAND proposes establishment of a Data for Action Training Activity for Family Planning (DATA-FP) program. The DATA-FP program would build capacity by increasing the number of people with the high-level skills needed for data system management at all levels—an ever-larger cadre of well-trained M&E personnel who can collectively collect, manage, analyze, interpret, disseminate, and facilitate use of family planning data. This is just one of several vertically oriented Gates Foundation programs to which the proposed DATA program could be applied—others could include DATA-NUT (nutrition), DATA-WASH (water, sanitation, and hygiene), and DATA-MNCH (maternal, newborn, and child health). Use across more program areas would unify data-oriented programming in countries and be a unique opportunity for the Gates Foundation to systematically promote mature and sustainable data systems, support a strong data culture in countries, and facilitate the training of enough staff/capacity to make it all possible.

### Program-Specific Recommendations: PMA2020

#### *1. Reorient and operationalize the program to better align with program objectives.*

Different stakeholders have varying expectations of the program's goals and objectives, which lead to widely ranging opinions on future directions and potential opportunities. The RAND team recommends that the Gates Foundation and PMA2020 implementers reexamine, revise, and reach consensus around the program's vision, goals, objectives, and activities, and that they then, in turn, operationalize their decisions. A good starting point is the set of PMA2020 objectives, including the four original objectives from 2013 and the four revised objectives from 2015. While PMA2020 has largely achieved its objective (from both 2013 and 2015) of carrying out annual, rapid-turnaround surveys that generate high-quality family planning data, it has not fully achieved its original objectives related to building capacity, promoting data use to meet local data needs, or integrating its surveys into country data systems, nor its revised objectives to build a sustainable business model, ensure sustainability of the platform, or integrate the survey into countries' M&E architecture. Clarifying PMA2020's goals and objectives for the future will be necessary to further define what successful data generation *and use* will look like, how to facilitate them, and how to measure them over time. These changes will clarify the extent to which the PMA2020 grantee organization is expected to both generate and facilitate the use of PMA2020 data. The RAND team recommends that the PMA2020 grantee play a key role in disseminating its data, interpreting them with and for decisionmakers, and sharing

the data with key advocates and others through dissemination meetings that also function as training workshops.

A strategic reexamination of goals and potential reorientation will also enable the Gates Foundation to make challenging decisions about future directions, all of which entail trade-offs. For instance, if the Gates Foundation decides to shift its emphasis, as in-country stakeholders have strongly advocated, toward subnational estimates over national estimates, then the survey can target certain regions or states of interest but will be required to stop marketing itself as being intended to produce national estimates to track progress toward a global goal (FP2020).

Alignment among different stakeholders will also help to address the tension between "vertical" family planning data generation and "horizontal" data system strengthening—ideally moving toward the idea of a more "diagonal" approach that accomplishes both (Frenk, 2010). A vertical approach fills important gaps in the availability and quality of family planning data and could address several of the stated needs of decisionmakers, such as more qualitative information; more data on quality of services; additional populations, such as adolescents and men; and more. A horizontal approach seeks to better integrate family planning data generation and use into a larger effort of building data maturity within countries across all health (and related) sectors.

The proposed DATA-FP program described above incorporates PMA2020 and Track20 and marries the virtues of horizontal data system strengthening with vertical family planning programming—i.e., the diagonal approach. The capacity-building feature increases the cadre of well-trained M&E personnel, while PMA2020 data generation remains vertically oriented around family planning narrowly or any other programs associated with non–family planning PMA2020 modules. By working within such a diagonal approach, PMA2020 could position itself to help build M&E capacity for family planning while also continuing to support the generation and use of high-quality data for family planning and potentially also other programs.

***2. Engage key partners in active data dissemination.***

RAND researchers recommend that PMA2020 actively engage key partners to strengthen its data dissemination efforts and further raise its visibility within program countries. There are definite advantages to having the PMA2020 principal investigators located primarily within academia. However, the RAND team heard loud and clear that the university-based location of the principal investigator contributes to the perception among in-country decisionmakers that the PMA2020 survey is "boutique," "academic," and "by and for researchers." The onus is on the PMA2020 in-country partners and their teams to deliberately build strong and lasting connections to governmental decisionmakers (including Track20 M&E officers), and other actual and potential users of its data (e.g., advocacy organizations, donors, NGOs, other researchers).

The evaluation revealed that in-country PMA2020 staff were not always invited to the national data consensus meetings organized by the Track20 M&E officer, suggesting a need for both programs to be more proactive about connecting around their common mission and finding efficient ways to communicate with a shared voice with key decisionmakers.

***3. Enhance PMA2020's survey design.***

There are many options to enhance PMA2020's design, but selecting one or more options will involve weighing different trade-offs and priorities. Based on our integration of family

planning data needs as articulated in stakeholder interviews with our analyses of the statistical properties of the PMA2020 survey, the RAND team offers six recommendations:

- Collect PMA2020 household survey data annually from the outset in a new PMA2020 country rather than every six months in the first two years; if surveys conducted every six months are important for capacity-building purposes, consider conducting them in different subnational jurisdictions.
- Support pilot efforts to implement both a targeted panel survey and a cross-sectional survey and adopt the panel survey component if the pilot testing proves promising.
- Use the resources freed up from decreasing the frequency of data collection to intensify explorations of other innovative but resource-intensive experiments—e.g., a pilot of conducting surveys by telephone, which may include polling.
- Consider enhancing efforts to produce robust subnational estimates to meet decisionmakers' data needs. This could be an important niche for PMA2020 going forward, should it be deemed a program priority. However, countries may have to work within given PMA2020 survey resources and make trade-offs in terms of where and how frequently to conduct subnational surveys. The RAND team also recommends implementing an optimal design (a sampling strategy that chooses the number of survey respondents in each cluster to minimize the variance of the population estimate of the indicator), potentially improving the precision of key indicators, increasing the efficiency of the PMA2020 survey, and reducing the necessary sample size for robust subnational estimates.
- Use the service delivery point surveys to their full potential. Stakeholders believe that the data from these surveys have not been effectively linked to household survey data (as originally intended), nor have they been used effectively to help validate routine service statistics or manage service delivery programs.
- Strengthen the technical support for data set users and ensure a streamlined and user-friendly data download process. For those who do not need or want to work with the full, raw data sets, PMA2020 should continue to support its valuable DataLab tool, which is a web-based data visualization program used to create customized charts with PMA2020 data.

***4. Broaden the PMA2020 platform and seek cost efficiencies to attract co-financing.***

Supporting a broader range of survey content (e.g., through rotating modules addressing other family planning–related topics or other priority health areas, such as nutrition, maternal and child health, and water/sanitation/hygiene) would, arguably, meet a broader range of stakeholder needs and increase the possibility of co-financing, which is an important sustainability enabler. The case against expansion involves the extra program costs that would be required to prepare new questions for collection using PMA2020's survey platform and what some stakeholders view as dilution of the family planning focus of PMA2020 as originally conceived and implemented. However, we recommend broadening the PMA2020 survey platform while redoubling efforts to reduce survey costs and seek co-financing.

### Program-Specific Recommendations: Track20

#### *1. Intensify focus on generating actionable data at different levels in increasingly decentralized health systems.*

The RAND team recommends that Track20 continue to emphasize the important country-level activities that country stakeholders wish to prioritize. In addition to having at least one M&E officer in the national MOH or comparable agency, more officers should be placed at subnational-level positions. Track20 should also continue to promote its FP Goals model and help improve countries' costed implementation plans. While these costed implementation plans are a source of pride within countries that are developing them, there is much room to improve their quality, feasibility, degree of country ownership, and use of family planning data.

#### *2. Improve and expand use of FPET.*

FPET is a core feature of the Track20 program and is one of the keys to its success. Track20 should (1) continue its work with countries to improve their routine service statistics, (2) continue to further understanding of the FPET methodology and theoretical basis for both the users of the tool as well as the consumers of the data (the decisionmakers), (3) expand the number of people who understand and are trained to use FPET, and (4) expand FPET's capacity to produce robust subnational estimates to meet the needs of decisionmakers.

#### *3. Optimize the M&E officer model.*

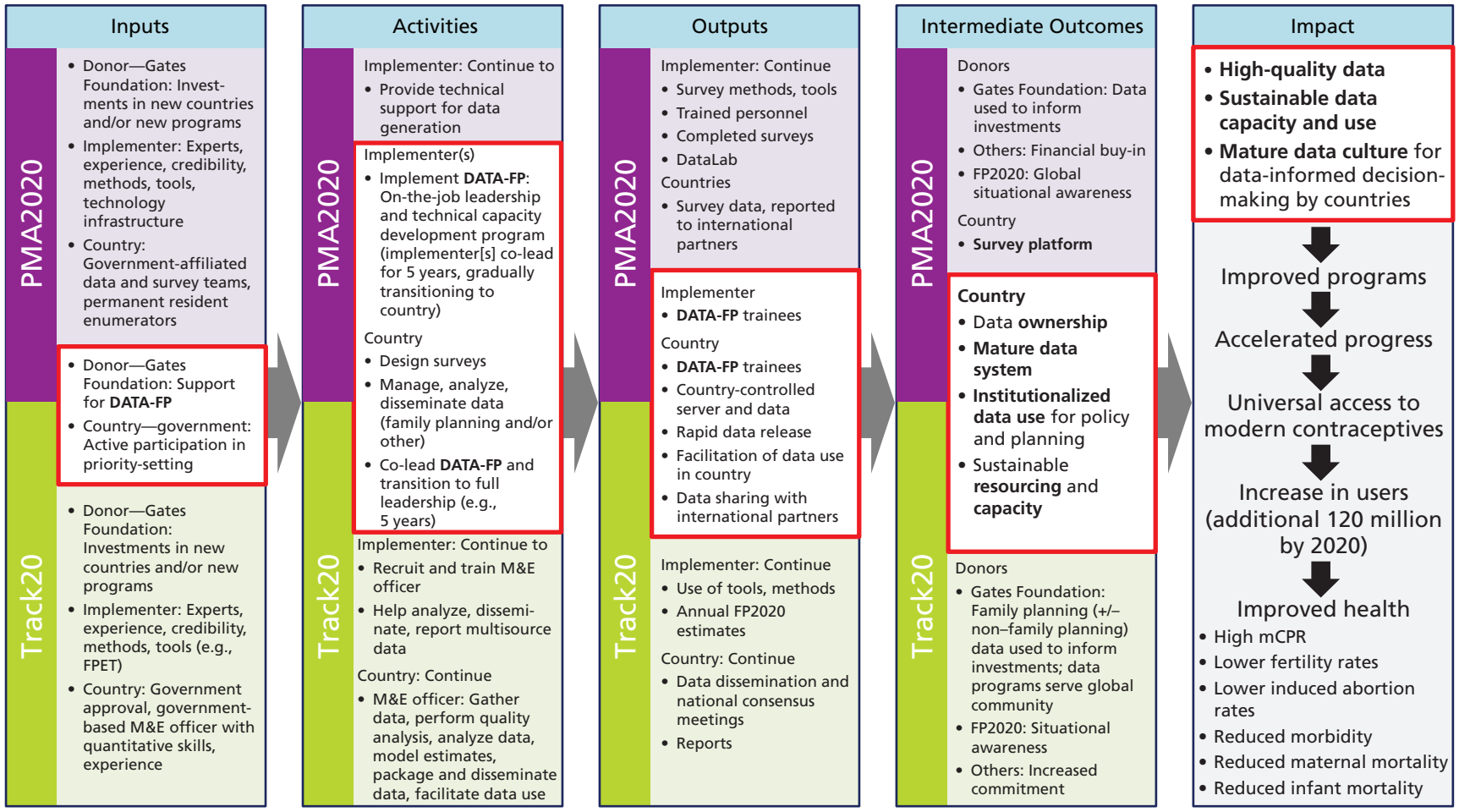
Another of Track20's strengths is its placement of M&E officers within existing governmental structures (primarily MOHs), facilitating their access to key decisionmakers in the family planning arena. The RAND team recommends that Track20 further optimize several dimensions of this unique model. First, Track20 should determine which of the several M&E officer financing models has yielded the best results and replicate this across other countries to the extent possible. Second, as noted above, Track20 should place additional M&E officers at subnational levels, as is occurring in some countries, and continue to test which partnerships with governmental offices yield the best results. Third, Track20 should place M&E officers at an organizational level at which they can be most effective (e.g., not buried within lower units in an MOH). The specific solution will not be the same in every country, but the success of the M&E officers depends on some of these structural factors.

Our interviews suggested that some M&E officers need additional support for their expected duties, including reasonable expectations about their workload; turnaround time for responding to data requests; further professional development, supervision, and guidance commensurate with their experience and skill level; and assistance with connections to policymakers. Track20 should continue to support the efforts of M&E officers to communicate effectively with stakeholders, including ensuring access to them and the tools they need to interpret analytic results and provide actionable messages in relation to those results. Building on existing partnerships with all relevant partners and with data users will help institutionalize the Track20 program within the larger data architecture of program countries.

### Conclusions

The RAND team's overarching recommendations for further empowering countries, strengthening data use, planning for and measuring data maturity and sustainability, and institutionalizing data capacity-building are captured in a logic model reflecting potential futures for these two family planning programs (Figure S.3). The figure also places these broad concepts in the

**Figure S.3**
**Logic Model for Potential Future Directions for PMA2020 and Track20 with Proposed DATA-FP**




RAND *RR2112-S.3*

context of the two programs we evaluated and shows how the proposed new DATA-FP program (outlined in red) could be integrated. Our overarching recommendations reflect ways to build on achievements to date and orient the programs to achieve even more into the future.

The Gates Foundation is uniquely well positioned to support DATA-FP by virtue of its commitment to empower developing country partners, its strong data system orientation, and its broad range of development programs. Government buy-in and expert technical assistance will continue to be required inputs for future PMA2020 and Track20 successes. Activities would include continuing data generation and analysis, strengthening dissemination efforts, focusing more broadly and intensively on data use to meet country needs, improving family planning service statistics, and potentially implementing the proposed DATA-FP program to help grow a cadre of well-qualified M&E personnel. Countries would assume greater leadership and management responsibilities for the data processes, including data presentations and reports. Intermediate outcomes would lead to impacts that include high-quality data to inform program planning, resource allocation, and management; a survey platform that meets a broader range of government and other stakeholders' needs; institutionalized and sustainable country data capacity and use; and a mature data culture.

PMA2020 and Track20 reflect the vision of Gates Foundation leadership, the commitment of participating countries, and the technical expertise of the grantee organizations. Their achievements to date are significant. We have offered recommendations for future program activities that could be implemented in the near term.

The RAND team is grateful for the opportunity to conduct this evaluation, and we thank the many people who informed the findings reported here. We wish the programs continued success in their efforts to provide high-quality family planning data to subnational, national, and global stakeholders while also empowering and enabling countries to mature and sustain their data systems.

# Acknowledgments

We would like to express our great appreciation and thanks for the invaluable information and insights from the numerous experts with whom we consulted in the United States and in PMA2020 and Track20 program countries over the course of this evaluation. We are particularly thankful for the guidance of colleagues from the Bill & Melinda Gates Foundation, under whose auspices the study took place, and, in particular, the direction and invaluable advice of Savitha Subramanian and Win Brown.

We are equally indebted to the numerous PMA2020 and Track20 staff in both the United States and program countries who not only provided valuable information about their respective program but also assisted us in facilitating consultations with other key stakeholders. We wish to thank in particular Emily Sonneveldt, Priya Emmart, John Stover, and Bob Magnani of Avenir Health, as well as Amy Tsui, Scott Radloff, and the team at Johns Hopkins University for responding thoughtfully to our various requests for information and logistical support.

Over the course of this five-month evaluation, we consulted with more than 260 stakeholders. While they are too numerous to thank individually by name, we wish to acknowledge and thank them all for taking the time to share candid information and insights with us.

Finally, we wish to thank Brian Briscombe, Kathryn Pitkin Derose, Jeanne Ringel, and Paul Koegel of RAND and Shawn Malarcher of the U.S. Agency for International Development for their careful quality assurance review and constructive feedback on the report.

## Abbreviations

| | |
|---|---|
| BKKBN | Badan Kependudukan dan Keluarga Berencana Nasional (Indonesian National Population and Family Planning Board) |
| CDC | Centers for Disease Control and Prevention |
| CMMI | Capability Maturity Model Integration |
| CoV | coefficient of variation |
| DATA-FP | Data for Action Training Activity for Family Planning |
| DATA-MNCH | Data for Action Training Activity for Maternal, Newborn, and Child Health |
| DATA-NUT | Data for Action Training Activity for Nutrition |
| DATA-WASH | Data for Action Training Activity for Water, Sanitation, and Hygiene |
| DfID | Department for International Development |
| DHIS | District Health Information Software |
| DHIS2 | District Health Information Software, version 2 |
| DHS | Demographic and Health Survey |
| DRC | Democratic Republic of the Congo |
| FETP | Field Epidemiology Training Program |
| FPET | Family Planning Estimation Tool |
| FP2020 | Family Planning 2020 |
| GPS | Global Positioning System |
| HISP | Health Information Systems Program |
| HIV | human immunodeficiency virus |
| HMIS | health management information system |

| | |
|---|---|
| ICC | intra-class correlation |
| IPUMS-DHS | Integrated Public Use Microdata Series—Demographic and Health Surveys |
| IT | information technology |
| IUD | intrauterine device |
| LAM | lactational amenorrhea method |
| Lao PDR | Lao People's Democratic Republic |
| mCPR | modern contraceptive prevalence rate |
| M&E | monitoring and evaluation |
| MEASURE Evaluation | Monitoring and Evaluation to Assess and Use Results Evaluation |
| MDG | Millennium Development Goals |
| MICS | Multiple-Indicator Cluster Surveys |
| MOH | ministry of health |
| NFHS | National Family Health Survey |
| NGO | nongovernmental organization |
| PMA2020 | Performance Monitoring and Accountability 2020 |
| SARA | Service Availability and Readiness Assessment |
| SDG | Sustainable Development Goals |
| TOC | theory of change |
| UN | United Nations |
| UNFPA | United Nations Population Fund |
| UNICEF | United Nations Children's Fund |
| USAID | U.S. Agency for International Development |
| WHO | World Health Organization |

CHAPTER ONE

# Introduction

Family planning enables people to voluntarily choose the number and timing of pregnancies, thus enabling countries to achieve manageable population growth levels, which in turn facilitate economic development. It is an important strategy for reducing infant and maternal mortality and enhancing social and economic development in low- and middle-income countries (Seltzer, 2002). At a summit hosted by the Bill & Melinda Gates Foundation and the government of the United Kingdom and supported by the United Nations Population Fund (UNFPA) in London in July 2012, the global community galvanized around an ambitious goal of achieving, by 2020, 120 million additional users of modern contraceptive methods (such as oral contraceptives, condoms, intrauterine devices [IUDs], and sterilization) in 69 of the world's poorest countries. The numeric target for the Family Planning 2020 (FP2020) goal was set based on national populations, each country's modern contraceptive prevalence rate (mCPR) at the time, and projected increases in mCPR. The global representatives at the summit recognized that achieving this ambitious goal will require major efforts by many different stakeholders to increase the demand for and broad population coverage of family planning commodities and services, as well as timely, accurate, and useful data to track progress on a more frequent basis than that provided through the Demographic and Health Surveys (DHSs) carried out in most of these countries every five years.

The Gates Foundation has brought vision, significant investment, and a focus on results to address selected strategic challenges, including education, global health and family planning. Melinda Gates has stated, "Contraceptives are one of the greatest antipoverty innovations the world has ever seen" (Keating, 2017), further noting how deeply touched she has been by the personal stories of women she has met around the world.

In 2013, the Gates Foundation launched two complementary programs intended to help monitor annual progress toward the FP2020 goal, with a total projected investment through 2020 of more than $50 million. The Performance Monitoring and Accountability 2020 (PMA2020) program was designed to *generate data* through (at a minimum) annual, rapid-turnaround, nationally representative surveys of households and service delivery points in nine countries (now 11), using mobile phone technology for data collection. The Track20 program was designed to help support global *standardization of family planning indicators* and, in 22 countries (now 37), *draw on data from various sources (including data from PMA2020 surveys) to produce estimates of those indicators* through Bayesian modeling.

In early 2017, at roughly the midpoint between the launch of PMA2020 and Track20 and the FP2020 target date, the Gates Foundation contracted with the RAND Corporation (hereafter, RAND) to assess its investments in these programs and identify potential future directions. The Gates Foundation specified that the evaluation should address four questions:

1. How well are the PMA2020 and Track20 programs functioning now?
2. What changes to the design of PMA2020 surveys might improve their use?
3. How do stakeholders view PMA2020 and Track20 (including their perceptions of the level of data maturity in program countries)?
4. Are the two programs sustainable?

This report describes the RAND team's evaluation of the PMA2020 and Track20 programs, conducted from April to September 2017. Following this introduction, Chapter Two provides background on the two programs and brief descriptions of other relevant family planning data sources. Chapter Three describes our methods for the evaluation. Chapter Four introduces the assessment frameworks we developed, including logic models for the two programs, a data maturity framework, and a sustainability framework. Chapter Five presents stakeholder views on family planning data needs. Chapters Six through Eight focus on PMA2020, including its goals, accomplishments, and challenges (Chapter Six); the results of our statistical analyses of PMA2020 surveys (Chapter Seven); and potential changes to the PMA2020 survey, drawing on stakeholder perceptions of its statistical properties and suggestions for modifications based on both stakeholder views and our independent statistical analyses (Chapter Eight). Chapter Nine then discusses the goals, accomplishments, and challenges for Track20, and Chapter Ten discusses the intersections between PMA2020 and Track20. Chapters Eleven, Twelve, and Thirteen address data use, data maturity, and sustainability of the two programs, respectively, reflecting a variety of stakeholder perspectives. Finally, Chapter Fourteen presents the RAND team's conclusions and recommendations. A series of appendixes contains more-detailed information on PMA2020's sampling procedures, relevant family planning indicators, other family planning data sources, logic models, the ratings related to data maturity and sustainability, brief overviews of the countries included in this evaluation, and our statistical analysis methods.

CHAPTER TWO

# Background

## Overview

For decades, family planning has been among the important strategies used by governments and donors to promote economic development. Family planning programs and the data to track program progress have evolved over the years. This chapter first provides background information about the PMA2020 and Track20 programs and then describes other relevant family planning data sources—which, taken collectively, represent the family planning data landscape and, thus, the context for the present evaluation.

## PMA2020

Stemming directly from the 2012 London summit, PMA2020 was designed to support nationally representative surveys to facilitate annual progress reporting of core FP2020 indicators.

In its original proposal,[1] the Gates Institute for Population and Reproductive Health at Johns Hopkins University (hereafter "Gates Institute" or "Baltimore team") specified four objectives for the five-year project period from April 2013 through March 2018:

1. Expand country-level monitoring capacity.
2. Integrate a rapid data collection system using mobile devices.
3. Improve data monitoring to strengthen service delivery.
4. Promote the use of data to respond to family planning needs at the community level.

The proposal aimed to support surveys in nine countries at the outset. The initial five-year grant was for approximately $15 million; a supplemental proposal in December 2015 requested an additional $29 million, bringing the total funding request for the 2013–2018 grant period to more than $40 million (to address similar objectives in roughly the same number of countries). In July 2015, the Baltimore team submitted a revised set of objectives:

1. Generate survey data.
2. Build a sustainable business model.
3. Progress toward survey sustainability and financing.
4. Integrate PMA2020 into countries' monitoring and evaluation (M&E) architecture.

[1] The original and supplemental proposals were shared with us by the Gates Foundation in May 2017.

The original proposal mentioned significant connections to the Track20 program, including training and technical support for M&E officers and the expectation that Track20 would include PMA2020 survey data when developing its national estimates.

PMA2020 surveys are typically carried out twice per year for two years and then annually thereafter. Most of these surveys produce national-level estimates for the desired indicators, though some countries have carried out PMA2020 surveys at subnational levels (e.g., specific states in Nigeria and India, Kinshasa in the Democratic Republic of the Congo [DRC]). The household survey samples are based on a probabilistic two-stage stratified cluster design, with a varying number of clusters of 30–35 households each. The sampling design is very similar to that used by the DHS, which is generally considered the gold standard for family planning surveys but is only carried out every five years and covers several health topics, with family planning covered in less detail than by PMA2020. Further details of the sampling design are presented in Appendix A.

The same household sampling frame is used for the first four rounds and then is refreshed. Reportedly, the refreshed samples are drawn from neighboring enumeration areas in order to retain the trained local resident enumerators rather than from entirely new samples of the entire population of the country or subnational jurisdiction.

PMA2020 surveys collect demographic information about households; information from women ages 15–49 in those households about their reproductive history and their utilization, demand, and choice of various contraceptive methods; and information from nearby service delivery points about access to and supply of commodities and services. These surveys capture most of the FP2020 core indicators—i.e., indicators of interest to the global family planning and development community. Some countries have included additional modules addressing development initiatives beyond family planning, such as water and sanitation, primary care, maternal and newborn care, and schistosomiasis (a parasitic disease carried by freshwater snails). More information about the core FP2020 indicators, definitions of PMA2020 indicators, and a comparison of indicators across key family planning surveys can be found in Appendix B.

PMA2020 survey rounds also include the three lowest levels of health facilities serving each enumeration area where the household surveys are conducted, varying by country but corresponding to health posts, health centers, and first-level hospitals. These surveys are referred to as service delivery point surveys. Data from households and service delivery points are collected over a six-week period; initial reports are published within about six weeks; and full reports and the data sets themselves are disseminated to the public six months after survey completion.

A unique feature of PMA2020 is that the individual (household and female) surveys are carried out by resident enumerators, who are women over the age of 21 from the surveyed area or nearby. While requirements vary per country, resident enumerators hold at least a high school diploma (with some working toward degrees in public health) and are remunerated for their PMA2020 work, but they are not full-time salaried health workers. Instead, their expertise lies in familiarity with the region and with use of mobile phones. The innovative use of mobile phones—the Mobile Assisted Data and Dissemination System—is another feature of PMA2020 surveys that figured prominently in the original Gates Institute proposal and in the country surveys.

As of January 2017, at least one round of PMA2020 surveys had been carried out in ten countries, with five rounds in five of those countries (Table 2.1). The RAND team's statistical

**Table 2.1**
**PMA2020 Survey History, by Country**

| Country | Launch Date | Number of Rounds of Data Collection | National or Subnational | Number of Enumeration Areas |
|---|---|---|---|---|
| DRC | January 2013 | 5 | Kinshasa, Kongo Central | 110 |
| Ghana | October 2013 | 5 | National | 100 |
| Kenya | July 2014 | 5 | National | 120 |
| Uganda | June 2014 | 5 | National | 110 |
| Ethiopia | March 2014 | 5 | National | 221 |
| Burkina Faso | December 2014 | 4 | National | 85 |
| Nigeria | October 2014 | 3 | Kaduna, Lagos, national | 302 |
| | | Round 4 is under way | | |
| Indonesia | August 2015 | 2 | National | 372 |
| Niger | August 2015 | 3 | Niamey, national | 80 |
| India | September 2016 | 1 | Rajasthan | 147 |
| | | Round 2 is under way | | |
| Côte d'Ivoire | August 2017 | Round 1 is under way | National | 87 |

analyses of surveys from selected countries (presented in Chapter Seven) included all survey rounds for which data were available as of July 2017, which did not include all rounds of data collected.

## Track20

Track20 was designed to serve as a global resource to monitor progress toward FP2020 goals, through standardization and reporting of national family planning indicators, and to also serve as a resource to countries to help develop their capacity to collate and analyze data and facilitate consensus around indicator values to be reported for global situational awareness. While the initial Track20 proposal specified working in 22 countries, as of spring 2017, the program was active in 37 of the 69 poorest countries targeted by FP2020.

According to its original proposal to the Gates Foundation,[2] Track20 aimed to achieve five objectives:

1. Standardize key family planning indicators for global-level monitoring of progress.
2. Build country-level capacity to monitor progress in achieving family planning goals and to use the results to improve programs.

[2] The original and supplemental proposals were shared with us by the Gates Foundation in May 2017.

3. Improve family planning monitoring data at national and global levels to strengthen service delivery systems: Identify and implement specific steps and strategies to improve country service statistics and develop methods to estimate key indicators using both household and service statistics data.
4. Estimate annual family planning expenditures by national programs and donors.
5. Issue annual global reports detailing progress and lessons learned.

One area of organic evolution for Track20 (and for PMA2020) has been an expansion beyond strictly FP2020 core indicators to now encompass a broader range of program-relevant and country-specific family planning data.

Track20 works with and within governments to collate, analyze, package, and create simple messages for disseminating family planning indicator data within the specific context of each program country. Avenir Health (hereafter Avenir) brought a successful model from global human immunodeficiency virus (HIV) monitoring into the family planning sphere as a core feature of Track20: the M&E officer. The Avenir team identifies a highly qualified individual in each country to serve as the country's Track20 M&E officer. Such individuals ideally have sophisticated quantitative skills, considerable experience related to data analysis and use, and very good people skills. They must be junior enough to be willing to get "into the weeds" with data but senior enough to speak with confidence and earn the respect of the many family planning stakeholders across their country. Typically, they either are already government employees or are placed in a relevant government office (e.g., within the ministry of health [MOH] or relevant statistics office; see Table 2.2). In roughly 20 countries, the Track20 program pays part or all of these individuals' salaries. Avenir believes that this model of embedded M&E officers within a strategic government office improves country ownership, data quality, and data use.

Avenir's activities include supporting M&E officers (e.g., helping them package data in an appealing way that highlights what the data mean); supporting the annual data consensus-building workshops; partnering with the United Nations Population Division to expand statistical modeling of family planning indicator estimates; tracking country-level family planning expenditures (combining data from three sources—the Kaiser Family Foundation, the World Health Organization [WHO] Commission on Information and Accountability, and the Netherlands Interdisciplinary Demographic Institute); and documenting and disseminating family planning monitoring information and tools. Avenir contracted with the developer of the Family Planning Estimation Tool (FPET) to adapt it for Track20 use. M&E officers use FPET to statistically model annual estimates for mCPR and mCPR demand satisfied, incorporating all available data that are of sufficiently high quality, both from service statistics and from surveys (including the DHS, national and other surveys, Multiple-Indicator Cluster Surveys [MICS], and PMA2020).

In broad terms, the use of data in countries has evolved. Initially, HIV data were used mainly for global reporting purposes, but countries also had a need for these data for planning, monitoring, and donor funding proposals. Several of the Track20 M&E officers worked previously in a similar role in their country's HIV program and thus have an appreciation of the need for and practical use of data in their country. Track20 data are used to assess overall progress in the family planning sphere and inform planning on a national level. They are also used to identify subnational areas where more program attention is needed. Ideal data use would inform, at a minimum, annual adjustments in family planning program investments

**Table 2.2**
**Summary of Track20 M&E Officers in the 15 Countries RAND Evaluated**

| Country | Year of Track20 Launch | Number of M&E Officers | M&E Officer Level | Source of M&E Officer Salary | Organizational Home for M&E Officer(s) |
|---|---|---|---|---|---|
| Burkina Faso | 2016 | 1 | National | MOH: 100% | MOH |
| Côte d'Ivoire | 2014 | 2 | National | Track20: 100% | MOH |
| DRC | 2014 | 1 | National | Track20: 100% | MOH |
| Ethiopia | 2015 | 1 | National | Track20: 100% | MOH/Directorate of Planning and Programming |
| India | 2014 | 5 | National, state (Bihar, Uttar Pradesh) | Track20: 100% national<br>Track20: 100% states<br>Additional technical adviser: 100%<br>Team lead (consultant): Track20 (based on hours charged) | One team lead, two national consultants (one full time, one part time) at Ministry of Health and Family Welfare, Nirman Bhavan State (Delhi); two at state level, based at Gates Foundation–supported technical service units in Bihar and Uttar Pradesh |
| Indonesia | 2013 | 1 | National | Track20 M&E officer: 100% (until 2017); then BKKBN 100%; Track20 technical adviser (1): 20% | Universitas Gadjah Mada (two people until 2017);<br>Now BKKBN: one person<br>Avenir/Track20: one person |
| Kenya | 2014 | 1 | National | Track20: 100% | Reproductive Maternal Health Services Unit, MOH Kenya |
| Lao PDR | 2016 | 0 | National | MOH and UNFPA | MOH and UNFPA. Track20 is training and providing targeted technical assistance based on government requests. There is no salary support and no "identified" Track20 person. Instead, they are using Avenir's mechanism to answer specific questions. |
| Niger | 2015 | 1 | National | Pathfinder International | Family Planning Division, MOH |
| Nigeria | 2014 | 3 | National, state (Kaduna, Lagos), consultant | MOH: 100%<br>Track20 pays for specific activities only | National: Reproductive Health Unit and M&E Unit<br>Lagos: Reproductive Health Unit<br>Kaduna: M&E Unit<br>Consultant: National and regions<br>(Now using consultant to work with the trained M&E officers and government authorities) |
| Pakistan | 2013 | 1 | National (works with provinces) | Track20: 75%<br>Population Council: 25% | Population Council (not MOH) |
| Tanzania | 2015 | 1 | National | Track20: 75%; FHI360: 25% | FHI360 (an NGO implementing family planning programming in Tanzania) |
| Uganda | 2014 | 1 | National | Track20: 100% | Office of the Assistant Commissioner, Sexual and Reproductive Health, MOH |
| Zimbabwe | 2015 | 1 | National | Track20: 100% | Ministry of Health and Child Care, Division of Family and Child Health |

SOURCE: This table was adapted from information provided by Priya Emmart of Avenir.

NOTES: BKKBN = Badan Kependudukan dan Keluarga Berencana Nasional (Indonesian National Population and Family Planning Board); Lao PDR = Lao People's Democratic Republic.

and more frequent (e.g., quarterly) attention to address local-level problems in family planning program implementation.

The original Track20 proposal has a section on sustainability, describing the importance of government commitment to FP2020. This commitment includes placing value on their data in general. In particular, government commitment includes instituting a full-time government M&E officer to help transition from a donor program toward sustainable country data capacity and data use and, ultimately, to guide family planning program strategy and implementation.

## Family Planning Data Sources

Most of the countries included in this evaluation have access to other important family planning data sources beyond PMA2020. We introduce these alternative data sources to provide context on the data landscape in which PMA2020 is operating and striving to prove its unique value to decisionmakers, as well as to summarize the data sources from which Track20 M&E officers produce their modeled estimates. The following sections provide brief descriptions of those most commonly used by countries, and Appendix C provides a more detailed comparison of the key features of different surveys, such as year launched, sponsoring organization, frequency of data collection, and sample size.

### Demographic and Health Survey

Established in 1984 by the U.S. Agency for International Development (USAID), the DHSs are typically carried out at intervals of five or more years, usually surveying between 5,000 and 30,000 households per country (DHS Program, undated). They produce nationally representative estimates, as well as estimates that are representative at the urban and rural levels and at subnational levels (departments or states). They have been conducted in all 15 countries included in our evaluation. Like PMA2020, the DHS design is a probabilistic two-stage stratified cluster design: The first stage is selection of the enumeration areas, drawn from census files, and the second stage is selection of a sample of households from those enumeration areas. DHS interim surveys are also conducted between major rounds, with the frequency varying greatly by country. These interim surveys do not cover all measures and have smaller samples than full DHS surveys, but they are nationally representative. The time from data collection to dissemination of findings from DHS surveys is about 20 months.

The DHS has been widely considered the gold standard for demographic and health data, for a number of reasons. The multistage sampling methodology is viewed as rigorous; the repeated waves of data collection over decades, largely using a standardized core questionnaire (i.e., serial cross-sectional surveys drawn from the same underlying populations and asking largely the same questions) allow for analysis of trends over time within countries and comparisons across countries; and the data collection procedures have been commended by experts in the field (Corsi et al., 2012). The large sample size also sets it apart from many other surveys.

Another strength of the DHS is the transparent and convenient release of the survey data, which are easily available for download and analysis from the DHS website and from websites such as StatCompiler (DHS Program STATcompiler, undated), which facilitates the creation of tables and graphs based on surveys from more than 90 countries, and Integrated Public Use Microdata Series—Demographic and Health Surveys (IPUMS-DHS, undated), which allows downloads of standardized DHS data from 96 surveys in 21 countries.

However, the DHS's notable weakness, which will be discussed in further detail in Chapter Six, is that it is generally conducted every five years, with an inconsistent data collection schedule across countries.

### Multiple-Indicator Cluster Surveys

Another program that collects data on family planning, among other topics, is MICS, funded by the United Nations Children's Fund (UNICEF, 2015). These surveys have been carried out in more than 100 countries since 1995, focusing on indicators related to the well-being of women and children, including child mortality, nutrition, literacy, water and sanitation, and family planning. National governments implement the survey with technical assistance and training by UNICEF. In 2009, the frequency of collection was increased from every five to every three years. While each round is different, surveys are considered comparable across years. Like the DHS, MICS is nationally representative. In some cases, MICS has been combined with another survey, such as the DHS or the Pan-Arab Project for Family Health.

In addition to child and family health indicators, MICS focused on providing data for the Millennium Development Goals (through 2015) and, now, the Sustainable Development Goals. The first MICS, in the mid-1990s, surveyed 60 countries; Round 2 was in the late 1990s and early 2000s; Round 3, conducted from 2005 to 2007, was in 49 countries; Round 4 was conducted from 2009 to 2012, with 55 countries; and Round 5 was conducted from 2013 to 2015 (Plowman and Fotso, 2013). The time to dissemination of findings from MICS accelerated between Rounds 3 and 4 and now averages 20 months. Round 6 of MICS commenced in October 2016.

### Service Availability and Readiness Assessment

The Service Availability and Readiness Assessment (SARA) is a WHO-designed and -led survey initiated in 2004 as Service Availability Mapping and revamped in 2011. It measures facility-level inputs and outputs of the health system, which can be used to measure progress in health system strengthening. The aim is to provide information "about whether or not a facility meets the required conditions to support provision of basic or specific services with a consistent level of quality and quantity" (WHO, 2017). The three primary categories of measurement are service availability (health infrastructure, personnel, and service utilization), general service readiness (the capacity of health facilities to provide services), and service-specific readiness (the ability of a facility to provide specific services, such as diagnostic capacity). SARA measures whether health facilities provide family planning services, the types of contraception offered, and the availability of training materials for health workers. WHO partners with USAID and country partners to scale up SARA implementation. SARA data are currently available in 13 countries: 11 in sub-Saharan Africa, one in Europe, and one in Central America. Of those countries, six were included in our evaluation: Burkina Faso, DRC, Ghana, Kenya, Tanzania, and Uganda.

### District Health Information Software

District Health Information Software (DHIS) is an open-source tool used to collect, report, analyze, and disseminate information from health programs (District Health Information Software, version 2 [DHIS2], undated[a]). Developed by the Health Information Systems Program (HISP) at the University of Oslo, it is used, country-wide or for smaller regions, in almost 50 countries. It is the primary national health management information system (HMIS) in

30 countries and is used by many NGOs for data collection (DHIS2, undated[b]). Development is coordinated by HISP and is supported by the Norwegian Agency for Development Cooperation; the U.S. President's Emergency Plan for AIDS Relief; and the Global Fund to Fight AIDS, Tuberculosis and Malaria. Data use components of DHIS2 (released in 2008) include geographic information system mapping, charts, pivot tables, communications features, and a platform to generate mobile data collection forms. With support from USAID and NGOs, several countries have transitioned commodity reporting to the DHIS2 platform (Otieno and Arunga, 2014). While many countries use the DHIS2 platform, others use other HMIS tools to collect family planning information. DHIS2 data are only available in a subset of the countries included in this evaluation (Burkina Faso, Ghana, India [the Bihar, Orissa, Maharashtra, Kerala, Punjab, Haryana, and Himachal Pradesh states], Kenya, Nigeria, and Uganda). We note this data source but do not use it for the validation analysis (Chapter Seven) because of the varied ways in which these data are collected.

CHAPTER THREE

# Methods

## Overview

Following review and approval by the RAND Human Subjects Protection Committee, we undertook the evaluation in steps, including document review, development of assessment frameworks and interview protocols, and interviews with stakeholders based in the United States and program countries. During the first month of the evaluation (April 2017), we reviewed documentation on the PMA2020 and Track20 programs and relevant literature related to the three assessment frameworks that we developed for this evaluation. To learn more about the two programs, we spoke with staff members from the Gates Foundation and the two grantee organizations. This preparatory work set the foundation for the interviews and statistical analysis of PMA2020 that served as the core data inputs for our evaluation.

The following three sections describe the methods for developing those assessment frameworks; for the planning, implementation, and qualitative analysis of stakeholder interviews; and for the quantitative statistical analysis of PMA2020.

## Assessment Frameworks

We developed and applied three assessment frameworks as part of this evaluation. The following sections describe the development of our logic models, data maturity framework, and sustainability framework.

### Logic Models

A **logic model** establishes a linear relationship between activities and impacts and the intermediate products and results between them (Bullen, 2013). It shows the trajectory of how *inputs* (resources) support program *activities* that in turn produce *outputs* (tangible, short-term products, services, or results) (O'Mahony et al., forthcoming), which result in *intermediate outcomes* (what is achieved using outputs) and, eventually, *long-term impacts* (desired long-term effects) (Gertler et al., 2011; Haggard and Burnett, 2006). We developed notional logic models to depict these relationships for PMA2020 and Track20 based on review of documents (e.g., objectives and activities specified in written proposals, materials from program websites) and consultations with the implementing organizations; the Gates Institute, which included a half-day site visit; and Avenir. These served as a foundation for several of the questions in our interview protocols. Drawing from stakeholder interviews, we created an empiric logic model for each program, indicating observed inputs, activities, outputs, outcomes, and trajectory toward

program goals. These are presented in Chapters Six and Nine. In the final chapter of this report, we again use logic models to depict proposed future directions for Gates Foundation family planning programming broadly and the two programs specifically. More information about logic models is found in Appendix D.

### Data Maturity Framework

Information and big data have become increasingly important inputs in a wide range of business and development operations, which has led to the development of analytic or **data maturity models** designed to evaluate and manage continuous improvement of data systems (Spruit and Pietzka, 2015). Data maturity refers to the extent to which high-quality data are collected, well managed, well governed, rigorously analyzed, shared, communicated, and, ultimately, used. It should be considered an evolutionary process. We performed a targeted literature review of existing data maturity models that could be relevant to the evaluation of the PMA2020 and Track20 programs, pulling from both health and other sectors. Models ranged in the detail of methodology that was publicly accessible: The University of Chicago has published its questionnaire and a model scorecard, while the world-renowned Capability Maturity Model Integration (CMMI) developed by Carnegie Mellon University is proprietary, with assessment services available for a fee. Key components of the models we reviewed are summarized in Appendix E.

Because none of the models we identified was sufficiently well aligned with the goals and organizational structures of PMA2020 and Track20, we drew from different models to develop a tailored composite data maturity framework that captures the maturity of organizational readiness, data systems, and data use for decisionmaking. We adapted the domains and areas covered by existing models to portray the particularities of the data generation and management processes of PMA2020 and Track20. We incorporated this framework into the in-country stakeholder interview protocols and applied it as a proof-of-concept tool to assess and help understand the data maturity of PMA2020 and Track20 in program countries. This first field application did not incorporate differential weights for the various data maturity elements. We tailored the questions to the particular stakeholder group (i.e., we only asked stakeholders who were sufficiently familiar with PMA2020 or Track20 to comment on and rate the program in particular areas). The details of RAND's data maturity framework are presented in Chapter Four, and the results from its application in program countries are presented in Chapter Twelve.

### Sustainability Framework

To develop a sustainability framework to guide our evaluation, we performed a focused literature review and then compiled a list of factors that could enable sustainable data systems and data use and could be measured to track progress toward this end. We organized these sustainability-enabling factors into four broad thematic categories—financial sustainability, technical sustainability, operational sustainability, and data culture. We then paired each enabling factor with at least one measure or indicator that could capture whether, or to what degree, each factor is being realized in a given country. While most of our measurements reflect self-assessed perceptions of progress or achievement, for some factors the proposed measures are directly quantifiable (e.g., existence of a multistakeholder coordination body, number of dedicated data system personnel). We selected factors and measures that we expected would be the most essential to assess in this evaluation and asked about them through both open-ended and structured questions during

our interviews with stakeholders in the 15 program countries. As with our data maturity framework, this initial field application of our proof-of-concept framework did not incorporate differential weights for the various sustainability-enabling factors. The details of our sustainability framework are presented in Chapter Four, and results from both the interviews and structured sustainability ratings are presented in Chapter Thirteen. More information about the sustainability enablers we considered in this evaluation is found in Appendix F.

## Stakeholder Interviews and Qualitative Analyses

The RAND evaluation included interviews with four groups of stakeholders: (1) U.S.-based stakeholders from the Gates Foundation, (2) U.S.-based stakeholders from the two grantee organizations (the Gates Institute for PMA2020 and Avenir for Track20), (3) U.S.-based external stakeholders from the PMA2020 External Consultative Group and other subject-matter experts in family planning and statistics, and (4) stakeholders in selected program countries (e.g., PMA2020 and Track20 program staff, government officials, bilateral and multilateral partners, NGOs).

### Interviews with U.S.-Based Stakeholders

We drew from the preparatory work to develop an interview protocol for U.S.-based stakeholders who were suggested by the Gates Foundation project officer. These included staff from the Gates Foundation and the two grantee organizations; members of the PMA2020 External Consultative Group who are based in such organizations as USAID, FP2020, UNFPA, and academic institutions; other family planning experts; and statistical experts. The interview protocols included questions that were similar to those we used with the stakeholders in program countries—the goals, achievements, and challenges related to PMA2020 and Track20. Some questions were specifically targeted to the Gates Foundation, PMA2020, or Track20 staff, but most were intended for all of the U.S.-based stakeholders. We conducted those consultations in person and by phone with 39 individuals from May through mid-July 2017 (Table 3.1). RAND interviewers took nearly verbatim notes, and one researcher manually extracted comments from all of the interviews by topic (e.g., goals, achievements, challenges, etc.). Using a deductive approach, two researchers independently identified key emergent themes for each topic and reached consensus, and these themes form the basis for results reported in Chapters Five, Six, and Eight through Thirteen.

### Interviews in PMA2020 and Track20 Countries

Together with colleagues at the Gates Foundation and the two grantee institutions, we finalized the 15 countries to be included in the evaluation. They include all 11 PMA2020 countries (ten of which also participate in Track20 and one of which only participates in PMA2020) and four that participate in Track20 only (Table 3.2).

Appendix G provides brief contextual information on the 15 countries included in this evaluation of PMA2020 and Track20. It draws from country statistics, information from the Gates Institute and Avenir, and observations from the RAND team member who conducted the interviews in each country. The countries vary greatly in population size, but all are classified as low income or lower middle income by the World Bank, and most rank quite low on the Human Development Index. Several of them have a rich history of relevant family planning

**Table 3.1**
**U.S.-Based Interviewees**

| Interviewee Affiliation | Number of Interviewees |
|---|---|
| Bill & Melinda Gates Foundation | 8 |
| Gates Institute (PMA2020) | 11 |
| Avenir (Track20) | 5 |
| PMA2020 External Consultative Group | 7[a] |
| Other | 8 |
| Total | 39 |

NOTE: One of these individuals was located in Canada but, for the purposes of this report, is included in the U.S.-based stakeholders category.

[a] An Avenir interviewee is an eighth External Consultative Group member.

**Table 3.2**
**Countries Included in the RAND Evaluation**

| PMA2020 and Track20 | Track20 Only | PMA2020 Only |
|---|---|---|
| Burkina Faso | Lao PDR | Ghana |
| Côte d'Ivoire | Pakistan | |
| DRC | Tanzania | |
| Ethiopia | Zimbabwe | |
| India | | |
| Indonesia | | |
| Kenya | | |
| Niger | | |
| Nigeria | | |
| Uganda | | |

surveys and data sources. Each RAND team member who completed interviews in a country summarized key features of the program(s) and his or her overall sense of notable strengths and opportunities for improvement, based on the collective feedback of interviewees. While their impressions are subjective, they provide an important country-specific contextual foundation for understanding the views of the in-country stakeholders that are reflected throughout this report.

Our review of program documentation and early discussions with Gates Foundation staff and PMA2020 and Track20 leaders informed the development of the interview protocols for program countries. Most questions were open-ended, but the protocol also asked stakeholders to provide numeric ratings for elements of the data maturity and sustainability frameworks

described briefly above and in more detail in Chapter Four. The frameworks served as proofs of concept for potential tools that might be used for periodic monitoring in the future.

### *Interviewee Selection and Recruitment*

During the second and third months of the evaluation (May and June 2017), we carried out interviews on site in 14 countries and by phone for one country (Pakistan). Both of the U.S.-based grantees facilitated contacts with their respective in-country staff to help the RAND team members arrange interviews with relevant stakeholders. We sought to conduct approximately eight to 12 interviews in each country with (1) relevant staff from PMA2020 (the principal investigator, data manager, information technology (IT) specialist, field supervisor, and resident enumerator); (2) Track20 staff (the M&E officer and the M&E officer's supervisor); and (3) government officials (family planning program managers or other MOH officials and statistics office personnel), bilateral donors (e.g., USAID, United Kingdom Department for International Development [DfID]), multilateral agencies (e.g., UNFPA), and NGOs working in family planning. In addition to purposively sampling from the aforementioned stakeholder groups, we performed snowball sampling in order to contact a broad range of perspectives on PMA2020 and Track20. Table 3.3 shows the number of interviews and interviewees by country and stakeholder type, as well as the number of completed data maturity and sustainability ratings.

Our eight team members who conducted in-country interviews are highly skilled in conducting semistructured interviews among international stakeholders and have extensive training in fields including health policy, global health, epidemiology, anthropology, public health program evaluation, and international relations. All team members had previous research experience, and three are native to three program countries included in our evaluation.

We developed semistructured interview protocols for each of the three broad stakeholder groups noted above: PMA2020 staff; Track20 staff; and all others, including government officials and bilateral, multilateral, and nongovernmental partners. The semistructured interview protocols included both stakeholder- and program-specific questions, as well as cross-cutting themes relating to the importance of family planning data; additionally, the semistructured format allowed stakeholders to elaborate on topics relevant to their work and experience. The Gates Foundation reviewed and approved the interview protocols. Interviews were designed to last approximately one hour.

We contacted representatives from each of the three stakeholder groups via e-mail and arranged an in-person interview at the convenience of the interview participants. When an in-person interview could not be scheduled or when travel to the country was not possible (as was the case with Pakistan), we conducted interviews by phone. Interview participants were notified of our confidentiality procedures and their rights with regard to participation in this evaluation and provided consent for the interview and recording prior to the start of the interview. Interviews were recorded, except for a small number for which recording was declined or a recording via telephone was not possible. RAND interviewers also took detailed written notes to capture interview data and any relevant observations. Interviews were conducted in English, except in the following countries, where interviews were conducted by a native or fluent speaker on the RAND team: Pakistan (Urdu); Burkina Faso, Côte d'Ivoire, DRC, and Niger (French); and Ethiopia (Amharic).

Interviews were recorded and securely uploaded to a third-party transcription service. Due to limited access to non-English transcriptionists, only those conducted in English were

**Table 3.3**
**Interviewees in PMA2020 and Track20 Program Countries**

| | Number of Individuals Interviewed, by Function | | | | | | | | Data Maturity Ratings | | Sustainability Ratings | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Country | PMA2020 Staff | Track20 Staff | Government | Bilateral | Multilateral | NGO | Total Interviewees | Total Interviews | # | % of Individuals | # | % of Individuals |
| Burkina Faso | 1 | 4 | 1 | | | 2 | 8 | 6 | 2 | 25% | 2 | 25% |
| Côte d'Ivoire | 3[a] | 4[a] | | | 2 | 2 | 10 | 9 | 7 | 70% | 7 | 70% |
| DRC | 2 | 2 | 1 | | | 3 | 8 | 8 | 7 | 88% | 6 | 75% |
| Ethiopia | 8 | 1 | 1 | 1 | 1 | | 12 | 11 | 11 | 92% | 11 | 92% |
| Ghana | 6 | (n/a) | 4 | 1 | | 1 | 12 | 11 | 11 | 92% | 11 | 92% |
| India | 4 | 5 | 7 | 2 | 7 | 4 | 29 | 26 | 20 | 69% | 19 | 66% |
| Indonesia | 7 | 4 | 7 | | 3 | | 21 | 14 | 8 | 38% | 7 | 33% |
| Kenya | 1 | 3 | 2 | 1 | 1 | 4 | 12 | 12 | 9 | 75% | 8 | 67% |
| Lao PDR | (n/a) | 3 | 3 | | 4 | 3 | 13 | 11 | 9 | 69% | 7 | 54% |
| Niger | 4 | 2 | 2 | | | 4 | 12 | 9 | 6 | 50% | 6 | 50% |
| Nigeria | 7 | 7 | 5 | | 1 | 7 | 27 | 24 | 22 | 81% | 22 | 81% |
| Pakistan | (n/a) | 1 | 6 | | 1 | 2 | 10 | 10 | 10 | 100% | 10 | 100% |
| Tanzania | (n/a) | 2 | 3 | 2 | 1 | 6 | 14 | 13 | 8 | 57% | 8 | 57% |
| Uganda | 7 | 2 | 4 | 2 | 2 | 9 | 26 | 17 | 10 | 38% | 5 | 19% |
| Zimbabwe | (n/a) | 2 | | 1 | 1 | 7 | 11 | 9 | 7 | 64% | 7 | 64% |
| Total | 50 | 42 | 46 | 10 | 24 | 54 | 225 | 190 | 147 | 65% | 136 | 60% |

NOTE: n/a = not applicable.

[a] One of the Track20 supervisors is also the PMA2020 principal investigator.

transcribed verbatim. Otherwise, detailed notes (including review of audio recording by the interviewer) were used to document interviews conducted in French, Amharic, and Urdu. RAND interviewers used their detailed written notes to review the English-language transcriptions for clarity and accuracy. All interviews conducted among PMA2020- and Track20-affiliated staff (e.g., PMA2020 principal investigator, Track20 M&E officer), as well as key government officials and members of bilateral organizations and NGOs, were transcribed. Any additional interviews were transcribed and analyzed if they were deemed to contribute highly relevant information relating to either program or to the state of data use in the program countries, at the discretion of the RAND interviewer.

#### *Observations at National Family Planning Data Consensus Meetings*

In addition to carrying out interviews to ask about program activities and achievements, we also wanted to see the programs in action. When possible, the RAND team member attended the country's national family planning data consensus meeting convened by Track20 staff, which allowed for direct observations of discussions, debates, challenges, opportunities, and progress surrounding the generation and use of family planning data. We were able to attend these meetings in five countries (Burkina Faso, Côte d'Ivoire, Tanzania, Uganda, and Zimbabwe). In countries in which the visit did not coincide with the consensus meeting, the team followed up with country stakeholders to elicit perspectives on the aforementioned topics surrounding family planning data. These observations presented a unique opportunity to observe rich discussions among key stakeholders (including local UNFPA, USAID, and DfID representatives; government officials; local academics; and others) and to generate key follow-up probes for subsequent interviews.

#### *Analyses of the Program Country Interviews*

We conducted a total of 190 interviews with 225 in-country stakeholders (Table 3.3; some interviews consisted of more than one interviewee). Based on contact with key U.S.-based and in-country program staff, we anticipated interviewing eight to 12 stakeholders per country. However, some countries had more robust and longer-term program implementation than other countries; thus, we adjusted the target number of interviewees according to the number of key stakeholders affiliated with each program and with family planning programming in the country—e.g., Lao PDR had a limited number of Track20-only stakeholders in one region, while Indonesia had both PMA2020 and Track20 stakeholders in multiple regions.

Of the 190 total interviews, 159 were deemed to be of sufficient recording and transcription quality to be coded for analysis. Thirty-one interviews were not transcribed because a recording was not possible or was of insufficient audio quality or because the interviewer determined that the interview did not contribute novel data to the evaluation (e.g., if two key stakeholders from NGOs shared very similar perspectives, only one interview was transcribed). In seven countries, all interviews were transcribed (either directly from the recording if in English or nearly verbatim by the interviewer in the non-English language of the interview and translated by the interviewer, resulting in detailed notes). In the remaining eight countries, the individual interviewers were considered the expert on the data collected, and they determined the selection of interviews to be transcribed for analysis.

All transcriptions and detailed interview notes were uploaded into Dedoose (SocioCultural Research Consultants, 2017), a cloud-based qualitative analysis software program that facilitates team-based coding and subsequent data analysis. We employed an inductive and

deductive approach to the development of a codebook, which accounted for key research questions covered within the interview protocols, as well as novel topics that emerged from the interviews. The codebook denoted the domains of interest to this evaluation (e.g., barriers to data use, strengths of the Track20 M&E officer, co-financing of the programs), and themes provided substantive patterns and examples within those domains. The project team reviewed the codebook and specified the proper application of codes. Five team members experienced in qualitative data analysis then coded the interviews. They established inter-coder reliability (i.e., consistency and consensus in the application of codes) within the Dedoose platform. We established inter-coder reliability (Cohen's kappa > 0.80[1]) for five exemplary codes out of the total codebook of 120 codes. While this proportion is lower than standard practice, we used the coded themes to extract key excerpts for a qualitative analysis of the interviews rather than using the codes in a quantitative analysis. The coding team met weekly over the course of June and July 2017 to resolve any discrepancies in the application of codes.

Upon completion of coding, the larger project team reviewed the coded excerpts for key themes. Themes were identified through well-established techniques, including repetition (e.g., if a theme was expressed more than three times) and comparing both within-country and across-country interviews for similarities and differences (Ryan and Bernard, 2003). In addition, we searched for themes, (e.g., facilitators of and barriers to data maturity and sustainability) that were common to both programs. We also focused on themes that were specific to particular types of stakeholders (e.g., PMA2020- or Track20-affiliated staff, members of bilateral organizations) or to particular countries. Together, these techniques gathered the range of key stakeholder perspectives and provided in-depth, contextual data to complement the quantitative findings from the data maturity and sustainability ratings. The results of these analyses are described in Chapters Five, Six, and Eight through Thirteen.

## Statistical Analyses of PMA2020

PMA2020 surveys are designed to provide nationally representative estimates of key family planning indicators. As the program has evolved, questions about the survey methodology have emerged. The Gates Foundation asked RAND to assess the statistical properties of the PMA2020 household surveys and consider alternative sampling schemes. Data for these analyses came from two sources. The RAND team downloaded PMA2020 and DHS data from the respective websites (http://www.pma2020.org and https://dhsprogram.com/data/available-datasets.cfm). The data were updated through July 12, 2017. We focused on four aspects of PMA2020:

- representativeness
- detectable margin of error
- sampling design
- frequency of data collection.

[1] Cohen's kappa is a robust, acceptable measure of inter-coder agreement. Cohen's kappa approaches 1 as agreement among coders increases (Haney et al., 1998). Cohen's kappa is calculated as $K = (Pa - Pc)/(1 - Pc)$, where $Pa$ is the proportion of units on which the coders agree and $Pc$ is the proportion of units for which agreement is expected to occur by chance.

Below, we explain how we assessed each of these topics. We present the findings from our analyses in Chapter Seven.

### Representativeness

In order to assess the **representativeness** of PMA2020, both for a given round and over time, we conducted two sets of analyses. Assessing representativeness means that we identify the extent to which the sample for a given PMA2020 round is similar, in terms of its demographic and economic characteristics, to a benchmark survey. The first analysis compares the distribution of the characteristics of the PMA2020 survey sample to the best available benchmark: the DHS, which is the gold standard for national surveys estimating family planning indicators. Because multiple rounds of the PMA2020 have been conducted in many countries, the second set of analyses compares the characteristics of the sample from one round to the next to determine whether systematic differences in the sampled population are observed over time. We note that changes in demographic measures between different rounds may be a result of secular trends and do not necessarily signify nonrepresentativeness.

#### *Comparison of PMA2020 to the DHS to Assess Sample Representativeness*

We compared PMA2020 to DHS survey data in five countries—Ghana, Ethiopia, DRC, Nigeria, and Kenya—that vary in population size and have multiple rounds of data available. Using the PMA2020 survey round that is closest in time to the most recent DHS data available (and, in the case of Nigeria, limiting our analysis of DHS data to the two states in which PMA2020 was conducted for Rounds 1 and 2, Lagos and Kaduna, and in DRC, limiting our analysis to Kinshasa), we compared the demographic and economic characteristics in the samples, using a weighted two-sample t-test for continuous indicators and a chi-squared test for categorical predictors. Because the different surveys are carried out by different entities and are weighted to largely different population sizes, all the weights used in the analyses were rescaled to be comparable.

#### *Representativeness of the PMA2020 Sample from Round to Round*

In order to assess the representativeness of the PMA2020 sample for generating estimates that can be compared over time, we conducted two-sample comparisons within PMA2020 data, comparing sample characteristics during different rounds in each country. In this case, the estimates for each round of data are calculated and then the differences from each round to the following round are reported, including the associated $p$ value for each comparison.

#### *Representativeness at More Granular Geographic Levels*

We conducted additional analyses comparing urban and rural areas, where possible, to understand possible sources of difference in characteristics between PMA2020 and the DHS or between different rounds in PMA2020.

### Detectable Margin of Error—Precision for a Specific Sample Size

For the sampling of the PMA2020 survey, we conducted statistical analyses to assess whether a sufficiently small (i.e., less than the 3-percentage-point target set by the program) **margin of error** could be observed with the country sample size, which, in turn, would increase the accuracy and validity of PMA2020. This analysis provides information on the **accuracy** of the different indicators, and, for those that are not the primary interest of PMA2020, it allows policymakers to assess how reliable the estimates for other indicators can be. In these analyses,

the 95-percent confidence interval was reported, defined as the critical value (1.96 for large sample sizes) multiplied by the standard error of the indicator.

The expected margin of error can be different from country to country and from indicator to indicator. The indicator that was originally used to determine sample size was the estimated margin of error for mCPR, which is FP2020's main outcome of interest (AvenirHealth, 2017). In general, the samples are powered, for each round, to detect an annual change in mCPR of less than an absolute 3-percentage-point margin of error at the national level, but some countries and regions used a larger detectable margin of error to calculate their sample size, indicating either that they expected rapid increases in this indicator or would accept smaller increases with less precision (Table 3.4) (PMA2020, 2014).

Because policymakers are likely interested in other PMA2020 family planning indicators beyond mCPR, RAND researchers estimated the **observed margin of error from the PMA2020 for different rounds and for different indicators that would be particularly useful to various decisionmakers,** with the goal of allowing informal comparisons of margin of error across rounds, countries, and indicators.

### Sampling Design

For each indicator of interest, the sampling weight impacts the precision of the estimate used for inference, which can increase the variance of the indicator. In addition to survey weighting (in many cases, post-stratification weights are used to adjust for nonresponse), the precision of each statistic can also be impacted by the clustered nature of the sample design, especially if there is a significant amount of correlation within the sampling units (referred to as *enumeration areas*). The amount of the clustering impact is specific to each indicator and depends on the variance of the indicator. Clustering results in a **design effect** greater than one, which has the impact of reducing the effective sample size in the data. In other words, *design effect* is the factor by which the variance of an indicator is deflated (and, therefore, the precision is overestimated) due to nonindependence within enumeration areas. If observations within a cluster are very similar, additional observations in that cluster might not necessarily add more information (or precision) to the estimates. Only the effective sample size of a survey, which is the actual sample size divided by design effect, will impact the statistical precision of the different statistics of interest.

Using a hierarchical linear model (even in the case of categorical variables, we used linear probability models), we estimated the intra-class correlation (ICC) for the different indicators of interest based on the sampling stratification used in the specific country design. We report both the ICC and the resulting design effect for the different indicators across countries. These estimates allowed us to infer the amount of similarity within enumeration areas. The calculated design effect for PMA2020 indictors, by round and by country, can be compared to the design effects observed in other data sets, such as the DHS. These calculations indicated whether some indicators are more correlated than others.

PMA2020 surveys are based on a stratified sample with the goal of recruiting 35 respondents per cluster (enumeration area), with a varying number of enumeration areas per country. Such design can be suboptimal if the variance of indicators of interest differs across different enumeration areas. For example, if an enumeration area has very low variability in the use of modern contraceptive methods, it will be best to select very few respondents in such an enumeration area, while for enumeration areas with high variability in the indicator of interest, more respondents will need to be selected to produce the same level of accuracy within

**Table 3.4**
**Level at Which Data Are Representative and Margin of Error, by Country, Used in Gates Institute Sample Size Calculations**

| Country | Level at Which Data Are Representative | Margin of Error Used for Original Sample Size Calculation |
|---|---|---|
| Burkina Faso | National | 2% |
| | Urban/rural | <5% |
| DRC | Kinshasa | <2% |
| | Kongo Central | <2% |
| Ethiopia | National | <2% |
| | Urban/rural | <3% |
| | 5 regions[a] | 5% |
| Ghana | National | <2% |
| | Urban/rural | <3% |
| India (Rajasthan) | State | 2% |
| | Urban/rural | 3% |
| Indonesia | National | <2% |
| | Urban/rural | <3% |
| | South Sulawesi | <3% |
| | Makassar district | 5% |
| Uganda | National | 2% |
| | Urban/rural | <3% |
| Kenya | National | <3% |
| | Urban/rural | 3% |
| | 9 counties[b] | 5% |
| Niger | National | <2% |
| | Niamey only | 3% |
| | Urban/rural | <3% |
| Nigeria | National | <2% |
| | Urban/rural | <2% |
| | 7 states[c] | <2–3% |
| Uganda | National | 2% |
| | Urban/rural | <3% |

SOURCE: PMA2020, 2017a.

[a] Addis Ababa; Amhara; Oromiya; Tigray; and Southern Nations, Nationalities, and People's Region.

[b] Bungoma, Kericho, Kiambu, Kilifi, Kitui, Nairobi, Nandi, Nyamira, and Siaya.

[c] Anambra, Kaduna, Kano, Lagos, Nasarawa, Rivers, and Taraba.

that enumeration area. In order to assess the efficiency in the design used in PMA2020 when compared to an optimal design that takes into account difference in variability between enumeration areas, we estimated the expected variance of an indicator if an optimal design were used and compared it to the variance under the current design.

Using the selected enumeration areas, and for each indicator, assuming that the estimated variance of each indicator is a good approximation of the population variance of each indicator in the enumeration areas, the RAND team used the **optimal design** scheme proposed by Cochran and William to choose the best probability design that allows for the best precision of each indicator (Cochran and William, 1977). Because precision is the inverse of variance (the smaller the variance, the larger the precision of the estimate), this sampling strategy chooses the number of survey respondents in each enumeration area that minimizes the variance of the population estimate of the indicator within the country. Practically, this is a strategy to increase efficiency because fewer respondents can be interviewed in enumeration areas with low variance without sacrificing precision.

For example, for traditional contraceptive methods, such as withdrawal and fertility-awareness approaches, which have little variation in prevalence among some enumeration areas, this optimization will require only taking a small number of observations in such areas. However, for other enumeration areas with an indicator showing large variation, a larger number of observations will be required to achieve the same level of accuracy. The optimal sample size for one indicator (e.g., mCPR) will not necessarily be optimal for another indicator (e.g., traditional contraceptive prevalence rate). In this evaluation, RAND researchers assessed the indicators for which variance can be minimized when using optimal design. The actual survey sample size can then be determined based on the indicator of main interest (such as mCPR) or the indicator that requires the largest sample size among different indicators of interest.

Additional detail on optimal probability design for a specific indicator and methods for calculating the sample size within enumeration areas that will result in the indicator's optimal sample are found in Appendix H.

### Frequency of Data Collection

PMA2020 survey data are collected in multiple rounds. The first four rounds are conducted roughly every six months, and all indicators are assessed at each round of data collection. To assess whether the statistical strength of PMA2020 is enhanced by such frequent observations, we conducted a **round-to-round comparison** of selected indicators and examined trends over time. Using the t-test for continuous indicators and the chi-squared test for categorical indicators, we assessed the difference between one round and all subsequent rounds for specific indicators. For simplicity, we reported the indicator estimates for round 1, the differences from round to round as available (for six-month differences, one-year differences, one-and-a-half-year differences, and two-year differences), and the $p$ value for the test statistics comparing rounds. We make inferences on whether estimates from one round to the next are different enough to warrant data collection every six months or every year. We compare across indicators and across countries. This helps to identify the optimal frequency of data collection and provides insight into whether the same data could be captured more efficiently (i.e., with less frequent surveys) and what the impact would be of less frequent data collection.

CHAPTER FOUR

# Assessment Frameworks

## Overview

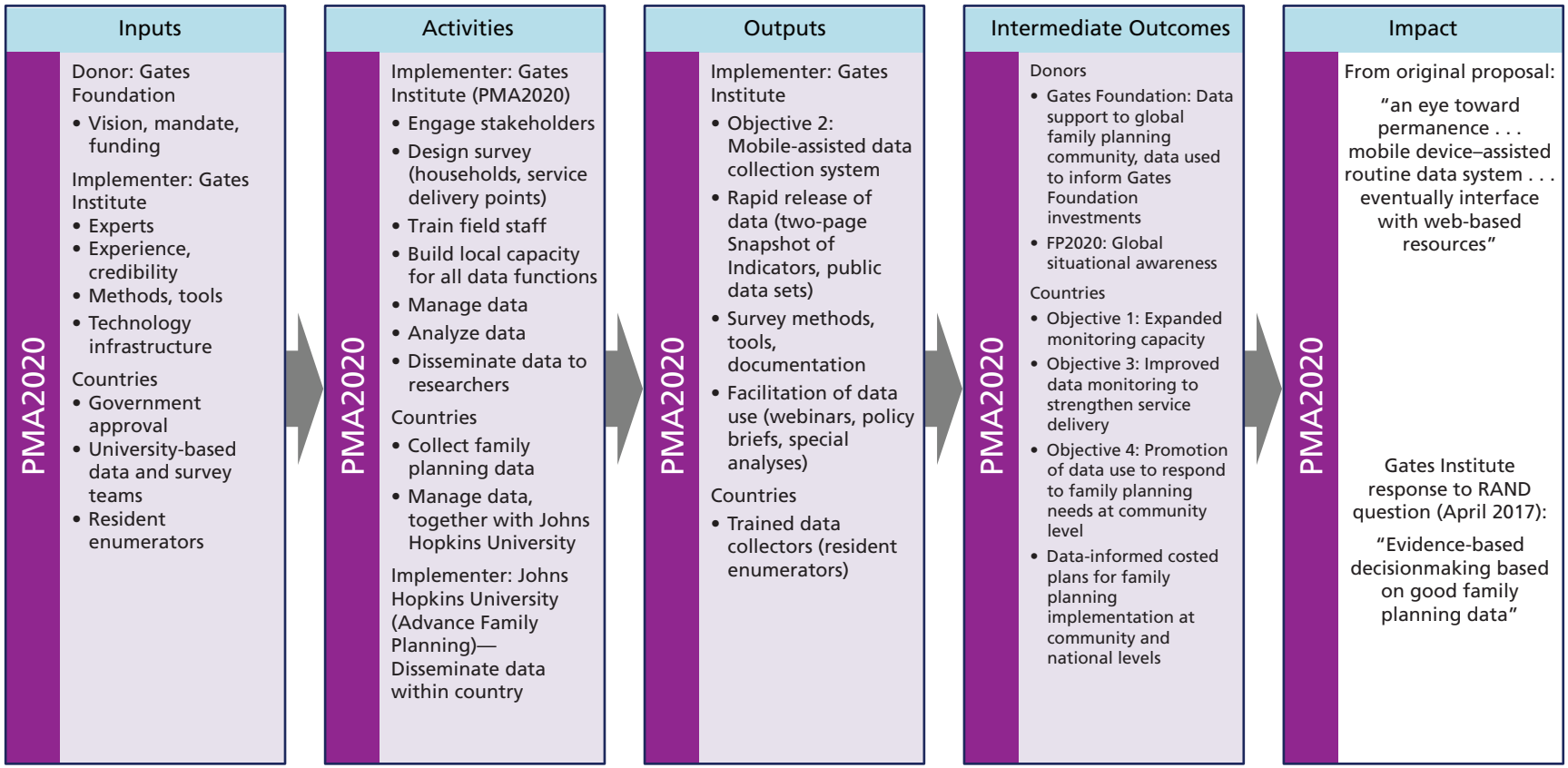
To guide our evaluation of PMA2020 and Track20, we developed three guiding frameworks: a logic model for each program, a data maturity model, and a sustainability framework. The following sections provide more detail on each.

## PMA2020 and Track20 Logic Models

Drawing from program documentation and initial discussions with the PMA2020 and Track20 grantee organizations, we developed an initial logic model for each of the programs (Figures 4.1 and 4.2). We used these models to guide the development of the stakeholder interview protocols and subsequently modified them to reflect empiric logic models drawn from observations from U.S. and country stakeholders (these are presented in Chapters Six and Nine). We also depict proposed future directions for the Gates Foundation, PMA2020, and Track20 in logic models presented in Chapter Fourteen. For more detail on the theoretical basis for logic models, see Appendix D.

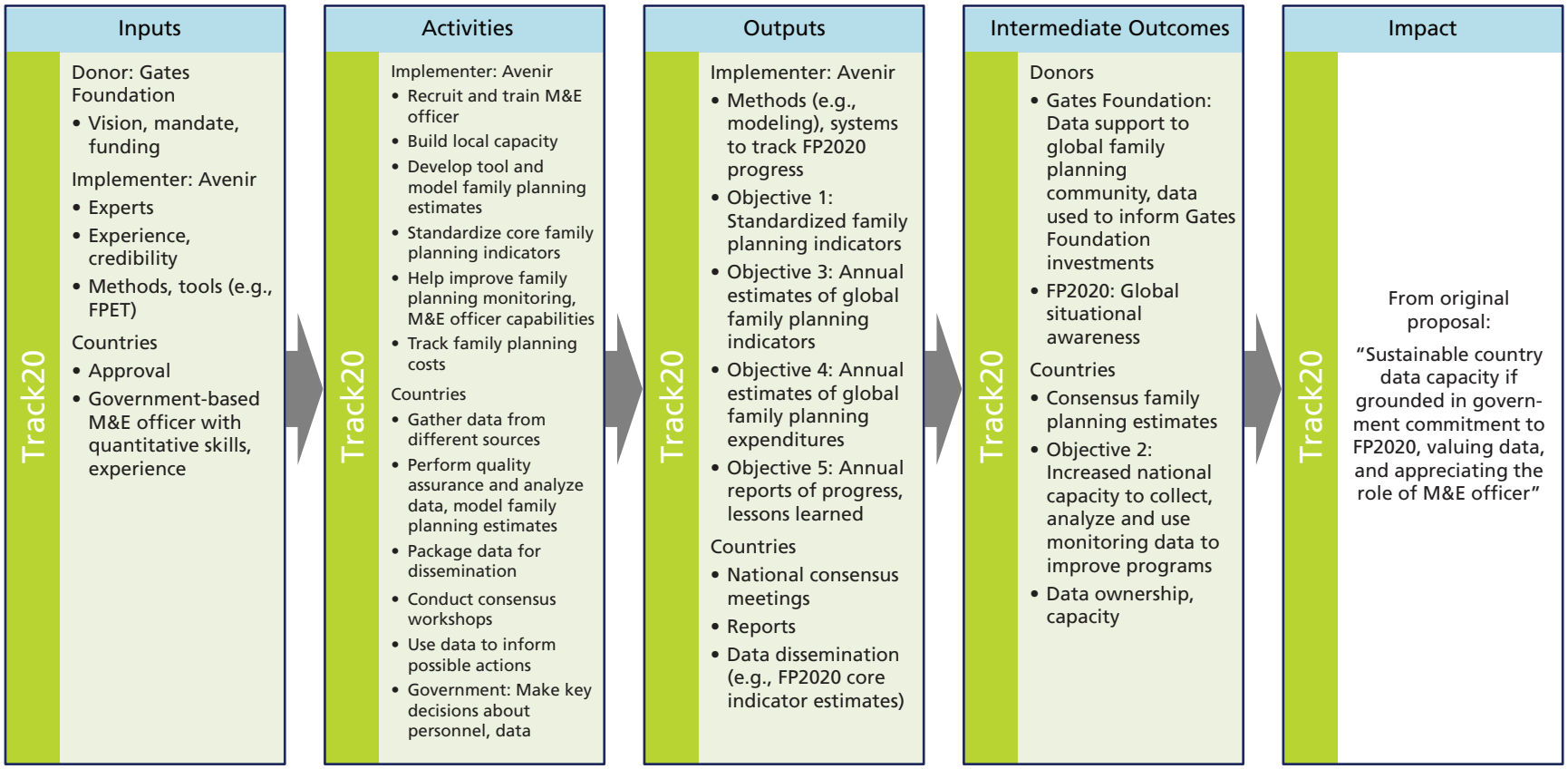
Program **inputs** reflect the donor, implementer, and country resources needed to support the program **activities** specified for both programs (e.g., from original proposals and as briefed to the RAND team by the two Gates Foundation grantees). **Outputs** reflect the tangible, short-term products of program activities (Marquis et al., 2016). Such outputs include methods and tools, data and indicator estimates, and data briefs and indicator tables that are disseminated to key stakeholders. Intermediate program **outcomes** are the results achieved using outputs (Gertler et al., 2011). Different from outputs, outcomes are less under program control due to external factors that affect them (Gertler et al., 2011). The PMA2020 and Track20 intermediate outcomes include data use by donors and countries, global situational awareness of progress toward the FP2020 goal, and increased country data capacity. Expected **impacts** reflect the desired long-term effects of PMA2020 and Track20 programs as expressed by the two grantees in their written proposals and during conversations with the RAND team. The two logic models show what the programs aim to achieve by providing data and evidence to decisionmakers through PMA2020 and Track20 efforts.

**Figure 4.1**
**Initial PMA2020 Logic Model**




RAND *RR2112-4.1*

**Figure 4.2**
**Initial Track20 Logic Model**




RAND *RR2112-4.2*

## The RAND Data Maturity Framework

The Gates Foundation asked RAND to develop and apply a framework to assess data maturity associated with PMA2020 and Track20 in program countries. While the present evaluation was not designed to attribute a country's data maturity specifically to the Gates Foundation–supported programs, the RAND data maturity framework provides a snapshot of the status of data maturity in those countries.

Our framework is tailored to PMA2020 and Track20. It uses a 10-point scale for rating maturity levels within three broader stages: beginning (scores 1–3), developing (scores 4–7), and advanced (scores 8–10). We adapted this structure from the Institute for Operations Research and the Management Sciences (see Appendix D) (INFORMS, 2017). Table 4.1 provides a general definition of each of the three basic maturity stages from the perspective of PMA2020 and Track20. The inclusion of the 10-point rating of maturity levels was intended to allow for a more nuanced reading on the variation in perspectives held by relevant staff and stakeholders within a single country and across countries.

The RAND data maturity framework (Table 4.2) includes three overarching domains; each domain is further disaggregated to generate a more detailed understanding of maturity within each one, for a total of 25 areas on which to assess, or score, maturity. The first domain covers **organizational readiness**, which includes areas related to staffing (number and qualifications of staff), staff buy-in (understanding of project and data relevance), leadership buy-in (leadership support for a data and analytics ecosystem), communication (quality and effectiveness among leaders and staff), and basic physical infrastructure. The second domain, **data systems**, is first divided into five subdomains—collection, management, analytics, governance, and institutionalization—with specific areas within each one. While data collection is more relevant for PMA2020, it can be used to assess Track20 as well if one considers the

**Table 4.1**
**RAND Data Maturity Framework Definitions**

| Maturity Levels | PMA2020 | Track20 |
|---|---|---|
| Beginning: 1, 2, and 3 | Standardized core questionnaire has been developed and data collection methodology established. Surveys take place semiannually or annually, but success is largely dependent on individual staff members leading the efforts. | Basic project management processes are established regarding data sources, analysis and reporting, and relationship building/collaboration with country stakeholders. However, the success of the project is largely dependent on individual staff members. |
| Developing: 4, 5, 6, and 7 | Process for training and oversight of resident enumerators is standardized and documented. Appropriate technology for data collection is used consistently. Sample is nationally representative. The survey life-cycle process is tailored to the country and clearly documented. | The process for data sources, analysis, and report development is documented, standardized, and applied regularly. Consensus-building workshops include key stakeholders, are tailored to stakeholder/country needs, and are clearly documented. |
| Advanced: 8, 9, and 10 | Data quality assurance measures are in place. Metrics are collected on barriers/obstacles and used to inform program/process improvements. Areas are identified for improved data collection innovations. Targeted modules have been adopted to respond to country needs. Data collection granularity (subnational level) is responsive to government needs. Stakeholder events effectively disseminate PMA2020 data. | Detailed measures for usefulness and uptake of data products are collected and used to ensure/improve the quality, dissemination, and use of data. Feedback is gathered from key stakeholders, and development of innovative data sharing/dissemination platforms responds to most/all stakeholder needs. |

**Table 4.2**
**RAND Data Maturity Framework for PMA2020 and Track20**

| | | | Data Maturity Level Description and Score | |
|---|---|---|---|---|
| **Domain** | **Subdomain** | **Area Description** | **Scores: Beginning (1–3), Developing (4–7), Advanced/Mature (8–10); (does not know = 0)** | **Score** |
| Organization | Readiness | Staffing | Are there enough qualified in-country PMA2020 project staff to carry out all data-related functions (collection, management, analytics)? | |
| | | | Are there enough qualified in-country Track20 project staff to carry out all data-related functions (collection, management, analytics)? | |
| | | Staff buy-in | How well do PMA2020 survey staff and supervisors understand the content and relevance of PMA2020 data? | |
| | | | How well do you feel that Track20 staff understand the content and relevance of data utilized by Track20? | |
| | | Leadership buy-in | To what degree do you feel that leaders of the organization where PMA2020 is based support PMA2020's family planning data and analytics? | |
| | | | To what degree do you feel that leaders of the organization/unit where Track20 is based support Track20's family planning data and analytics? | |
| | | Communication | How effective do you feel that workflow communications are among leaders/supervisors and staff of PMA2020? | |
| | | | How effective do you feel that workflow communications are among Track20 staff and supervisors? | |
| | | Infrastructure | How reliable is the basic physical infrastructure (IT, electricity, etc.) required for regular function of data systems managed by PMA2020? | |
| | | | How reliable is the basic physical infrastructure (IT, electricity, etc.) required for regular function of data systems used/managed by Track20? | |
| Data systems | Collection | Technology | How reliably does the technology function for PMA2020 data collection and management including programming? | |
| | | Response rates and completeness | To what degree are response rates high for household and facility surveys? | |
| | | Data quality | What do you feel is the level of quality of data collected by PMA2020 (e.g., how valid, reliable, and trustworthy)? | |
| | | | What do you feel is the level of quality of data from sources used by Track20 (e.g., valid, reliable, trustworthy)? | |
| | | Data sufficiency | How appropriate is the quantity of data collected by PMA2020 compared to data needs in [country]? (on a scale from 1 [very mismatched—either way too little or too much] to 10 [very well matched]) | |
| | | | How appropriate is the quantity of data used by Track20 compared to data needs in [country]? (on a scale from 1 [very mismatched—either way too little or too much] to 10 [very well matched]) | |
| | | Geographic granularity | To what extent does the the level of geographic granularity (local, state, national) of PMA2020 data meet policymaker needs (scale from 1 [definitely does not meet needs] to 10 [definitely meets needs]) | |

**Table 4.2—continued**

| Domain | Subdomain | Area Description | Data Maturity Level Description and Score<br>Scores: Beginning (1–3), Developing (4–7), Advanced/Mature (8–10); (does not know = 0) | Score |
|---|---|---|---|---|
| | | | To what extent does the level of geographic granularity (local, state, national) of data used by Track20 meet policymaker needs in [country]? (scale from 1 [definitely does not meet needs] to 10 [definitely meets needs]) | |
| | Management | Quality assurance processes/ practices | How well developed are data quality assurance processes/practices to ensure high-quality data for analysis? | |
| | | Quality improvement orientation | To what extent are PMA2020 data managers oriented toward identifying and implementing data quality improvements? | |
| | | | To what extent are Track20 data managers oriented toward identifying and implementing data quality improvements? | |
| | | Privacy protection | To what extent do PMA2020 surveys protect respondents against identification of individuals? | |
| | | | To what extent do the surveys or other data sources used by Track20 protect the survey respondents against identification of individuals? | |
| | | Integration | To what extent are survey (or service) data collected by PMA2020 integrated with past surveys and/or service statistics or other sources? | |
| | | | To what extent are survey (or service) data used by Track20 integrated with past surveys and/or service statistics or other sources? | |
| | | Documentation | What is the quality/adequacy of documentation of data sets collected by PMA2020 to facilitate data use by others? | |
| | | | What is the quality/adequacy of documentation of data sets used by Track20 to facilitate data use by others? | |
| | Analytics | Technology | How reliably do the hardware and software necessary for PMA2020 in-country analytics function? | |
| | | | How reliably do the hardware and software necessary for Track20 in-country analytics function? | |
| | | Analytics capabilities | What is the level of in-country analytics capabilities among those responsible for PMA2020 analyses? | |
| | | | What is the level of analytics capabilities among those responsible forTrack20 analyses? | |
| | | Analytics quality | How appropriate are PMA2020 analytic methods? | |
| | | | How appropriate are Track20 analytic methods? | |
| | | Analytics utility | To what extent are PMA2020 analyses useful in meeting the needs of data users (e.g., policy-/ decisionmakers, program managers)? | |
| | | | To what extent are Track20 analyses useful in meeting the needs of data users (e.g., policy-/ decisionmakers, program managers)? | |

**Table 4.2—continued**

| Domain | Subdomain | Area Description | Data Maturity Level Description and Score<br>Scores: Beginning (1–3), Developing (4–7), Advanced/Mature (8–10); (does not know = 0) | Score |
|---|---|---|---|---|
| | Governance | Data use policy and accessibility | To what extent are standards applied to govern the use of PMA2020 data? | |
| | | | To what extent are standards applied to govern the use of Track20 models and data? | |
| | | Security policy/ protection | What is the degree of adherence of PMA2020 data systems to security/protection standards? | |
| | | Standards development and adoption | How well do PMA2020 data sets that are released to the public conform to quality standards? | |
| | | | To what extent do Track20 analytic tools that are released to the public conform to quality standards? | |
| | Institutionalization | Policies | To what degree do government policies facilitate family planning data collection, management and use in [country]? | |
| | | Data collection processes | To what degree are data collection processes well established (institutionalized) in [country]? | |
| | | Data management processes | To what degree are data management processes well established in [country]? | |
| | | Data analytics processes | To what degree are data analysis processes well established in [country]? | |
| | | Data sharing processes | To what degree are data sharing processes well established in [country]? | |
| | | Data communications processes | To what degree are data communications processes well established in [country]? | |
| Data use | Use | Valuation | To what degree do PMA2020 data satisfy the needs of decisionmakers to inform their policies, programs, and/or plans? | |
| | | | To what degree do Track20 analytic tools and estimates satisfy the needs of decisionmakers to inform their policies, programs, and/or plans? | |
| | | Ownership/ stewardship | What is the perceived degree of ownership (by country) of PMA2020 survey and service data collection and data sets? | |
| | | | What is the perceived degree of ownership (by country) of Track20 analytic tools and data sources? | |
| | | Stage of change toward regular use of data for informed decisionmaking | For this question (only): Please indicate which of the following (1 to 6) best describes where you are regarding the use of PMA2020 data: (1) Aware of the data; (2) Intend to use the data but have not yet; (3) Understand the data but do not use them yet; (4) Confident in my ability to use the data; (5) Occasionally use data; (6) Regularly use the data | |
| | | | For this question (only): Please indicate which of the following (1 to 6) best describes where you are regarding the use of Track20 estimates: (1) Aware of the data; (2) Intend to use the data but have not yet; (3) Understand the data but do not use them yet; (4) Confident in my ability to use the data; (5) Occasionally use data; (6) Regularly use the data | |

sourcing and integration of other surveys (e.g., the DHS, MICS) as the data collection process. The third domain, **data use** for decisionmaking, focuses on the progress toward sustained, regular use of PMA2020 data and Track20 estimates, as well as analyses produced by both programs. As noted in the table, each item is rated (scored), mostly on a scale from 1 through 10, with one item scored on a scale from 1 through 6. We asked in-country stakeholders to rate each item (for programs with which they are familiar). We calculated average scores by element, by country, and by type of stakeholder. Detailed results are presented in Chapter Twelve.

## The RAND Sustainability Framework

Based on the early focused literature review, we developed a sustainability framework for PMA2020 and Track20 that organizes what we have termed *sustainability-enabling factors* into four broad thematic categories—financial sustainability, technical sustainability, operational sustainability, and data culture (Table 4.3). The table highlights (in bold font) several of the factors and associated measures that we believed were the most essential to this evaluation and, therefore, that we assessed in countries through both open-ended and structured questions. While we believe that all the sustainability-enabling factors are important, we strove to minimize the burden on the interviewees rather than ask about all of them. We used this framework to explore in-country stakeholders' perceptions on sustainability; detailed results are presented in Chapter Thirteen.

**Table 4.3**
**RAND Sustainability Framework: Enabling Factors and Associated Measures/Indicators**

| Category | Sustainability-Enabling Factor | Measure |
|---|---|---|
| Financial | **(Co-)Financing** | Is the program co-financed by more than one donor or with the country government? (Yes/No)<br>If no: Is there any discussion on co-financing? (Yes/No)<br>Percentage of program funded by donors |
| | **Domestic resourcing** | Percentage of program funded through domestic resources (financial and/or in kind) |
| Technical | **Data system and analytics capabilities** | Perceived level of capabilities of technical data support and analytic personnel (1 = very low to 5 = very high) |
| | **Hardware and software maintenance** | Perceived availability of qualified vendors/technicians (1 = very low to 5 = very high)<br>Reliable mechanism is in place for maintenance of (a) IT hardware and (b) software maintenance (1 = disagree strongly to 5 = agree strongly for [a],[b]) |
| | Education: Integration of relevant curricula into educational system | Perceived efficacy of data-oriented curriculum |
| | Training: Internationally accredited training facilities | Number of health data system management and/or analytics certification or training programs available in (or to) country |
| | Personnel pipeline: Human resources system with training and career track for data systems and analytics professionals, including professional development and mentorship | Demonstrated career track for data professionals (e.g., government positions at progressively higher levels)<br>Perceived availability of professional development opportunities/mechanisms for data systems and analytics professionals |
| Operational/programmatic | **Leadership buy-in** | Perceived efficacy and support of relevant country leader(s) for data collection and use (1 = very poor to 5 = very good) |
| | **Country ownership of data** | Degree to which country perceives that it owns data and information systems (1 = very low to 5 = very high) |
| | **Satisfaction of policymakers' data needs** | Degree to which policymakers depend on data to inform programming (1 = not at all to 5 = very high)<br>Satisfaction with content, format, timeliness, and granularity of data (1 = very low to 5 = very high) |
| | **Accountability** | Perceived governmental accountability for improved use of data to inform programs and policy (1 = very poor to 5 = very good) |

**Table 4.3—continued**

| Category | Sustainability-Enabling Factor | Measure |
|---|---|---|
| | **Stakeholder buy-in (planning)** | Perceived stakeholder involvement in planning for data generation among those in: (a) local communities, (b) subnational government, (c) national government, (d) civil society, (e) donors (1 = none to 5 = very involved for each)<br>Existence of a multistakeholder planning body (Yes/no)<br>If existent, perceived efficacy of a multistakeholder planning body (1 = very low to 5 = very high)<br>Perceived cooperation in planning (1 = very low to 5 = very high)<br>Perceived cultural acceptability of project goals and methods (1 = very low to 5 = very high) |
| | **Use of local expertise** | Perceived degree of engagement of government offices (MOH, national statistics office), local organizations, or universities to perform data collection, operations, and analytics (1 = very low to 5 = very high) |
| | **Planning for sustainability** | Perceived degree to which program design, implementation, and performance monitoring address sustainability |
| | **Collaboration (implementation)** | Perceived cooperation across governments, donors, and civil society in program implementation<br>Perceived existence of country-driven coordination (rather than donor-driven) |
| | **Country ownership of program** | Degree to which programming operates out of MOH (whether self-contained within MOH or through collaborations) |
| Data culture | **Data accessibility** | Perceived accessibility of PMA2020 data and Track20 estimates (1 = very low to 5 = very high) |
| | **Trust in data accuracy** | Perceived trust in PMA2020 data and the accuracy of PMA2020 data and Track20 estimates<br>(1 = very low to 5 = very high) |
| | **Data impact on outcomes** | Perceived contribution of PMA2020 data and Track20 estimates to achievement of program outcomes (1 = very low to 5 = very high) |
| | **Institutionalization of data use** | Perceived degree to which PMA2020 data and Track20 estimates are regularly used by program managers and policymakers in the country<br>(1 = very low to 5 = very high) |
| | **Conducive policy/regulatory environment** | Degree to which country or organizational policy facilitates and does not impede collection, sharing, and use of program-relevant data |
| | **Data impact on services** | Documented or perceived relevance or impact of data on family planning program services |

CHAPTER FIVE

# Stakeholder Views on Family Planning Data Needs

## Overview

This chapter addresses the family planning data needs of decisionmakers in the 15 countries included in this evaluation. We asked decisionmakers and PMA2020 and Track20 program staff about their data needs for decisionmaking, including their desired data frequency and geographic granularity (i.e., providing national versus subnational estimates). The purpose was to explore whether PMA2020 and other family planning data sources are producing the kind of data that program managers want and need.

## Desired Frequency of Family Planning Data

**The desired frequency depends on the indicator and its use.** When considering the ideal frequency of data related to family planning, respondents from the different program countries clearly differentiated between routine service statistics and surveys such as PMA2020. They emphasized several factors that influence the desired frequency: how frequently one expects the indicator in question to change; what interventions are being undertaken as a result of the data and on what timeline; what types of decisions and actions are to be informed by the data; the cost of data collection; and the audience for the data. For instance, respondents said that annual estimates are sufficient for reporting to the global community, determining annual budgets, and policymaking, but that quarterly data are preferred for day-to-day program implementation and tracking family planning commodities. A governmental representative in Tanzania had a similarly nuanced view: Monthly availability of data estimates on commodities and essential medicines is ideal; indicators to track progress and plan interventions should occur every three to six months; and larger surveys to monitor high-level trends, such as total fertility rate, adolescent birth rate, and year at first marriage and birth, could be reported every five years.

**Service statistics are needed more frequently—quarterly or even monthly.** Decisionmakers cited the importance of frequent, near-real-time service statistics to evaluate the supply of commodities and change service delivery approaches as needed. We asked what the ideal frequency of reporting service statistics might be, and respondents most often noted that service statistics are reported monthly and typically reviewed either monthly or quarterly. Respondents felt that the more real-time the data are, the higher quality they tend to be. Additionally, a Tanzanian respondent from the Health Information Management System described how the minister had begun requesting monthly rather than quarterly service statistics, prompting

the development of an efficient, rapid-cycle process for collecting, cleaning, summarizing, and providing the data to the minister on a monthly basis.

**Most stakeholders feel that PMA2020 household surveys should be annual.** When asked specifically about the frequency of family planning household surveys such as PMA2020 (in applicable countries), a few respondents supported data collection every six months, particularly for the purposes of "midterm review" of programs. However, various PMA2020 staff registered an opposing view, noting the resource intensiveness and fatigue on the part of the PMA2020 team that resulted from this frequency during the first two years of the program in each PMA2020 country. Most respondents felt that annual surveys were sufficient, as programs are typically reviewed and rebudgeted on an annual basis. Some respondents advocated for tailoring the frequency of data collection to the individual country's context; for instance, Indonesian respondents noted that the total fertility rate is already low, at 2.4, making frequent data collection on this indicator unnecessary (and, indeed, PMA2020 staff noted that this indicator may be dropped altogether from the survey).

Similarly, U.S.-based stakeholders, PMA2020 staff, and other decisionmakers noted that it takes time to intervene and for indicators to change. Very few respondents advocated for data collection more frequently than yearly, and some even mentioned that one year may not be long enough to detect a meaningful change in certain indicators after an intervention.

While there was consensus that many of the indicators were not expected to change every six months, some indicators may warrant such frequent collection—for example, contraceptive method mix and switching (between methods). Questions about switching contraceptive method, however, are not consistently asked in PMA2020 surveys. In contrast with indicators from the household surveys, more-frequent data from the service delivery point surveys were noted to be useful. One individual suggested conducting PMA2020's household surveys annually but its service delivery point surveys more frequently.

Even staff at the Gates Institute were skeptical of the value of such frequent data collection. They have removed a question regarding total fertility as a result—this indicator is not likely to change on a six-month or even annual basis.

**The desired frequency of data depends on the frequency of decisions to be made.** Rather than asking which indicators need to be collected frequently, it is better to reframe the question to ask about the frequency of decisions to be made and what data are needed to inform those decisions.

In Chapter Seven, we present the analysis of the change in various indicators from round to round to compare empirical results with the stated data needs of stakeholders above.

## Desired Geographic Granularity of Family Planning Data

While decisionmakers at different levels need different levels of geographic granularity, **most in-country stakeholders called for collecting subnational data.** The overwhelming majority of respondents across all countries and stakeholder groups expressed a strong desire for subnational estimates in addition to national estimates. Stakeholders in Indonesia, Nigeria, and Pakistan mentioned the decentralization of the governmental structure, including the MOH, as a reason for needing more district- and provincial-level family planning data. Specifically, budgeting, allocation of resources, and developing workplans (such as costed implementation plans) typically occur at the district level. Thus, respondents need district-level data "to make

the right recommendations to smaller geographic areas" and because "no one is programming for a whole country."

Respondents in several countries (India, Ethiopia, Nigeria, and Ghana) repeatedly emphasized that national estimates "mask" the true variation and "huge disparities" within the population and they want to "zoom in on where the problems really are." In India, a "vast" country, even state-level estimates "are not going to be useful . . . because the averages don't tell anything to the policymaker." Similarly, a respondent from Ghana noted, "The real challenges are hidden in the statistics. If you do the subregional you actually get it closer and closer to the realities."

**This finding, however, may not be as pertinent for smaller countries.** There were a few respondents, mainly from countries with smaller populations, such as Zimbabwe (a country with less than 9 percent of the population of Nigeria), who advocated for national estimates only. National estimates were felt to be important for planning (for instance, related to commodities), resource allocation, and following big-picture trends to track achievements. More commonly, respondents acknowledged that both national and subnational estimates have a role in decisionmaking, with respondents commonly stating that national estimates are important for budgeting, policy, and monitoring trends, but in order to target and successfully implement interventions, and encourage accountability at local levels, subnational data are required.

## Summary

Overall, the desired frequency of data collection varied by type of data. Stakeholders called for more-frequent service statistics (quarterly or even monthly) and less-frequent survey data. Specifically, they find annual PMA2020 surveys to be sufficient for their purposes. They also find service statistics to be immensely useful but of variable quality. Data users expressed a clear need for subnational family planning data in order to inform local decisionmaking. These perspectives inform the analyses performed in Chapter Seven, the proposed options for modifying PMA2020's design described in Chapter Eight, and, ultimately, the recommendations offered in Chapter Fourteen.

CHAPTER SIX

# PMA2020 Goals, Accomplishments, and Challenges

## Overview

We sought stakeholder views of PMA2020's goals, accomplishments to date, and the challenges it faces going forward, including its utility to FP2020 and to the 11 countries in which it is now implemented. In general, the stakeholders we interviewed at the Gates Foundation, elsewhere in the United States, and in-country felt that PMA2020 had laid the foundation for collecting high-quality data on family planning. They valued the fact that the PMA2020 survey was fielded annually and that data were quickly available for monitoring family planning activities.

However, stakeholders also felt that PMA2020 had yet to meet some of its original goals—in particular, promoting use of data for family planning needs at the local level. They viewed this evaluation as an opportunity to refine the survey's goals and uses and to more clearly distinguish it from other family planning data sources.

In this chapter, we summarize stakeholder views, highlighting key themes that emerged from the interviews. We conclude by presenting an empiric logic model, based on stakeholder assessments, reflecting revisions to the initial model presented in Figure 4.1.

## Goals of PMA2020

The original objectives of PMA2020 proposed by the Gates Institute were to (1) expand country-level monitoring capacity, (2) integrate a rapid data collection system using mobile devices, (3) improve data monitoring to strengthen service delivery, and (4) promote the use of data to respond to family planning needs at the community level. While the first and third objectives were also objectives for Track20, the Gates Institute spelled out its own specific activities and intended achievements. As understood by the other U.S.-based stakeholders, including the PMA2020 External Consultative Group, the program aimed to (1) provide more-frequent family planning data to complement the DHS, (2) provide national-level estimates of key family planning indicators (but also produce provincial- or state-level data where possible), and (3) report these data to the global community to monitor progress toward achieving FP2020's goal of 120 million new users of modern contraception by 2020. They viewed data use by country decisionmakers as a secondary, albeit important, goal.

Stakeholders had different perspectives on these goals. Gates Foundation staff described enthusiasm within the foundation about potential innovative directions for PMA2020, beyond simply serving as a tool to collect family planning data more frequently. However, their opin-

ions differed about which of the potential directions should be pursued, as well as the target consumers for the PMA2020 data. More than one Gates Foundation staff member noted that the **primary objective of the survey is "good data" that are "hopefully being used"** to monitor performance of various national- or subnational-level efforts and to ensure accountability of government decisionmakers, as the name *PMA2020* suggests.

The strategic vision for and scope of PMA2020 have been under discussion across the Gates Foundation, the External Consultative Group, and within the Gates Institute. Whereas the External Consultative Group recommended in 2016 that PMA2020 "stay the course" in focusing exclusively on family planning, they noted in 2017 that the vision for the scope of PMA2020 should be clarified, recognizing that non–family planning content might offer appealing opportunities for co-financing of surveys, year-round retention of resident enumerators, and shifting the perception of PMA2020 as a strictly vertical (content-focused) program to also include a more horizontal (system-focused) orientation.

Discussions at the PMA2020 External Consultative Group meetings in 2016 and 2017 reflect a consensus that the scope of PMA2020 needed to be clarified and refined (PMA2020, 2016a; PMA2020, 2017e). In both years, the group emphasized opportunities for innovation in survey content and methods; greater focus on data use in countries; and the need to reduce survey costs, regardless of survey scope. The idea of innovation was also raised in individual interviews and is under active consideration by the Gates Institute team—the potential role of PMA2020 as a testing ground for innovation in survey content (e.g., family planning questions or modules, non–family planning modules) and methods (e.g., sampling schemes, data collection). One of the U.S-based stakeholders noted that the Gates Institute team "is built for that," and another commented that there have been "lost opportunities to be more innovative."

Suggested innovations to both the purpose and the survey design include adapting PMA2020 to be "some sort of continuous survey with flexible, rotating modules" and/or potentially experimenting with different designs to make it a more useful tool for evaluating specific family planning programs. The goal would be to clarify what interventions do or do not work well, in what subpopulations, and why, especially at the local and other subnational levels.

**Stakeholders disagreed about whether PMA2020 should serve primarily the global family planning community or country-specific needs.** There was disagreement about whether PMA2020 is, or should be, mainly focused on producing data at the country level to be reported to the global community to monitor progress toward a common goal, if it should be focused on priority geographies within countries, or if it should produce data to improve specific family planning programs that may operate at a very local level. For instance, when asked about examples of "actionable" PMA2020 data, one stakeholder provided two examples that are quite local in nature—namely, information on where stock-outs of contraceptive supplies were occurring and where family planning–related information is or is not provided to women.

We also noted varying opinions about the extent to which PMA2020 data are intended to serve the interests of the global family planning community to measure progress toward the FP2020 goals and/or also to serve the interests of countries to guide and improve their family planning programs. Stakeholders questioned whether the survey could both produce national estimates of key indicators and signal potential gaps in family planning programs that countries can use at national or even subnational and local levels. One Gates Foundation respondent commented that if both could be achieved, it would be the "holy grail" for PMA2020.

In-country stakeholders explained that the PMA2020 program promised to release data every six months in order to present a representative national picture even when it does not operate in all of a country's regions. The rapid data collection and analysis allows family planning managers to assess progress almost in real time before information becomes obsolete. When asked about the intent of PMA2020, country principal investigators focused on its role for annual data collection in service of (global) FP2020 goals. They argued that its data are "more or less an alternative to the DHS" and a reliable data source for governments to make family planning decisions.

Gates Foundation stakeholders articulated secondary goals of PMA2020 that serve local needs, including employing, empowering, and building capacity in a cadre of resident enumerators; serving as a proof of concept for a rapid data collection platform that could be expanded beyond family planning; and serving as a testing ground for optimizing the cost-effectiveness of such a platform, with ultimate goals of lowering unit costs of the surveys and obtaining early financial buy-in from other partners.

**PMA2020 data are not well integrated with other in-country data.** Some stakeholders expressed their views about what PMA2020 is *not.* According to one interviewee, the Gates Foundation needs a more "strategic perspective on the role of surveys" and how they intersect with routine data systems. Another noted that PMA2020 was always very "vertical" (meaning that it was narrowly focused on family planning) and was never designed to strengthen data systems on the ground. One U.S.-based stakeholder described PMA2020 as "expensive, boutique, and not sustainable." If PMA2020 continues to be fairly vertical, not mainstreamed in a country's routine data collection systems, and without a clear strategy for sustainability, the Gates Foundation staff felt that it will "continue to be seen as a vertical approach feeding the FP2020 beast." From their perspective, at least, it appears to have been a challenge to combat the perception that PMA2020 data are "mainly for global visibility" and for use by the Gates Foundation or FP2020 rather than by the countries themselves.

**Demand for more actionable, local data is strong.** Both Gates Foundation staff and other U.S.-based stakeholders felt that the goals of PMA2020 had evolved; among in-country respondents, PMA2020's goals were viewed as more consistent over time. In general, national family planning estimates are now seen as less valuable than subnational estimates because the hypothesis that more national data would have an impact has not been proven. With time, it has also become increasingly apparent that more frequent data are not necessarily better or needed, as some indicators do not change rapidly enough to warrant semiannual or even annual data collection.

PMA2020 is facing increasing demands by countries to move to more subnational and more actionable local data. One External Consultative Group member commented that the new target audience should be country stakeholders, noting in particular that "institutionalizing demand at the country level will also require being responsive to country needs in the design of questionnaire and the content that is covered."

There is also interest, although not universal, in expanding beyond family planning data to meet the data needs of decisionmakers—both internally within the Gates Foundation and within countries—who are concerned with how family planning fits into larger maternal, child, and newborn health agendas. Perhaps reflecting that interest, both Gates Foundation and Gates Institute staff wanted more direction from the other regarding an overarching vision for PMA2020, key priorities, and the way forward.

## Accomplishments of PMA2020

**PMA2020 has demonstrated that its proof-of-concept survey can collect rigorous, detailed family planning data more frequently and provide results within a shorter time frame than other surveys.** In particular, PMA2020 has achieved its second original objective of integrating a rapid data collection system using mobile devices. The Gates Institute touts the speed (six weeks for data collection, six weeks for first data outputs, six months for final data sets for the public), frequency (semiannually for two years, then annually), quality, flexibility, affordability, and sustainability of its surveys, as well as the linkage between households and service delivery points in each enumeration area. Most stakeholders felt that PMA2020 has achieved one of its original objectives: proving that it is possible to train a cadre of resident enumerators to rapidly collect statistically valid family planning data using mobile phones; have local staff working with the grantee institution to provide rapid turnaround of quality-checked data; and produce estimates that are comparable (to a large extent) to those from the DHS, but more frequent.

Specifically, PMA2020's use of resident enumerators not only serves data collection needs but also helps to empower those women with new skills, social stature, and employment. As of April 2017, the program had trained more than 1,700 resident enumerators, fielded 36 survey rounds, and conducted more than 350,000 household and service delivery point interviews in ten countries. While recruiting and retaining qualified resident enumerators has proven challenging in rural areas, the level of trust engendered among survey respondents who share a language and culture with the resident enumerators is seen as an important benefit.

Gates Foundation staff generally view PMA2020 as an innovative model platform that could be expanded beyond family planning to tackle larger challenges in global health. Indeed, it has already created enthusiasm for the many applications of this data collection methodology. Of note, those who are directly involved with supporting its implementation by the Gates Institute note its limitations—namely, its overall cost, its lack of horizontal integration into other data collection efforts (family planning or otherwise), and its limits on promoting actual use of the data generated by the survey. Additional challenges are discussed below.

In-country respondents were generally positive about the program, commenting that "it requires a lot of energy, and, more than energy and ambition, it requires commitment." Country decisionmakers appreciated its frequency and called it "a reliable data source . . . that we can also call on for quality data and decisions." One NGO leader remarked, "I know what it takes to conduct a very good survey, and the PMA2020 survey is a good survey." Finally, respondents praised PMA2020's ability to produce more-precise state-level estimates than MICS "because of the relatively large sample sizes" per state.

Indeed, the accomplishments to date of PMA2020 have been significant (especially considering the small number of staff at the Gates Institute), although one respondent added the qualifier "if the goal is vertical [family planning] data" that are not well integrated with other family planning and non–family planning data sources.

**PMA2020 is beginning to demonstrate its value to decisionmakers.** Both U.S.-based and in-country stakeholders commented on several characteristics of PMA2020 that set it apart from other surveys, most notably the DHS: its frequency of data collection, rapid turnaround of results, in-depth family planning data, and linkage between household and service delivery point data.

**Annual fielding of the PMA2020 survey** is seen as an important advantage. Respondents overwhelmingly noted the annual nature of PMA2020 surveys as a comparative advantage over the DHS. Nearly everyone who commented on the DHS noted that its infrequent data collection, every five years, was a limitation. The Gates Institute and other U.S.-based stakeholders repeatedly emphasized the value of more frequent estimates than the DHS provides. Similarly, the frequency of the data appeals to government and NGO stakeholders, who find the DHS data "not very revealing because it's every three to five years."

**Stakeholders praised PMA2020's rapid turnaround and predictable timetable.** U.S.-based stakeholders commended the two-page reports that are disseminated rapidly after each round, the public nature of the data sets, and the speed with which they are released after being assembled and cleaned. In-country decisionmakers had a similar view. Some complained that delivery of DHS or MICS data is unpredictable, and they appreciated that they knew when to expect PMA2020 updates. A government employee commented, "I think the biggest thing that they have going for them is the fact that they can turn things around quickly, the fact that the data can be available in six months."

**PMA2020 is seen as providing "rich" information and more detail on key family planning indicators** than other surveys (such as providing a breakdown of method mix within mCPR). It also asks questions relating to quality of care (such as the percentage of women who are counseled on side effects of different forms of contraception), which decisionmakers view as very useful for tracking program progress and assessing targeted interventions and as a much-needed complement to aggregated service statistics that only provide supply-side data. More than one in-country stakeholder noted that the depth of information on family planning that PMA2020 provides makes it an important complement (although not replacement) for the DHS.

Stakeholders also considered PMA2020 in comparison with another important source of family planning data: service statistics, often collected in an HMIS tool like DHIS2 and considered to be of variable quality and accuracy in different countries. PMA2020 was generally viewed as a complement to this frequently updated, useful, but aggregated source of data that provides information on service receipt (i.e., supply but not demand). The main comparative advantage that PMA2020's household survey has over service statistics is that it provides rich information at the individual level—for instance, it collects data on demand for services as well as the perceived quality of services received. Important uses of PMA2020's service delivery point surveys are to fill in the gaps in the service statistics' quality and completeness and to triangulate with data that are common to both.

While it is unclear whether the analyses fully harness the data linkages between households and facilities in each enumeration area, U.S.-based stakeholders noted the **linkage between household and service delivery point data** as another comparative advantage of PMA2020 because it aims to relate family planning supply (service delivery points) to demand (household) data.

## Challenges for PMA2020

The achievements of PMA2020 are significant. However, it is less clear that the program has fully realized its first, third, and fourth original objectives—to expand monitoring capacity in countries, to improve data monitoring to strengthen service delivery, and to promote the use of

its data to respond to community-level needs. PMA2020 faces several challenges as it strives to achieve its original objectives. We discuss issues related to these original objectives next. Challenges related to sustainability (modified objectives 2 and 3) and integration into countries' M&E architecture (modified objective 4) are discussed in more detail in Chapters Twelve and Thirteen, which address data maturity and sustainability, respectively.

#### Objective 1: Expanding Monitoring Capacity

PMA2020 initially proposed to expand in-country capacity to monitor progress toward FP2020 goals, in a "parallel but joint" effort with Avenir Health. Expanding capacity meant improving the ability of country M&E officers, stakeholders, and others to conduct evidence-based decisionmaking around family planning programming and resource allocation. Avenir (through Track20) assumed sole responsibility for the M&E officer mode. PMA2020's role in expanding monitoring capacity appears to have shifted away from M&E officers and has been limited mainly to the data collectors (i.e., resident enumerators) and the small PMA2020 country teams, mostly within universities.

Country ownership of the PMA2020 survey and capacity-building are mutually reinforcing and interdependent. As they become more comfortable with the operational details of PMA2020, **in-country stakeholders want more control over the PMA2020 survey**. Most stakeholders praised the frequent interaction with the Gates Institute when it was preparing to launch the surveys and planning the trainings for data managers and resident enumerators. However, some stakeholders said that they did not feel that their views were adequately considered by the Gates Foundation or the Gates Institute during initial planning for the surveys. Respondents in one country also disagreed with the decision to host data on the Gates Institute's servers, preferring to host the data locally and ensure "true security." These respondents would also like access to the server where raw data are stored in order to do their own data analysis.

Relatedly, because of the burden of translating the PMA2020 survey into sometimes multiple local languages and then training resident enumerators to administer those different versions, some in-country stakeholders recommended a more decentralized model in which changes to the survey no longer needed to occur at the Gates Institute. This would permit last-minute changes that arise during training and would limit delays.

On a positive note, in some countries, the close working relationship between the PMA2020 staff and the MOH has promoted domestic ownership and transformed the PMA2020 survey into the primary source for data to track family planning indicators. (Country ownership, as one of several data maturity and sustainability-enabling factors, is discussed in more detail in Chapters Twelve and Thirteen.)

#### Objective 3: Improving Data Monitoring to Strengthen Service Delivery Systems

As with Objective 1, this objective was originally intended to be jointly accomplished with Track20 and centered around, among other activities, improving the quality of service statistics and strengthening service delivery systems by forging collaborations with other surveys, such as the DHS and MICS, as well as country HMISs, to build integrated systems for FP2020 indicator reporting. Once again, strengthening service statistics evolved to fall squarely in the purview of Track20, while PMA2020's contribution to strengthening service delivery systems is mainly the service delivery point surveys, which are conducted in parallel with household surveys. As noted by multiple stakeholders, the service delivery point surveys

and household surveys are not yet systematically linked so that they can be used as envisioned, nor are they being systematically used to triangulate with routine service statistics and, thus, serve as a quality check. In-country stakeholders gave few examples of the usefulness of the service delivery point surveys, and it is unclear to what extent these surveys are being used to concretely improve service delivery.

#### Objective 4: Promoting Use of Data to Meet Community-Level Needs

The fourth original objective for PMA2020 was "promoting the use of data to respond to family planning needs at the community level." Most stakeholders felt that this objective had not been fully realized, noting, for example, various challenges with helping in-country decisionmakers (including program leadership, policymakers, and MOH representatives) understand the PMA2020 data and use them for planning purposes. The Gates Institute is considering ways to meet demand for subnational PMA2020 data but tends to leave advocacy and data use to others. Chapter Eleven discusses data use in more depth.

**PMA2020's collaboration with local universities is commendable.** Many respondents praised the academic partnerships as a strength of PMA2020. The Gates Institute has long-standing relationships with the researchers and staff, and the researchers there represent a potential audience for (one way of) using the data. Universities are seen as unbiased, apolitical, and potentially more reliable than governmental bodies. Local universities are also logical places to build in-country capacity for data use and analysis, in comparison with partnering with governments. Respondents observed that governments had high turnover, sometimes lacked commitment, and faced challenges with sustaining interest over time. However, a common observation was that **universities are not always well linked to the MOH family planning programs** that would most benefit from PMA2020 data. Lack of linkage negatively impacts the survey's visibility and decreases government buy-in when the MOH sees itself as a passive recipient of data rather than an active contributor to the process.

**In some countries, PMA2020 is still struggling to prove its value and, more specifically, demonstrate its usefulness to decisionmakers.** The Gates Institute team felt that both the U.S.-based and country PMA2020 teams could do more to support data dissemination and use in countries—for example, by promoting data use, hosting more data workshops, and producing more publications. They noted that they had focused on data generation during the first four years and now hope to expand the focus to these additional areas. They initially felt that just getting the data out to the public would be enough to promote data use; they learned that this is not the case.

PMA2020's External Consultative Group discussion in 2017 suggested a "general agreement on shifting toward country stakeholders as PMA2020's primary audience, as opposed to global-level FP2020 monitoring, as originally envisaged." The discussants noted that "it will be important to focus on **institutionalizing demand for PMA2020 at the country level** and positioning PMA2020 as part of or [a] complement to a country's health information system."

These comments were consistent with the perspectives of Gates Foundation staff, who noted that PMA2020 may be limiting itself by focusing narrowly on FP2020 indicators. These staff reflect a desire to create a **data culture** by enhancing country demand for and use of data, including the possibility of broadening the scope beyond family planning data (while being careful not to "overload PMA with too many things"). There was interest among Gates Foundation respondents for PMA2020 to "dissociate" itself from the FP2020 core indicators (which

many, including in-country stakeholders, view as top-down measures) and market itself as a tool for countries to do more-continuous or regular surveys in a different way.

Several PMA2020 staff (particularly in Nigeria) highlighted the desire for a reorientation toward more **actionable information** for family planning programs. This reorientation would require a focus not just on data generation but also on data *use*. For instance, a PMA2020 leader in Nigeria said,

> We have to find ways of making our data useful to [end users] and go beyond collecting data to actually . . . working with them to make the data useful to them. That is something that, I think, initially we're not prepared for. We are hearing more and more requests to improve in that direction and we are willing to take on the challenge. We just need to build in the capacity to train people that will need the data, to work with them more closely. We're beginning to do that.

This individual went on to say that her colleagues were preparing to attend a Track20-sponsored training workshop to present on PMA2020. She noted, "I'm very happy that we have very good quality data, but if nobody uses it, what have we done all of this for?"

### Other Challenges

While PMA2020 has achieved varying levels of visibility, trust, and, perhaps, respect in different countries, in other countries, **PMA2020 still faces challenges to its visibility and reputation**. An example of success is from Burkina Faso: At first, the MOH was skeptical of data not produced "in house" and needed to be sure that the process and results were rigorous and reliable. PMA2020 worked to collaborate with the National Statistics Institute and to better interface with the MOH. PMA2020 staff joined with NGOs to form a reproductive health technical working group, which meets regularly to advance common goals. Once the rigor of the PMA2020 approach was accepted by key decisionmakers (after the second or third round of data collection), PMA2020 became the survey that decisionmakers in Burkina Faso seek out first.

However, in other countries, PMA2020 is less well known and has a less well-established reputation, particularly after just a couple of rounds of data collection. In Ghana, a perception of PMA2020 is that "the researchers do that." Similarly, while dissemination activities occur in the PMA2020 program countries, efforts to engage other key partners and decisionmakers vary by country. For instance, an NGO representative in Nigeria stated,

> I think they need to engage more. They need to give the PMA survey the type of stature . . . that is given to the DHS. The DHS, for example, used to involve people from far and wide, people from the other development agencies, the other development partners, people from government, people from . . . . I know they do this, but I think they need to do better . . . it will give more visibility, more trust and more respect and more use for the data.

**PMA2020 faces the challenge of comparisons to the well-established, respected DHS.** One of the challenges faced by local PMA2020 champions is the existence of alternative family planning–related surveys, either those supported by domestic agencies or an overwhelming preference for alternatives to PMA2020, most notably the DHS. Decisionmakers most frequently compared PMA2020 to their country-specific DHS surveys, which are viewed as the "go-to" survey, "the most authentic source of data in the country," and the "gold stan-

dard." Indeed, the DHS was universally discussed as a trusted, reliable source of family planning data, at least in part because it has been conducted for more than 30 years. In countries where PMA2020 operates in only some of the states or provinces, it has struggled to gain the same traction as the DHS. As a decisionmaker in Uganda commented, "Honestly, other than PMA giving us annual estimates . . . even now people still quote the DHS."

In Rajasthan, India, for example, despite an interest in scaling up PMA2020 to more provinces, the government is currently promoting the dissemination of the National Family Health Survey (NFHS), which operates under the DHS program, as well as installing agile systems to support the HMIS infrastructure. A respondent representing the Population Council in India noted that "while more data is better for decisionmaking and planning purposes," the different estimates produced by the NFHS and PMA2020 (particularly the sterilization estimates) have caused confusion among policymakers.

However, the DHS is carried out only every five years; PMA2020 was designed to provide more-frequent, and more-detailed, family planning data for monitoring progress toward the FP2020 goal, with more-rapid turnaround of results than the DHS and most other surveys. Stakeholders recognized and acknowledged the value of annual family planning data, and a majority of in-country stakeholders noted that PMA2020's comparability to the DHS (described as its estimates being "aligned with" those in the DHS) is one of its key strengths.

**PMA2020's survey design poses challenges to using its data.** Several U.S.-based stakeholders commented on comparisons of estimates among the PMA2020, DHS, and MICS surveys. Most of them felt that PMA2020 should complement rather than substitute for or compete with the other surveys, suggesting that earlier successful efforts to make (newer) MICS more consistent with the (existing) DHS might serve as a model for reconciling inconsistencies between PMA2020 with other surveys as well. One stakeholder noted that PMA2020 could serve as a "useful annual triangulator" against other family planning data sources. NGO representatives from several countries expressed interest in redesigning PMA2020 as a panel survey that could track progress over time (rather than inadvertently resampling households).

In-country stakeholders noted that PMA2020 sometimes undersamples regions with the greatest need for family planning services or only assesses part of the country. These stakeholders had several technical suggestions to improve the usability of PMA2020 for further analyses by those directly working with the data (typically, researchers external or internal to the country, or perhaps statisticians within MOHs or Track20 M&E officers). These suggestions include standardizing the variables across country surveys, improving the provision of technical support, improving the data documentation that accompanies the PMA2020 surveys, and troubleshooting technical issues with the download button for the DataLab tool (which is a web-based data visualization tool that allows users to build customized charts and download up-to-date PMA2020 data).

Chapter Eight discusses other approaches related to sampling design through which PMA2020 surveys could be made more useful to decisionmakers, and Chapter Eleven discusses the facilitators of and barriers to use of PMA2020 data in more detail.

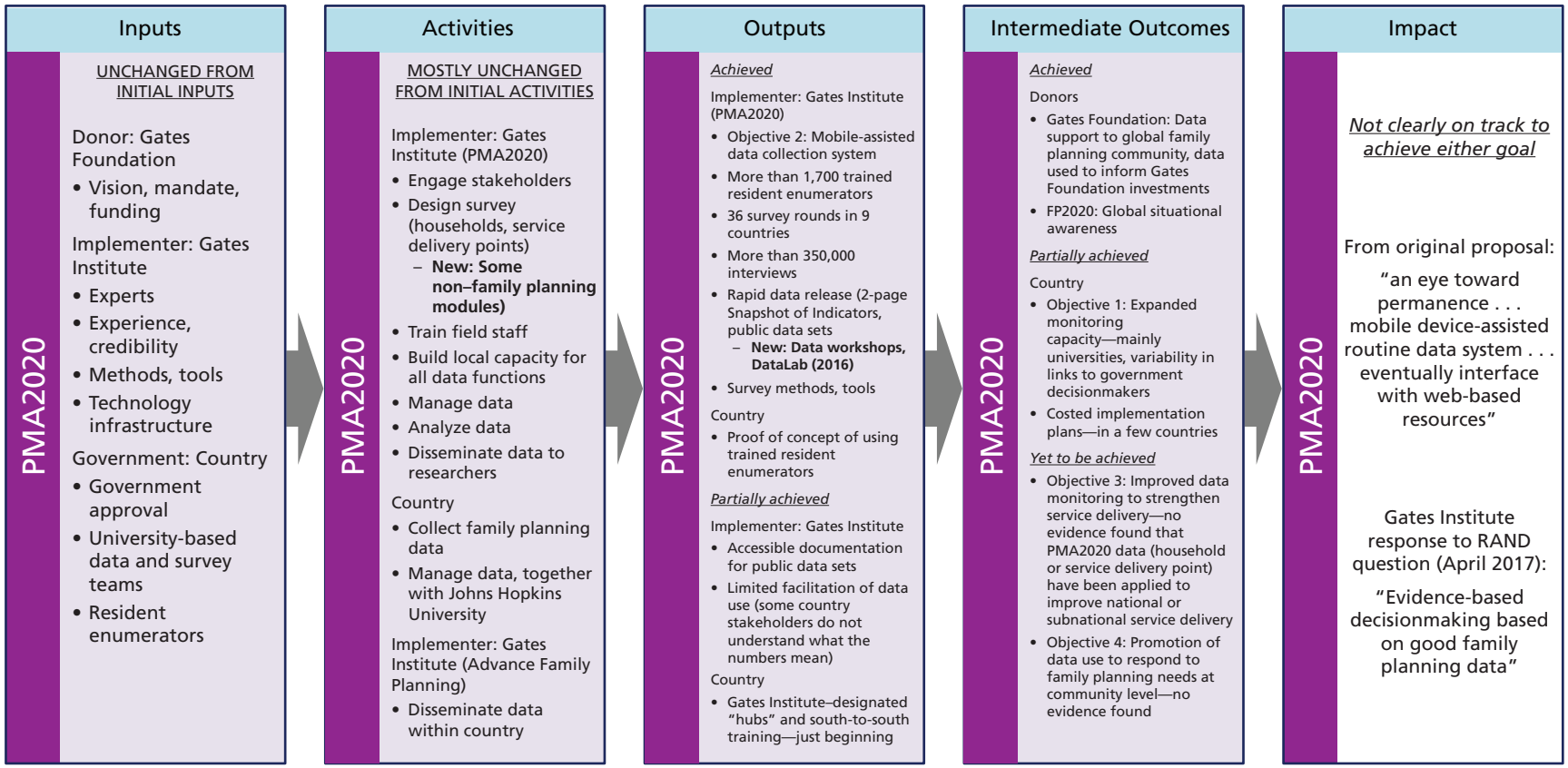
## Summary

Interviews with key stakeholders in the United States and program countries suggest that the PMA2020 program has laid an important foundation for collecting high-quality family plan-

ning data in 11 countries to date, with greater frequency and faster turnaround than previously available through other comparable surveys. The Gates Institute has carried out many of its proposed activities, including integrating a rapid data collection system using mobile devices; providing more frequent, more in-depth data to the global family planning community; and enhancing global awareness of the goals of FP2020.

Nonetheless, PMA2020 falls short of achieving some of its original objectives: building monitoring capacity (outside of academia), fully harnessing service delivery point surveys, and enhancing evidence-based decisionmaking at the community level based on good family planning data. It also has yet to achieve several of its 2015 objectives: building a sustainable business model (including financial sustainability of PMA2020 surveys) and integrating PMA2020 into countries' M&E architecture. These gaps are highlighted in the empiric logic model that was informed by views from 200 stakeholders from program countries, the Gates Foundation, the Gates Institute, and other U.S.-based subject-matter experts (Figure 6.1). The desired outputs and outcomes that have not yet been achieved, or have been partially achieved, reflect opportunities for future actions, and we address these issues in our conclusions and recommendations (Chapter Fourteen). This evaluation thus provides an opportunity to reassess objectives, to acknowledge the notable achievements of PMA2020 to date, and to build on those achievements to achieve clearly defined goals for the future.

**Figure 6.1**
**Empiric PMA2020 Logic Model**




RAND *RR2112-6.1*

CHAPTER SEVEN

# Statistical Properties of PMA2020

## Overview

The previous chapter described the goals of, accomplishments of, and challenges for PMA2020. In this chapter, we describe our findings of the analyses of statistical properties of the PMA2020 household survey, focusing on five countries. These quantitative analyses will be integrated with the qualitative perspectives of stakeholders described in the next chapter to inform potential design changes to the survey. As described in detail in Chapter Three, we explored representativeness of the data, including comparisons to a comparable survey and across rounds of this survey. We estimated margin of error for key indicators; examined ICC, which reflects the homogeneity within groups that are surveyed (e.g., enumeration areas) in a clustered design such as that used by PMA2020; and explored the impact of changing the frequency of data collection on detection of changes in rates of key indicators. In this chapter, we focus on the overall results across the five countries we examined in detail. PMA2020's methodology is described further in Appendix A, and additional detail is found in Appendix H, including selected methods and results from the country-specific analyses.

mCPR has been the key indicator of interest for PMA2020. It is the indicator around which the chosen country-level sample size of the survey has been based, and it is reported in PMA2020 materials. However, as will be addressed in the next chapter with regard to stakeholder views, there is concern that this indicator tells only part of the story of a woman's contraceptive use because the rate of use of modern contraception, which often changes over one's lifetime, may reflect desire to prevent pregnancy, desire to space pregnancies, a lack of information about different contraceptive methods, a lack of access to family planning services, or another reason. Therefore, we included several indicators other than mCPR in this analysis, including overall contraceptive prevalence, unmet need, and specific methods within the method mix.

As described previously, PMA2020 currently is conducted in ten countries and has recently launched in an eleventh. Even though there are considerable similarities in the survey from country to country, there are also substantial differences. Some have only been conducted in portions of the country, and the number of rounds varies from country to country. For this analysis, we selected five countries that vary in population size and have multiple rounds of data available: Ghana, Ethiopia, DRC, Nigeria, and Kenya (Table 7.1). Ghana, Nigeria, Ethiopia, and Kenya all had conducted at least three rounds of surveys for which data were available for our analysis. DRC is the only country for which there were five rounds of data available as of July 2017.

**Table 7.1**
**Countries Included in the Statistical Analyses**

| Country | Number of Rounds of Surveys with Data Available as of July 2017 |
|---|---|
| Ghana | 4 |
| Ethiopia | 4 |
| DRC | 5 |
| Nigeria | 3 |
| Kenya | 4 |

Sampling frames used for the PMA2020 in different countries were based on available census data for the countries. The survey sampling frame will be more representative of the population if the country has a fairly recent census to use as the basis for drawing the survey sample. For two of these countries, data were only available for subsets of the country: In DRC, this analysis includes only Kinshasa, and in Nigeria, the data include the states of Kaduna and Lagos.

Below, we briefly summarize our country-specific findings and then synthesize our findings across the five countries.

## Summary of Country-Specific Analyses

### Ghana

The data from Ghana show that there were some statistically significant differences in the demographic and socioeconomic characteristics of the population surveyed in PMA2020 when compared with the DHS, even though the absolute differences were very small. At the same time, from round to round, similar populations were sampled, with just a few demographic and socioeconomic characteristics differing in some rounds. Overall, PMA2020 surveys resulted in very small margins of error for mCPR, which is the main indicator of interest, and small margins of error for the other indicators we examined. Correlation within clusters led to large design effects for many indicators, meaning that the effective sample size was decreased. At the same time, the margin of error that can result from the design effect was still very small. There was also evidence that adopting more-efficient sample designs, such as optimal design (i.e., sampling fewer respondents in enumeration areas that have low variance; this is described in more detail in Chapter Three and Appendix H), can improve the precision of the different indicators and lead to a greater ability to detect differences between different groups in the survey. When we examined the rates of change for key family planning indicators over defined time intervals (six, 12, 18, and 24 months), we found that there were statistically significant changes in the indicators every six months. Therefore, conducting the PMA2020 survey every year could miss some of the rapid changes that are happening in the population related to family planning.

### Ethiopia

The population sampled in Ethiopia was significantly different from the DHS sample, though the DHS sample was from 2011, three years before the first round of PMA2020 in the country. Major differences include rate of electricity (PMA2020's sample had 20 percentage points more access to electricity), mobile phone use (29 percentage points higher in the PMA2020 sample), and socioeconomic status (PMA2020's sample was less likely to be in the lowest quintile and more likely to be in the highest, perhaps explaining the differences in electricity and mobile phone use). Round to round, the demographic differences were small and often not statistically significant; age was stable, as was the percentage married. There was some fluctuation in rates of access to electricity and mobile phones, which was statistically significant, but the absolute differences were small. Margins of error for all indicators in Ethiopia were very small—1.0 percent or less for all indicators in all rounds. Design effects were between 3 and 4 for mCPR. There was limited potential for improvement in precision using optimal design. Few significant differences were seen in indicators between six-month rounds; one-year differences were significant for mCPR and any contraceptive user. Some specific methods, like use of male condoms or the lactational amenorrhea method (LAM), did not significantly change even between Round 1 and Round 4 (i.e., over an 18-month interval), though response rates for specific methods were less than 30 percent, compared with the 99-percent response rates for the major indicators.

### DRC

In the Kinshasa region of DRC, the DHS population was more similar to PMA2020's Round 2 data than Round 1's, though Round 1 overlapped with the DHS's data collection period. For most demographic measures we examined, differences were statistically significant and often large, with PMA2020 having a much lower reported rate of access to electricity, television, and mobile phones. Round-to-round differences within PMA2020 data were small and rarely significant except for a decrease in access to electricity of 13 percentage points in Round 3, but access to electricity increased somewhat in Rounds 4 and 5. Margins of error were small—2 percent or less for all major indicators we analyzed. Margins of error were very small even for specific contraceptive methods in all rounds except Round 1, when a few were greater than 3 percent (male condoms and the rhythm method). Design effect was smallest in the first round and then increased; design effect for mCPR was greater than 3 for all rounds after the first. The potential gain in precision for mCPR with optimal design was still limited after Round 1. Changes from round to round for major indicators were significant from Round 3 to Round 4; otherwise, changes over six-month intervals were mostly not statistically significant, but changes over one-year and longer intervals were. DRC had five rounds of data available, so we were able to examine two-year differences, which were significant for all major indicators; however, the enumeration areas did not change, so these data do not offer insight into the impact of "refreshing" the sample—in other words, moving to an adjacent enumeration area for the subsequent round of data collection.

### Nigeria

The demographic and socioeconomic characteristics of the sample in the PMA2020 survey for Nigeria closest to the most recent DHS are quite dissimilar from the DHS data, with the DHS population being more wealthy, less likely to be married, and more likely to have received more education. When comparing these characteristics from round to round, however, they were

relatively similar. The estimates based on PMA2020 data from Nigeria show small margins of error. Findings suggest potential improvement in efficiency with optimal design. Differences in rates of several key indicators between the first two rounds of PMA2020 (which were separated by a six-month interval) were significant, but there was little change in the six-month period from Round 2 to Round 3. Because PMA2020 was only conducted in a few regions of Nigeria in the first few rounds, analyses were limited. Analyses of the national-level data that are now available and of data from the recently collected Round 4 will be helpful for making informed decisions about Nigeria's PMA2020 data moving forward.

#### Kenya

Overall, the data from Kenya's PMA2020 surveys raise questions about representativeness compared with DHS data, as well as about round-to-round consistency in some observed demographic and socioeconomic characteristics. However, like with other countries, small margins of error were found for key family planning indicators; unlike other countries, the design effects in Kenya's PMA2020 surveys were small, suggesting little need for Kenya to change its sampling approach to achieve a larger effective sample size. Our analysis of changes in indicators from round to round suggests that collecting annual data may be sufficient to detect statistically significant changes in mCPR and other indicators.

## Synthesis of Findings Across the Five Countries

In some of the five countries we considered, we found differences in the demographic and socioeconomic characteristics of the PMA2020 survey populations compared with the same characteristics in the most-contemporaneous DHS data. For example, as noted above, in Ethiopia, the rate of electricity and mobile phone access were significantly different, with much higher rates of access in PMA2020 surveys; and in DRC, for which we only used DHS data from Kinshasa to align with the PMA2020 survey, access to electricity and socioeconomic status were also different, with the PMA2020 population having less access and lower socioeconomic status. However, these differences may not lead to biases in the family planning indicators, especially in cases where the characteristics for which we observed such differences are not confounded with the indicators. That is, observed differences in the population characteristics may not necessarily mean that they will have different rates of contraceptive use or of other indicators. However, where differences are seen in family planning indicators between the DHS and PMA2020, those differences could be explained by some of the population differences. For example, a higher percentage of married women might explain a higher rate of contraception use; a better educated population could also explain more modern contraceptive use. Observed differences in population characteristics are not necessarily straightforward to explain. They could be due to a number of factors: changes in the population between the time of the DHS and the PMA2020 surveys (though we attempted to compare interviews that were conducted as contemporaneously as possible), interventions targeting one of these measures (such as mobile phone access or education) that affected the DHS or PMA2020 sample differently, differential nonresponse or the way questions were understood or asked or other problems with data, or many other reasons. Our results can only suggest that in these countries, the populations that were sampled differed in some significant ways.

The populations sampled by PMA2020 from round to round were relatively similar in most cases. There were some differences between some rounds in specific countries, and in some cases these differences could reflect the effects of economic development and modernization—e.g., more penetration over time of mobile phones. But overall, we found that the surveyed population was relatively constant on these observed characteristics. This is an indication of comparable populations within countries on which the samples were collected, which suggests that changes in key family planning indicators cannot be attributed to the possibility of changes in the sample population. However, not all characteristics of the population can be measured, and these inferences of similarity are only applicable to measured characteristics. Unmeasured characteristics can still explain observed differences.

We explored the design effect of indicators of interest in all five countries. The design effects were large in many cases; in some cases, they were greater than 5.0 (Table 7.2). These results imply that the effective sample size is much smaller than the observed sample size because of the correlation within clusters. That is, even if a sample size seems large, when correlation within enumeration area and, thus, design effect is high, the effective number of respondents is much lower, meaning that resources are being used to interview additional people who are not adding meaningful data. One way to avoid the issue of the large design effect would be to select more enumeration areas and sample fewer respondents per enumeration area, producing the same total sample size but increasing the *effective* sample size.

Despite some large design effects, the margins of error for mCPR and for other key indicators considered were quite small—generally less than the 3-percent target and usually less than 2 percent. Across countries, rounds, and indicators, margins of error were typically quite small for calculations performed on the full survey sample. Of course, estimates obtained for subsamples—e.g., for regions of the country—will be based on a smaller number of cases and will have larger margins of error.

At the same time, the potential improvement in efficiency by implementing optimal design indicated that in a number of cases, particularly in Nigeria, there was potential benefit to decreasing the number of observations in certain enumeration areas to improve efficiency. However, this result should be weighed against practical issues that can arise when trying to assign different numbers of observations for different enumeration areas. To implement differ-

**Table 7.2**
**Design Effect and Optimal Design for mCPR, by Round**

| | Round 1 | | | Round 2 | | | Round 3 | | | Round 4 | | | Round 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain |
| Ghana | 0.065 | 3.36 | 78.9% | 0.071 | 3.69 | 74.1% | 0.099 | 5.46 | 84.0% | 0.081 | 5.14 | 88.5% | | | |
| Ethiopia | 0.088 | 3.83 | 91.9% | 0.077 | 3.46 | 92.1% | 0.087 | 3.89 | 93.3% | 0.091 | 4.00 | 92.3% | | | |
| DRC | 0.007 | 1.24 | 56.2% | 0.057 | 3.72 | 94.9% | 0.084 | 4.78 | 93.0% | 0.055 | 3.53 | 92.4% | 0.047 | 3.07 | 91.7% |
| Kenya | 0.053 | 2.59 | 91.3% | 0.065 | 3.28 | 91.2% | 0.062 | 3.25 | 92.9% | 0.069 | 3.75 | 91.5% | | | |
| Nigeria | 0.086 | 3.65 | 72.9% | 0.100 | 4.61 | 80.1% | 0.111 | 5.00 | 78.5% | | | | | | |

ent numbers per enumeration area, a varying number of households in each enumeration area would have to be conveyed to resident enumerators who may be covering more than one enumeration area. There is a trade-off between the benefits of optimal design and the complexity of implementing it.

Lastly, we explored the changes in indicators over different time intervals to draw conclusions about the optimal frequency of the surveys. In most cases, changes in mCPR over six-month intervals were not statistically significant; for most countries, only one of the six-month intervals showed statistical significance. Countries did differ in this regard, with Ghana demonstrating statistically significant round-to-round changes in two out of three six-month intervals available. But in general, our results suggest no consistently significant changes in mCPR or other major indicators over the six-month intervals. We looked at these differences separately for urban and rural settings to examine whether the ideal frequency differed by setting, but we did not find different results based on setting (results not shown). When we explored changes over intervals of a year or longer, we did find consistently statistically significant differences. Thus, our findings suggest utility in repeating the survey every year, but not necessarily more frequently than that. We explored six-month, yearly, and less frequent intervals (Table 7.3) to reflect what is being done in the field and what is available in the data. We also report the significance of changes across all rounds for a number of indicators of interest—i.e., the $p$ value of an F-test for each indicator (Table 7.4)—demonstrating that almost all indicators differ significantly across the rounds of data that are currently available. Of note, the samples we compare in our analyses are of the same enumeration areas for each round. This should lead to less random change over time than if each round had sampled a new enumeration area. Thus, it is likely that our findings reflect real changes in the indicators over time.

PMA2020 could experiment with other intervals (e.g., nine months) for the household surveys to see where efficiency and statistical significance meet. As we recommend elsewhere, asking some family planning questions less often does not necessarily mean that the PMA2020 household survey would be conducted only on a yearly basis; other questions could be asked between annual rounds. Options for alternative designs are discussed in Chapter Eight.

## Summary

Analysis of the PMA2020 data offers insight into how well the survey is meeting its goals. While each of the PMA2020 countries is different, and, thus, it is impossible to extrapolate the results from the countries we examined to all other program countries, our analyses of five different countries with multiple rounds of data offers several key findings for consideration.

Our results raise questions about population representativeness of PMA2020 surveys in the five selected countries (as inferred by comparison to DHS data), though these population differences will not necessarily impact indicator estimates. Analyses suggest that margins of error for major indicators are small and meet the program's stated goals for mCPR, with similarly small margins of error for other estimates. ICCs and design effects, however, are large in some countries for some indicators, which could be addressed by increasing the number of enumeration areas selected and decreasing the number of households sampled within those enumeration areas or clusters. In many cases, precision could be improved by using optimal design, which would sample a different number of households in different enumeration areas depending on the local variance of an indicator. Lastly, findings regarding frequency of data

**Table 7.3**
**Between-Round Differences in mCPR for Different Time Intervals**

| | 6-Month Differences | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Difference Between Rounds 1 and 2 | | Difference Between Rounds 2 and 3 | | Difference Between Rounds 3 and 4 | | Difference Between Rounds 4 and 5 | |
| | Difference | *P* Value | Difference | *P* Value | Difference | *P* Value | Difference | *P* Value |
| DRC: Kinshasa | –0.28 (–4.28, 3.71) | 1.000 | 0.94 (–3.10, 4.98) | 0.969 | 4.03 (1.27, 6.78) | 0.001 | –0.05 (–2.78, 2.67) | 1.000 |
| Ethiopia | 1.25 (–1.60, 4.10) | 0.673 | 2.48 (0.74, 4.21) | 0.001 | 0.20 (–1.53, 1.93) | 0.991 | | |
| Ghana | 0.42 (–1.69, 2.52) | 0.957 | 3.40 (1.29, 5.52) | 0.000 | 5.36 (3.27, 7.44) | 0.000 | | |
| Kenya | –1.26 (–4.00, 1.48) | 0.637 | 6.25 (3.53, 8.97) | 0.000 | –0.55 (–3.27, 2.17) | 0.954 | | |
| Nigeria: Kaduna and Lagos | 5.38 (3.95, 6.81) | 0.000 | 0.19 (–1.64, 2.01) | 0.969 | | | | |

| | 12-Month Differences | | | | | |
|---|---|---|---|---|---|---|
| | Difference Between Rounds 1 and 3 | | Difference Between Rounds 2 and 4 | | Difference Between Rounds 3 and 5 | |
| | Difference | *P* Value | Difference | *P* Value | Difference | *P* Value |
| DRC: Kinshasa | 0.66 (–2.07, 3.39) | 0.965 | 4.97 (0.95, 8.98) | 0.007 | 3.97 (1.21, 6.73) | 0.001 |
| Ethiopia | 3.73 (0.84, 6.62) | 0.005 | 2.68 (1.01, 4.35) | 0.000 | | |
| Ghana | 3.82 (1.73, 5.91) | 0.000 | 8.76 (6.66, 10.86) | 0.000 | | |
| Kenya | 4.99 (2.25, 7.72) | 0.000 | 5.70 (2.97, 8.42) | 0.000 | | |
| Nigeria: Kaduna and Lagos | 5.57 (3.77, 7.37) | 0.000 | | | | |

| | 18-Month Differences | | | |
|---|---|---|---|---|
| | Difference Between Rounds 1 and 4 | | Difference Between Rounds 2 and 5 | |
| | Difference | *P* Value | Difference | *P* Value |
| DRC: Kinshasa | 4.68 (1.99, 7.38) | 0.000 | 4.91 (0.89, 8.93) | 0.008 |
| Ethiopia | 3.93 (1.08, 6.78) | 0.002 | | |
| Ghana | 9.18 (7.09, 11.26) | 0.000 | | |
| Kenya | 4.43 (1.70, 7.17) | 0.000 | | |

| | Two-Year Differences | |
|---|---|---|
| | Difference Between Rounds 1 and 5 | |
| | Difference | *P* Value |
| DRC: Kinshasa | 4.63 (1.93, 7.33) | 0.000 |

NOTES: Differences between rounds are reported with 95-percent confidence intervals and *p* values. *P* values less than 0.05 are shown in pink for statistical significance. *P* values were calculated using Tukey-Kramer adjustment to account for multiple comparisons.

**Table 7.4**
**Indicator Estimates and Significance of Changes Across Rounds**

| | Ghana | | | Ethiopia | | | DRC: Kinshasa | | | Nigeria: Kaduna and Lagos | | | Kenya | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Indicator** | **First Round** | **Most Recent Round** | ***P* Value for Differences (All Rounds)** | **First Round** | **Most Recent Round** | ***P* Value for Differences (All Rounds)** | **First Round** | **Most Recent Round** | ***P* Value for Differences (All Rounds)** | **First Round** | **Most Recent Round** | ***P* Value for Differences (All Rounds)** | **First Round** | **Most Recent Round** | ***P* Value for Differences (All Rounds)** |
| **Round** | **1** | **4** | | **1** | **4** | | **1** | **5** | | **1** | **3** | | **1** | **4** | |
| Contraceptive user: any | 15.38 (14.20, 16.56) | 28.07 (26.89, 29.25) | 0.000 | 22.83 (20.83, 24.83) | 27.84 (26.93, 28.75) | 0.000 | 33.26 (31.57, 34.94) | 42.28 (40.55, 44.00) | 0.000 | 10.89 (9.77, 12.01) | 19.02 (17.88, 20.16) | 0.000 | 41.73 (40.24, 43.22) | 47.72 (46.24, 49.19) | 0.000 |
| Contraceptive user: modern | 14.23 (13.11, 15.36) | 23.41 (22.29, 24.53) | 0.000 | 22.57 (20.59, 24.55) | 26.50 (25.61, 27.40) | 0.000 | 16.26 (14.91, 17.61) | 20.89 (19.50, 22.28) | 0.000 | 10.37 (9.31, 11.43) | 15.94 (14.87, 17.01) | 0.000 | 41.55 (40.06, 43.03) | 45.98 (44.52, 47.45) | 0.000 |
| Contraceptive user: traditional | 1.15 (0.67, 1.63) | 4.66 (4.19, 5.14) | 0.000 | 0.25 (–0.16, 0.67) | 1.34 (1.15, 1.53) | 0.000 | 17.00 (15.63, 18.36) | 21.39 (19.99, 22.79) | 0.000 | 0.52 (0.07, 0.97) | 3.08 (2.63, 3.54) | 0.000 | 0.18 (–0.13, 0.49) | 1.81 (1.50, 2.11) | 0.000 |
| Total unmet need (spacing and limiting) | 27.46 (26.20, 28.73) | 21.60 (20.33, 22.88) | 0.000 | 16.08 (14.40, 17.75) | 16.24 (15.48, 17.00) | 0.868 | 17.72 (16.40, 19.04) | 14.57 (13.21, 15.93) | 0.000 | 26.13 (24.83, 27.43) | 22.57 (21.25, 23.89) | 0.000 | 18.75 (17.65, 19.84) | 13.59 (12.51, 14.67) | 0.000 |
| Current use of pill | 4.30 (3.42, 5.18) | 15.50 (13.81, 17.19) | 0.000 | 7.72 (5.33, 10.12) | 8.67 (7.68, 9.65) | 0.000 | 6.46 (5.44, 7.48) | 3.11 (2.50, 3.71) | 0.000 | 22.12 (20.43, 23.80) | 2.80 (2.24, 3.36) | 0.000 | 5.13 (4.32, 5.94) | 5.74 (4.94, 6.55) | 0.000 |
| Current use of emergency contraception | 0.35 (–0.11, 0.80) | 6.39 (5.51, 7.26) | 0.000 | 0.10 (–0.57, 0.77) | 0.61 (0.33, 0.89) | 0.105 | 2.02 (1.47, 2.58) | 1.57 (1.24, 1.90) | 0.000 | 0.88 (0.24, 1.52) | 0.61 (0.39, 0.82) | 0.180 | 0.23 (–0.04, 0.50) | 1.06 (0.79, 1.33) | 0.000 |
| Current use of female condom | 0.15 (–0.06, 0.36) | 1.09 (0.68, 1.49) | 0.000 | 0.10 (–0.12, 0.33) | 0.05 (–0.04, 0.14) | 0.968 | 0.15 (–0.09, 0.40) | 0.25 (0.11, 0.40) | 0.763 | 2.31 (1.76, 2.85) | 0.13 (–0.05, 0.31) | 0.000 | 0.17 (0.05, 0.29) | 0.17 (0.05, 0.29) | 0.927 |
| Current use of female sterilization | 0.28 (–0.02, 0.57) | 1.83 (1.27, 2.39) | 0.000 | 0.82 (–0.12, 1.76) | 0.71 (0.32, 1.09) | 0.109 | 0.87 (0.50, 1.24) | 0.35 (0.13, 0.57) | 0.087 | 0.42 (–0.16, 1.00) | 0.33 (0.13, 0.52) | 0.396 | 1.41 (0.93, 1.89) | 2.12 (1.65, 2.60) | 0.000 |
| Current use of implant | 2.06 (1.31, 2.81) | 15.09 (13.65, 16.54) | 0.028 | 16.35 (12.52, 20.17) | 22.96 (21.38, 24.54) | 0.001 | 2.74 (1.67, 3.81) | 4.25 (3.61, 4.88) | 0.000 | 10.88 (9.28, 12.49) | 3.14 (2.60, 3.67) | 0.000 | 7.45 (6.47, 8.43) | 11.49 (10.51, 12.46) | 0.000 |

**Table 7.4—continued**

| Indicator | Ghana | | | Ethiopia | | | DRC: Kinshasa | | | Nigeria: Kaduna and Lagos | | | Kenya | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) |
| **Round** | **1** | **4** | | **1** | **4** | | **1** | **5** | | **1** | **3** | | **1** | **4** | |
| Current use of injectables (1-month forms if differentiated between 1- and 3-month versions) | 5.22 (4.59, 5.85) | 3.64 (2.44, 4.84) | 0.000 | 71.35 (66.87, 75.83) | 61.88 (60.03, 63.72) | 0.000 | 8.56 (7.49, 9.63) | 2.20 (1.57, 2.83) | 0.000 | 41.31 (39.06, 43.55) | 5.20 (4.45, 5.94) | 0.000 | 23.65 (22.32, 24.98) | 21.51 (20.19, 22.83) | 0.000 |
| Current use of IUD | 0.42 (0.11, 0.72) | 1.96 (1.38, 2.54) | 0.000 | 2.17 (0.79, 3.56) | 2.21 (1.63, 2.78) | 0.947 | 0.67 (0.30, 1.03) | 0.50 (0.29, 0.72) | 0.116 | 4.07 (3.18, 4.96) | 0.61 (0.31, 0.90) | 0.000 | 2.05 (1.53, 2.57) | 2.69 (2.17, 3.20) | 0.000 |
| Current use of LAM | 0.04 (–0.24, 0.31) | 1.76 (1.23, 2.29) | 0.000 | 0.26 (–0.54, 1.06) | 0.64 (0.31, 0.97) | 0.121 | 1.42 (0.93, 1.90) | 0.31 (0.03, 0.60) | 0.000 | 1.18 (0.58, 1.78) | 0.23 (0.03, 0.43) | 0.002 | 0.49 (0.27, 0.71) | 0.31 (0.09, 0.53) | 0.001 |
| Current use of male condoms | 1.31 (0.63, 1.99) | 12.89 (11.59, 14.19) | 0.000 | 2.15 (0.93, 3.36) | 1.42 (0.92, 1.92) | 0.069 | 24.48 (22.69, 26.27) | 8.97 (7.92, 10.03) | 0.000 | 16.53 (14.72, 18.35) | 3.85 (3.25, 4.45) | 0.000 | 2.42 (1.78, 3.05) | 4.03 (3.40, 4.65) | 0.000 |
| Current use of male sterilization | 0.00 (–0.07, 0.07) | 0.00 (–0.07, 0.07) | 0.011 | 0.00 (0.00, 0.00) | 0.00 (0.00, 0.00) | | 0.00 (0.00, 0.00) | 0.00 (0.00, 0.00) | | 0.00 (–0.18, 0.18) | 0.03 (–0.02, 0.09) | 0.854 | 0.00 (–0.02, 0.02) | 0.00 (–0.02, 0.02) | 0.216 |
| Current use of other traditional method | 0.10 (–0.17, 0.36) | 1.62 (1.12, 2.12) | 0.000 | 0.01 (–0.22, 0.24) | 0.16 (0.07, 0.26) | 0.054 | 1.57 (0.63, 2.51) | 3.20 (2.65, 3.75) | 0.000 | 3.47 (2.72, 4.22) | 0.45 (0.20, 0.70) | 0.000 | 0.18 (0.04, 0.33) | 0.13 (–0.01, 0.28) | 0.007 |
| Current use of rhythm method | 0.84 (0.17, 1.51) | 15.29 (14.00, 16.58) | 0.000 | 1.18 (–0.41, 2.76) | 4.01 (3.36, 4.66) | 0.000 | 43.02 (40.78, 45.27) | 17.13 (15.80, 18.46) | 0.000 | 0.99 (0.04, 1.95) | 1.17 (0.85, 1.49) | 0.933 | 0.25 (–0.10, 0.60) | 1.66 (1.31, 2.00) | 0.000 |
| Current use of standard days method | 0.38 (0.08, 0.68) | 1.49 (0.91, 2.07) | 0.000 | 0.11 (–0.19, 0.41) | 0.13 (0.00, 0.25) | 0.395 | 0.00 (–0.32, 0.32) | 0.47 (0.28, 0.66) | 0.133 | 0.95 (0.54, 1.36) | 0.21 (0.08, 0.35) | 0.001 | 0.32 (0.11, 0.52) | 0.31 (0.10, 0.51) | 0.000 |
| Current use of withdrawal | 0.64 (0.06, 1.23) | 11.08 (9.95, 12.21) | 0.000 | 0.10 (–0.74, 0.94) | 0.99 (0.65, 1.34) | 0.241 | 6.48 (4.90, 8.05) | 8.50 (7.57, 9.43) | 0.000 | 4.87 (3.39, 6.35) | 2.86 (2.37, 3.35) | 0.012 | 0.18 (–0.07, 0.43) | 0.78 (0.53, 1.02) | 0.000 |

**Table 7.4—continued**

| Indicator | | Ghana | | | Ethiopia | | | DRC: Kinshasa | | | Nigeria: Kaduna and Lagos | | | Kenya | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) |
| **Round** | | **1** | **4** | | **1** | **4** | | **1** | **5** | | **1** | **3** | | **1** | **4** | |
| Unmet need (categories) | 1. Unmet need for spacing | 19.86% | 21.90% | 0.000 | 12.71% | 12.51% | 0.000 | 25.55% | 16.00% | 0.000 | 25.26% | 22.16% | 0.000 | 13.35% | 11.74% | 0.000 |
| | 2. Unmet need for limiting | 8.12% | 8.46% | | 7.67% | 7.00% | | 12.13% | 5.04% | | 8.11% | 7.18% | | 11.58% | 6.51% | |
| | 3. Using for spacing | 10.72% | 26.48% | | 25.33% | 30.01% | | 23.86% | 44.21% | | 9.55% | 14.74% | | 28.49% | 36.28% | |
| | 4. Using for limiting | 4.14% | 9.23% | | 14.87% | 13.55% | | 16.90% | 15.11% | | 4.83% | 8.64% | | 27.13% | 28.45% | |
| | 7. No unmet need | 12.50% | 19.38% | | 18.77% | 23.28% | | 10.83% | 11.30% | | 27.91% | 29.87% | | 10.98% | 10.09% | |
| | 9. Infecund or menopausal | 15.16% | 14.55% | | 20.64% | 13.64% | | 10.74% | 8.34% | | 24.33% | 17.42% | | 8.47% | 6.93% | |
| | 97. Not sexually active | 29.50% | 0.00% | | | | | | | | | | | | | |
| Have you ever heard of IUD: Collected only in later rounds | | | 49.55 (48.19, 50.91) | | | | | | 47.44 (45.54, 49.34) | 0.613 | | 42.36 (40.89, 43.84) | | | | |
| Current use of diaphragm | | 0.10 (0.00, 0.19) | 0.07 (–0.11, 0.25) | 0.773 | | | | | | | | | | | | |

**Table 7.4—continued**

| Indicator | Ghana | | | Ethiopia | | | DRC: Kinshasa | | | Nigeria: Kaduna and Lagos | | | Kenya | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) | First Round | Most Recent Round | *P* Value for Differences (All Rounds) |
| **Round** | **1** | **4** | | **1** | **4** | | **1** | **5** | | **1** | **3** | | **1** | **4** | |
| Current use of foam/ jelly | 0.00 (–0.07, 0.07) | 0.20 (0.06, 0.34) | 0.028 | | | | | | | | | | | | |
| Current use of Primolut N tablet (norethisterone) | | 4.09 (2.95, 5.23) | 0.712 | | | | | | | | | | | | |

NOTES: *P* values of less than 0.05, signifying statistical significance, are highlighted in pink. Confidence intervals are shown in parentheses. We report the value as well as the percentage of respondents who gave a response because the percentage of questions answered varied widely for the specific methods; when all women responded, the rate would reflect the rate of use in the population, but in cases with smaller response rates, the rate may reflect the percentage of contraceptive users using a certain method. P values were calculated using Tukey-Kramer adjustment to account for multiple comparisons.

collection show that, at least in most countries we explored, the intervals between rounds could be extended to 12 months without loss of information regarding changes in rates for key indicators. Waiting longer than 12 months, however, would mean a potential loss in information, as changes at the one-year intervals were almost always statistically significant. This could vary in other countries, depending on the beginning rate of mCPR and the intensity of interventions designed to increase modern contraceptive use. On the other hand, these findings are consistent with stakeholder views on the ideal survey frequency from the full sample of PMA2020 countries.

The next chapter presents potential changes to the survey, taking into account these findings and the views of stakeholders we interviewed.

CHAPTER EIGHT

# PMA2020 Survey Design

## Overview

In the previous chapter, we analyzed the statistical properties of PMA2020 surveys based on data from five countries with multiple survey rounds. In this chapter, to complement those quantitative findings, we describe stakeholder views of PMA2020's survey design. We then propose potential changes to PMA2020 more broadly, based on both the statistical analyses reported in Chapter Seven and qualitative input from in-country and U.S.-based stakeholders described here.

## Stakeholder Perceptions of PMA2020's Survey Design

U.S.-based and in-country stakeholders raised several questions about PMA2020's design and sampling methodology, including concerns about (1) sampling associated with repeated surveys, in particular returning to the same enumeration area and, in some cases, the same households from one round to another; (2) presenting changes over time versus point estimates; (3) presenting estimates as nationally representative; and (4) focusing on the mCPR indicator to monitor performance toward family planning goals.

### Challenges of Conducting Repeated Surveys

**Are repeated PMA2020 surveys truly cross-sectional?** PMA2020 is described as a two-stage cluster sample, but several stakeholders commented that starting with the second round of data collection, resident enumerators return to the same communities. In Round 1, they are interviewing different households, but increasingly they will be conducting repeat interviews, making PMA2020 conceptually more like a panel. Resampling from adjacent enumeration areas is intended after four rounds, making the design of rounds following the first survey somewhat of a hybrid between a longitudinal panel and a cross-sectional survey, rather than a set of truly independent serial cross-sectional surveys. Because the resampling reflects a hybrid, it is difficult to know how to analyze these data.

U.S.-based stakeholders also expressed concern about bias and representativeness associated with resampling from the same or adjacent enumeration areas. One U.S.-based demographer acknowledged the desire to retain the trained resident enumerators but said that there needs to be a solution to changing or rotating areas "without creating bias." Another stakeholder said that this is the reason PMA2020 is not cross-sectional—because if a woman is not

in the enumeration area in the first round, there is no chance of her being in the sample at any subsequent point (unless in an adjacent enumeration area in Round 5).

Such issues have introduced confusion about whether PMA2020 is meant to be a panel by design or whether the repeated sampling of the same households is problematic (e.g., it may introduce bias such as the Hawthorne effect, which is described in more detail later in this section). This confusion is illustrated by in-country perspectives on the PMA2020 survey.

Respondents in both Ethiopia and Côte d'Ivoire praised PMA2020 as being "longitudinal" because "we will be able to follow the evolution of people in the same population," a view that suggests a misunderstanding or mischaracterization of the survey design. Perhaps because of the repeated sampling of the same areas and sometimes the same households, we found that a number of respondents considered PMA2020 to be a panel survey, using exactly that language and praising the "interesting angle and information" gleaned from that design. In contrast, other respondents recognized that repeated sampling in the same enumeration areas is problematic specifically because PMA2020 is not intended to resample the same households.

**What is the effect of returning to some of the same households?** A related issue, which increases in salience with each round, is a concern regarding an intervention effect. The number of women who have already been surveyed increases in each round, reaching as high as roughly 30 percent by Round 4 in the surveys conducted so far. As one stakeholder put it, it is possible that "there is knowledge already in the area and contraceptive use increases in the area as a result of the survey." Another stakeholder framed this as a "Hawthorne effect"—as people become more and more familiar with the survey, the act of measuring their behaviors actually influences those behaviors (or, at least, their responses). Thus, the measured rate of contraceptive prevalence, for example, may not be the rate in other locations that were not already exposed to the survey multiple times. The later surveys may, in fact, be measuring the impact of the initial surveys and not the prevalence it would have found otherwise. In addition to this intervention effect, resampled women may give answers in a way that is systematically different from new women.

The issue of resampling surfaced in discussions of the current sampling strategy as well. There were concerns about how resampling affected responses, including concerns about fatigue among respondents and concerns about response accuracy related to expectations or social desirability. One research coordinator worried that fatigue led people to refuse to answer the survey. Multiple coordinators suggested changing enumeration areas more often to avoid "survey fatigue" or so as to not "tire the same people." However, despite these perceptions, a paper published by the Gates Institute provides evidence refuting this concern: Household and female response rates were nearly uniformly high across all rounds, and refusal rates were low (Zimmerman et al., 2017).

**PMA2020 surveys have overlapping populations in their cross-sectional surveys.** Another critique from stakeholders is that the PMA2020 survey presents point estimates as if it were a completely independent cross-sectional survey. Sampling from the same enumeration areas for the first several survey rounds was a motivation for the quick repeated surveys, but this results in high numbers of women who are resampled over and over. Furthermore, many of the interviewees noted that the current design and analysis do not focus on measuring and tracking trends.

One stakeholder described PMA2020 as a "survey within a demographic surveillance site," rather than a true probability sample. Results show "more about the cluster and less about how it represents the country," although many acknowledged that local data may be what

are actually needed by in-country data users. One respondent suggested that PMA2020 was originally designed to conduct sentinel surveillance—i.e., sampling from the same enumeration areas although from different households with each enumeration area—which is a new approach in family planning. Such an approach aims to observe trends within the same area rather than to generate point estimates intended to represent the entire underlying population. However, when PMA2020 began to emphasize that it was a cross-sectional survey (i.e., intended to generate serial point estimates), the interpretation of the indicators became confusing for users. From this stakeholder's perspective, PMA2020 "should never be putting out point estimates," though the confusion may have to do with the population included in the estimates. As long as it is clear what population is being sampled, PMA2020 estimates can be considered truly cross-sectional from a statistical perspective.

While PMA2020 produces numerous reports and briefs, those reports focus on point estimates over time. At least one individual commented that PMA2020 should be thinking about more comprehensive modeling and statistical analyses, such as analyses of trends, associations, and subpopulations, not just the simple statistical tests that it currently conducts, in order to increase its rigor (and, indeed, some new work on analyzing the data in more detail is being carried out through various contracts within the Gates Foundation family planning portfolio).

### National Representativeness

**Are PMA2020 surveys truly nationally representative?** Both in-country and U.S.-based respondents challenged the national representativeness of the PMA2020 estimates. While few of the PMA2020 countries purport to be sampling nationally and producing national estimates (as opposed to sampling within a few states), there is a perception surrounding PMA2020 that its goal is to produce national-level data. For example, stakeholders from Burkina Faso, Ethiopia, Kenya, and Côte d'Ivoire all said that results should be representative at the national level: As one put it, "It's true [that] with the calculations we do the results will be nationally representative," and the existing sampling technique was praised as ideal for this goal.

Other stakeholders whom we interviewed, however, expressed concern about the estimates or stated outright that they felt that the estimates collected from a single state or a few regions of the country could not be nationally representative. Some of these concerns came from stakeholders in the same countries where others praised the national estimates, highlighting wide variation in expectations from PMA2020 data.

This tension is inherent in PMA2020's publicly facing materials; in many cases, PMA2020 markets itself as providing national estimates, such as on its methodology web page: "PMA2020's primary aim is to collect a nationally representative sample of data from households and service delivery points in selected sentinel sites . . . . A nationally representative number of clusters or enumeration areas will be sampled in each program country" (PMA2020, 2017b).

**Subnational surveys may be PMA2020's niche.** As discussed in Chapter Five, in-country stakeholders overwhelmingly agreed that for local governments, subnational data are critical. Thus, the fact that PMA2020 is not nationally representative (i.e., is only operating in certain parts of some countries) might actually be an opportunity: "Since countries want subnational data, perhaps PMA should abandon the goal of data for national estimates," particularly if the DHS is providing national estimates. They offered various suggestions: Once the DHS provides national estimates, PMA2020 could follow up with a targeted survey of

specific regions to identify gaps that the DHS was not able to pinpoint. To collect subnational data, PMA2020 might consider surveying part of the country every year and only survey the entire country once every three years or so. The members of the External Consultative Group concluded in April 2017 that

> A promising approach to balance costs and country needs for subnational data for decision making would be to focus on strategic subnational data collection in areas of top priority for country governments as well as where change is anticipated to be measured, such as where large-scale interventions are focused (PMA2020, 2017e).

These approaches would solve both the problem of confusion among data users about whether PMA2020 data are nationally representative and the problem of direct competition with the DHS.

#### A Focus on mCPR

Given the important role of mCPR in calculating new users for FP2020 goals, mCPR plays a central role in PMA2020: It is the indicator on which all sample sizes are based, and it is the indicator that is reported in most detail in the two-page briefs. However, in the RAND team's conversations with U.S.-based stakeholders, the mCPR indicator was criticized as a "blunt tool." For example, it does not by itself distinguish between those using contraception to limit pregnancies and those using it to space pregnancies, which are different purposes. To fully understand this indicator, additional context is required about the individual, such as whether she is married, sexually active, or desiring more children. PMA2020 explores some of these subgroups in its analyses, but the overall focus is on mCPR. U.S.-based stakeholders felt that mCPR (a composite indicator of modern contraceptive use) is more informative when combined with information about method mix (i.e., the distribution of different contraceptive methods that are being used rather than a simple yes or no question).

## Potential Changes to the PMA2020 Survey

PMA2020's current sampling design (described in detail in Appendix A) meets country and FP2020 needs in many ways, including the fact that it is quite similar to the approach of the DHS, though with frequent, multiple rounds on a smaller scale. Thus, it is intended to provide more–rapidly available data using fewer resources. Like the DHS, it is a multistage clustered design using respondents nested in households and households nested within enumeration areas. However, Gates Foundation staff and other experts and stakeholders with whom we spoke felt that PMA2020 could be more than just a smaller and quicker version of the DHS (with data available sooner), as many perceive it to be now. Indeed, stakeholders suggested a number of alternatives to the current PMA2020 design.

Below, we analyze the advantages and disadvantages of potential changes to (1) **sample design** for both initial and subsequent rounds of data collection, (2) **frequency** of data collection, (3) **methods** of data collection, (4) PMA2020 survey **content and length**, and (5) the use of **facility-based data**, including suggestions for optimizing both service delivery point surveys and service statistics. This analysis informs recommendations offered in Chapter Fourteen.

### Potential Alternative Sample Designs

While some PMA2020 staff and decisionmakers commented that the current PMA2020 sampling strategy is "fine as is," "very good," and "the most feasible sampling technique," many other in-country stakeholders suggested potential changes to the current sampling strategy, including (1) replicating "the strategy that is employed by DHS," including possibly expanding sampled areas, to maximize comparability and to ensure that one can have "both national estimates and some subnational estimates," although we note that this would be a significant trade-off, as precision would decrease unless the sample size were increased significantly; (2) making PMA2020 a true panel survey; and (3) better aligning PMA2020 with service statistics. The following potential alternative sampling schemes were informed by these perspectives, as well as by the statistical analysis presented in Chapter Seven.

When evaluating the sampling schemes, we considered those parameters identified as important to the Gates Foundation, including cost, timeliness, continuing to use the resident enumerator model, and generating national and subnational estimates.

#### *Sample Design for the First Round of Data Collection*

PMA2020 currently uses a multistage sample design, in which sampling is done sequentially across several hierarchical levels. The first level is the enumeration area, next are households, and then all eligible women within selected households are interviewed. This approach has a number of advantages—for example, "it provides good coverage, is simple to implement, and allows for control of field-work quality" (Aliaga and Ren, 2006). However, it also has important drawbacks—it requires weighting to correct for different individuals having different probabilities of selection, and the clustering of individuals within levels can lead to a design effect larger than 1 and, thus, a larger variance when compared with a simple random sample (Lavrakas, 2008). We now discuss two alternatives to the current design.

##### Option: Implement an Optimal Cluster Design—Adjust Cluster Size

In clustered designs with sampling weights, the level of variability of an indicator of interest can dictate whether a larger sample size is needed in particular clusters. In the extreme case in which the correlation of the indicator observations within the cluster is 1 (i.e., all observations are identical), sampling only one observation from such cluster will provide the same information as sampling 10, 50, 100, 1000, . . . observations in such cluster. On the other hand, when there is substantial variation of the indicator within a cluster, selecting more observations within a cluster can potentially provide more information and, thus, improve the precision of the estimators. PMA2020 could therefore use an **optimal cluster design**, in which homogeneity versus heterogeneity of the main indicator of interest within clusters is taken into account when determining the sample size within each cluster. The sample sizes may, therefore, vary by cluster. An optimal cluster design has the potential to improve the precision of an estimate. In our analysis of the PMA2020 sampling design presented in Chapter Seven, we calculated the improvement in variance that can be obtained when using an optimal design compared with the current design.

##### Implications for PMA2020

The implications for PMA2020 of using optimal cluster design would be a change in the number of households selected in each enumeration area. Some enumeration areas might need many fewer interviews; others might need more. Precision would increase. This might require

adjustment of the number of resident enumerators and their assigned location(s), which would have cost implications.

#### Option: Simple Random Sampling

For PMA2020, simple random sampling would entail randomly selecting households and then, within those households, selecting some or all women of reproductive age to be interviewed. This approach produces smaller variance of indicator estimates. However, because sampling no longer includes clusters, it requires covering the entire country (e.g., possibly going to a distant location to interview one selected respondent), with serious implications for logistics and costs. Furthermore, this approach raises concerns about subnational and subgroup representativeness. As noted by the PMA2020 program, "Simple random sampling is not preferred for conducting a national level survey due to budget and logistical constraints" (Zimmerman, 2017). We concur with this assessment.

### *Sample Design for Later Rounds of Data Collection*

PMA2020's current design generates estimates each year, looking both at point estimates and at trends. Some stakeholders took issue with PMA2020 claiming to be a serial cross-sectional survey that can generate point estimates because they interpreted a cross-sectional design to mean that each round is independently sampled, which is not the case in the PMA2020 survey, in which the same enumeration areas are sampled in subsequent rounds. Instead, they describe this repeat sampling from the same or adjacent areas as "sentinel surveillance." At the same time, PMA2020 is not designed as a longitudinal survey of the same individuals. This tension could be addressed by a different or mixed design, combined with clear communication about what PMA2020 surveys represent and how the data should be analyzed and interpreted.

Even after the design for the first round of data collection is established, there are a number of ways a survey could proceed in subsequent rounds regarding selecting the sample:

1. **Preserve the current approach to PMA2020 for Rounds 2–4**—i.e., select a new random sample of households in the same enumeration areas as in Round 1. While this option has the drawback of potentially resampling certain households, the benefits are that the same resident enumerators could continue to work in their local areas and would not have to travel as extensively. There are also additional costs of mapping and listing new enumeration areas chosen for each round.
2. Implement a time series of cross-sectional surveys but **"refresh"—that is, select a new sample by sampling from new enumeration areas each round**. The DHS takes this approach, though for rounds much further apart than PMA2020's.
3. **Preserve the current approach for Round 5**—that is, refresh the sample by choosing a sample from enumeration areas adjacent to those used in Round 4. This approach has the advantage of being a new, independent sample, but, by selecting adjacent enumeration areas, theoretically, the same resident enumerators could conduct the survey. However, these are not completely new *random* samples because each individual within the population does not have the same likelihood of being selected. Of note, this approach has been used in Round 5 in Ethiopia, Kenya, and Uganda. In Ghana, completely new enumeration areas were selected, not necessarily adjacently.[1] In DRC, the original enu-

[1] This was not done intentionally by PMA2020 but was due to decisions by the country statistics bureau.

meration areas were used for Round 5 because the enumeration areas there were considered big enough that there was little concern about repeatedly surveying the same women.

4. Lastly, there is the option of **combining some of these approaches**, as PMA2020 does by staying in the same enumeration area for Rounds 1 through 4 and then refreshing enumeration areas for Round 5, or some other combination of the above, such as refreshing enumeration area selection at some other frequency. This is the path that PMA2020 has chosen.

Separate from selecting enumeration areas, a survey could consider a different approach: Rather than staying in the same place or changing enumeration areas, the same individuals could be intentionally resampled. This would be considered **a panel design**, in which the effort aims to interview the same respondents as in Round 1. We examine the panel design option more closely for PMA2020's consideration. A panel survey—i.e., a longitudinal survey that surveys respondents on two or more occasions—can give information about trends or changes in characteristics of a given population, including at the individual level, over time (Lavrakas, 2008). PMA2020's current sampling approach could be considered a panel survey of the selected enumeration areas. However, it effectively results in only a small panel-like subpopulation of households and individuals, as by the fourth round of data collection, many women have already been surveyed (15 percent in most countries, according to the Gates Institute; in Uganda's Round 4, for example, almost 30 percent of women had been surveyed by PMA2020 at least once before, and 33 percent in Ghana's Round 4). In the course of this evaluation, we learned that the Gates Foundation, the Gates Institute, and the PMA2020 External Consultative Group have had numerous discussions about the potential benefits and drawbacks of a panel approach. The arguments below are informed by those discussions and also represent our additional analysis. We present three variations of the panel design option.

### Option: Maintain the Status Quo

PMA2020 has chosen to refresh enumeration area lists after four rounds because of a balance between the need for new enumeration areas and the logistical complications of changing enumeration areas. More information about the benefits and drawbacks of this design will be available with more Round 5 data from countries where enumeration areas are refreshed, and there was some expressed concern about only refreshing with contiguous enumeration areas. However, given the constraints and considerations for PMA2020, maintaining the current refresh schedule is a reasonable option, taking into account the rate of repeat surveys by Round 4 (approaching 30 percent in some countries) and the importance to the program of maintaining local resident enumerators.

### Option: A Straightforward Panel Survey

#### *Strengths of This Approach*

A panel survey, with some or all participants being surveyed repeatedly over time, is intended to accurately measure changes over time for individual respondents, though not necessarily for the population as a whole. Depending on design, it could also allow for retaining trained resident enumerators, as most (but not all) respondents' locations are unlikely to change significantly. Indeed, in many ways, PMA2020 is well positioned to do this, and it has started piloting a new maternal, newborn, and child health module with this design. This approach

would distinguish PMA2020 from the DHS. For example, longitudinal data could be very useful for studying such contraceptive dynamics as discontinuation or method switching, which are of great interest to many of the stakeholders RAND interviewed and which cannot be well understood with DHS data or current PMA2020 data.

*Potential Weaknesses*

A panel approach presents a number of challenges. The question of cost hinges on how large the panel is, how often people move and have to be located, and whether conducting a panel allows PMA2020 to achieve savings elsewhere, such as by replacing the regular cross-sectional survey or allowing it to be conducted less often. However, the main difficulty is that panel surveys require a more complicated design to maintain representativeness over time (Lavrakas, 2008). If the panel does not begin as representative, the data will not provide valuable information about the entire population. But even if the initial sample is representative, the samples in subsequent rounds may not be, depending on the difficulty of tracking people down and related loss to follow-up, maintaining age representativeness as the sample ages, and maintaining representativeness with regard to the population overall, specific enumeration areas, or something else. There is also an issue of what to do if a household splits or changes: Women's surveys are currently nested in households, and in a longitudinal approach, this nesting would not be constant (for example, should a young woman marry between rounds and then move to another house to live with her husband, she would no longer be nested in household A with her mother and father but rather would now be in household B, which might already have a number of women, and then tracking her survey over time given household indicators from a different household could be complicated). Given the relatively short time intervals between PMA2020 rounds, these concerns about migration and household changes are somewhat less of a concern than if the rounds were conducted less frequently. The data to specifically examine this concern are not yet available but should be examined once more surveys have been carried out over a longer period of time. Finally, because the same people are asked questions in each round of a panel survey, there is concern that ideas or information provided in prior rounds of questioning might influence answers in later rounds (the Hawthorne effect, which is of particular concern, given the short time intervals between PMA2020 rounds). This has already been articulated as a concern about the current design, given the up-to-30-percent rate of repeat surveys.

### Option: A Rotating Panel

A variation of a straightforward panel design is a rotating panel, in which each panel lasts for a shorter time. An advantage of this approach is that "by restricting the duration of each panel to a shorter period, problems of attrition are reduced and representativeness is more easily maintained" (Lavrakas, 2008). However, such a design would lose its longitudinal nature, for example, to explore contraception use over the course of a woman's life, which is a major advantage of the panel approach. If the goal is to track changes in use over years rather than decades, this limitation poses less of a problem.

### Option: Implement Both a Panel and a True Cross-Sectional Survey

As some stakeholders suggested, the twin but sometimes conflicting desire for both cross-sectional and longitudinal data could be mediated by implementing a **targeted panel component separately from a truly cross-sectional survey**, in which some data are collected repeatedly from a panel while other data are collected from newly selected samples. This design

could help address some of the issues in the PMA2020 design noted previously, such as the intervention effect resulting from repeated surveys. But if interviewers are still returning to the same population, that bias has not been eliminated, and there are concerns regarding mixing methods of data collection for interpretation of the data. Furthermore, unless the overall sample size is increased to maintain the cross-sectional component and add the panel, there will necessarily be a decrease in precision.

#### Implications for PMA2020

Without a significant change to its design, PMA2020 could elect to identify a subset of its respondents to be part of a longitudinal panel. These women could be asked more or different questions and analyzed separately over time for a closer look at factors contributing to trends and change. Other analyses looking for point estimates would not include women who had been repeatedly sampled, and this could vary based on indicator as well. A similar idea is already being piloted in some form with a maternal, newborn, and child health module.

PMA2020 would require an increase in sample size to maintain accuracy of the remaining sample independent of the panel—i.e., if panel women are deleted from some calculations because they are interviewed more frequently, there would need to be an overall increase. However, they would also be able to give different kinds of information.

At the same time, cost could be moderated by a decreased frequency of data collection for the larger population for the items for which the panel is considered informative but not representative. However, this might have negative implications for employment of resident enumerators.

Lastly, a panel component could allow for the questions regarding "why" that many stakeholders suggested was something that PMA2020 was uniquely positioned to answer. The statistics are hard to interpret without context, and qualitative conversations with women like those conducted by resident enumerators might be able to answer some of the causation and intentionality questions, particularly in a population with a relationship to the project and the interviewer.

### Options Related to Frequency of Data Collection

As noted earlier, PMA2020 collects data every six months for the first four rounds and then continues yearly thereafter. So far, almost half of PMA2020 countries have conducted five rounds of surveys since 2013. There are several options for modifying PMA2020's current frequency of data collection.

#### *Option: Implement a Rolling Sample*

A rolling sample is one with nonoverlapping panels that can be combined over different time frames to create samples of different sizes. This option "enables a single survey to serve multiple purposes" (Alexander, 2002). This is similar to what the DHS is testing in Senegal and Peru—it is known as the *continuous DHS* (Rutstein and Way, 2014; DHS, undated). A knowledgeable stakeholder described the effort in Senegal as "an experiment to understand the benefits and limitations of the continuous approach in a lower-capacity country." The data will be reported annually to give national and urban/rural estimates and then aggregated for longer periods to give further results.

However, the yearly samples alone might not be large enough to provide reliable national or subnational estimates. The cost for the five annual surveys is expected to be similar to the

cost of one DHS (DHS, undated; Rutstein and Way, 2014). An evaluation is planned in 2018 at the end of the five-year experiment.

The American Community Survey is possibly the most well-known example of a rolling sample, with a monthly sample size of 250,000 housing units to produce five-year cumulations for small areas "at the same level of sampling reliability as the long-form census sample" (U.S. Census Bureau, 2009). This means that while each year the data give a five-year summary, that number is updated every year and shows change over smaller periods of time (U.S. Census Bureau, 2009).

### Strengths of This Approach

"The flexibility of the rolling sample design comes from the opportunities it provides to make different tradeoffs between spatial, temporal, and demographic detail" (Alexander, 2002).

PMA2020 is already designed to conduct frequent surveys with relatively small samples, particularly if the model of using resident enumerators is continued. Reframing the PMA2020 survey as a rolling sample would allow it to provide national estimates less frequently without decreasing the frequency of data collection for other topics of interest. Smaller populations could be targeted in each round, with the totals aggregated for annual estimates.

In addition, in this model, instead of a single survey, multiple surveys could be designed, each to be administered more frequently, but some with fewer questions and others with more. Topics expected to change less frequently would be queried yearly or less often, while the topics changing more often could be asked about during every cycle.

### Potential Weaknesses

Though this approach has been tested in two countries as part of the continuous DHS, one stakeholder who has worked in Peru feels that it has worked well there because of the support of a well-developed census bureau. This respondent expressed concern that this would not be feasible in sub-Saharan Africa. And, in fact, after 2018, the continuous DHS experiment will not continue in Senegal. According to an involved stakeholder, "The kind of infrastructure that we have to work with didn't work for the continuous survey in Senegal."

A second major issue is interpreting the findings over this moving time frame. The longer time frames for interpretation mean that some indicators would be expected to change. "The rolling moving average of single-year estimates is very hard to interpret," said a stakeholder. However, PMA2020 needs to add and subtract those single-year fluctuations to estimate the next year's, even though "the single-year estimates are not stable" and therefore are not reported.

### Implications for PMA2020

Just because rolling samples do not seem to be working well for the continuous DHS in Senegal does not mean that they would be infeasible for PMA2020. But there are challenges of continuous data collection, including the costs of keeping resident enumerators throughout the year and potential seasonal limitations in some countries. (On the other hand, some consider employment and empowerment of resident enumerators an important achievement to date that should be continued into the future.) Furthermore, it would be useful to gather certain data elements more often, while some others are probably already collected more often than needed. Continuous data collection would need to be coupled with a careful design specifying which questions would be asked, when, and of whom, as we discuss later in this chapter.

#### *Option: Decrease the Frequency of Household Surveys*

Both stakeholders and our statistical analysis in Chapter Seven suggested the potential of decreasing the frequency of data collection from households—from every six months to annual—without losing value. Decreased frequency would decrease cost and respondent and resident enumerator burden, though it might make the resident enumerator model less feasible (if PMA2020 continues to focus narrowly on family planning) because they would have long breaks between rounds. Explicitly taking into account resident enumerator preferences is part of running a program that relies heavily on the resident enumerators. Decreased frequency would also potentially delay detection of changes in rates, but that depends on the actual changes in rates of the key indicators.

### Method of Data Collection

Another dimension of survey design is how the data are collected. Surveys can be conducted face to face, by mobile phone, over the Internet, or by mail. PMA2020 has elected to conduct face-to-face surveys (and enter data using mobile phones), which has both advantages and disadvantages. Advantages include the ability to establish a personal relationship with the interviewee, to collect information about time spent on the survey, to assess the person's involvement and understanding, and to achieve a higher probability of completion. At the same time, face-to-face data collection is more costly than other methods. PMA2020 asks many sensitive personal questions, and there is conflicting evidence about whether individuals are more willing to share this kind of information with someone they know or anonymously.

#### *Option: Continue Primarily Face-to-Face Surveys*

A face-to-face survey could be conducted by a random interviewer, as is done in many surveys. This is the easiest and most efficient approach. Alternatively, as in PMA2020, an interviewer can be specifically selected to be local and even to have a relationship with the interviewee. This may increase trust and response rates or may impact answers in less desirable ways. PMA2020 recently reported a first exploration of this dynamic in five countries (Burkina Faso, Ethiopia, Ghana, Kenya, and Uganda) in a methodological paper by Hawes et al. (2017). Although there was a suggestion at a recent PMA2020 External Consultative Group meeting to consider moving away from the resident enumerator model, the Gates Institute is committed, for now, to continuing to use resident enumerators. A model without resident enumerators would lose the individual-relationship aspect of the current design but would allow more geographic flexibility and could be considered in the future.

#### *Option: Incorporate Surveying from Afar*

Spurred by evolving technology, surveys are now increasingly conducted from a distance. Traditionally, this was done by telephone, but now surveys can also be conducted by a mobile phone via text message or over the Internet. Due to limited access to computers and the length of the PMA2020 questionnaire, conducting the complete survey over the Internet is not feasible. However, another option is potentially adding polling or a polling-like component to the PMA2020 survey. We explore polling in more detail next.

#### Polling

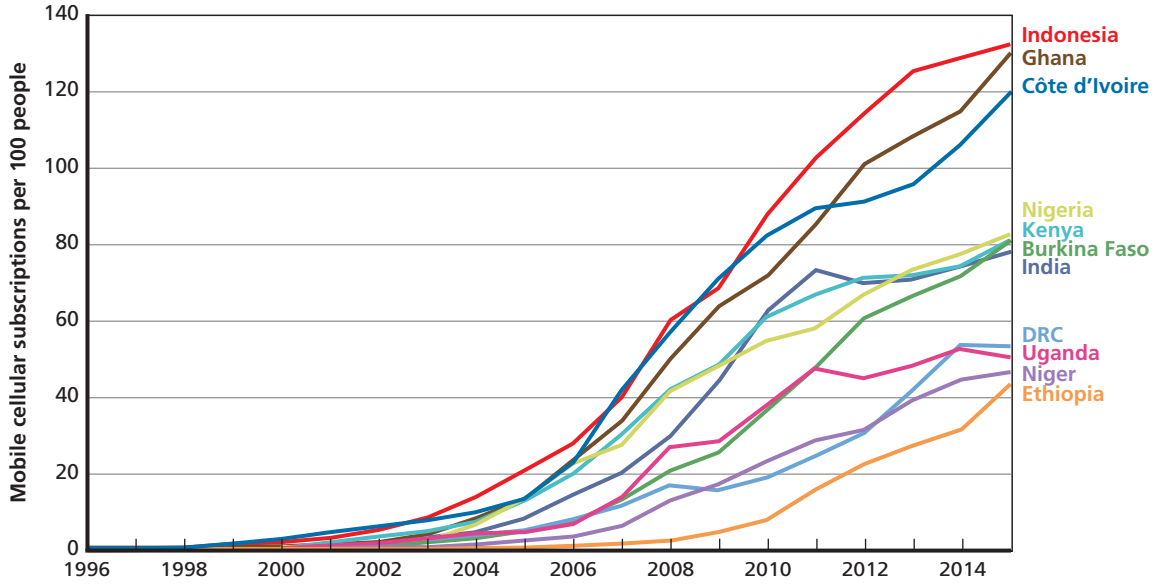
A poll is a public opinion survey designed to collect information about people's opinions, preferences, perceptions, attitudes, and evaluations of public opinion issues (Lavrakas, 2008). It is usually a fast way to get answers to short, targeted questions, which can be changed as needed.

Polling might work best for PMA2020 in combination with another design or for a specific purpose, rather than as a replacement for the full survey. For example, if an intervention were implemented to increase the use of modern contraceptive methods in a specific region, a PMA2020 poll could assess the impact of the policy within a specified time, which might be shorter than the routine six- or 12-month intervals of the PMA2020 survey. The timing of polling is usually not predetermined long in advance; instead, it is decided on and conducted as the need occurs based on the societal conditions that require the update of a certain belief (e.g., rate of mCPR) about the population.

Polling can be done in person, but, increasingly, at least in more-developed countries, it is done using phone or even text messages. However, different polling methods have their own challenges. For example, when it comes to polling via text messages, considerations would be whether the sample with reliable texting access is representative of the population of interest. Indeed, in much of the world, access to cell phones is increasingly ubiquitous, but survey implementers would still need to investigate such concerns as whether there are age or socioeconomic groups who would be more or less facile with the mobile phone and more or less likely to have a charged phone in a place with network access.

Some stakeholders have suggested that a poll could be an opportunity for private-sector involvement or, perhaps, for a possible public-private partnership arrangement. However, polling may be a more viable approach in some countries than in others, based on the penetration of mobile phone use. Figure 8.1 shows a rapid increase in mobile cellular subscriptions in PMA2020 countries, beginning around 2006.

**Figure 8.1**
**Mobile Cellular Subscriptions per 100 People in PMA2020 Countries, 1996–2015**




SOURCE: International Telecommunication Union, World Telecommunication/ICT Development Report and database, 2017.

RAND *RR2112-8.1*

### Strengths of This Approach

The strength of a poll is that it is fast, is easy to implement, and provides quick answers to specific questions of interest in a population. Because it is short, it avoids respondent burnout. Because it can be implemented quickly, it can occur as needed, which is especially valuable when obtaining data immediately is important.

### Potential Weaknesses

Polling also presents challenges. The most important one is maintaining the representativeness of the sample. In general, polling methods provide the possibility of selection bias, which is not usually alarming if polls that are repeated many times, using different methods, return the same inference. With some recent exceptions, election polling is a good example.

For PMA2020, the use of polling would require more scrutiny of representativeness since polling would not necessary be repeated quite as frequently. In addition, because polling can only concentrate on very few questions, this approach could restrict the survey's use for other purposes beyond inferences about the mCPR.

Respondents' opinions or self-report of attitude, often regarding something timely and in the news, can be biased by the media's presentation of an issue. For this reason, PMA2020's use of polling should concentrate on questions of interest that would be less likely to evoke strong reactions from respondents.

### Implications for PMA2020

For PMA2020, polling could be used effectively if there are very specific questions that need fast answers. For example, questions could probe reactions to a shortage of a supply or to a change in family planning policy. Polling could provide information on changing situations in a time frame faster than the current PMA2020 six-month interval. However, polling is only appropriate for very targeted information (since the number of questions needs to be very small and they must require little explanation), and, as such, polling results may not support certain detailed analyses.

Representativeness needs to be paramount, and all efforts should be made to prevent the polls from turning into a selection of convenient samples that do not represent the population of interest.

In addition, due to the sensitivity of topics that PMA2020 covers, the range of questions that could be conducted by poll might be limited. Interviewers may be hesitant to ask questions about sexual behavior over the phone or via text message; however, interviewees may be more willing to answer questions and/or endorse behaviors that they would not want to admit in person. This potential trade-off warrants further research and may be an opportunity for PMA2020 to contribute to generalizable knowledge about survey methodology in developing countries, especially in the realm of family planning.

## Survey Content and Length

Content and length are considerations in any survey design. PMA2020 is obviously committed to collecting family planning data, but it has recently expanded its reach into other topics, and in-country stakeholders have expressed interest in diversifying further. As discussed elsewhere, there are advantages to using this platform to gather other non–family planning data; at the same time, there is concern about diluting the clarity of the program's purpose. In addition, the method of collection needs to be matched to the content; sensitive information may not be

collected well over the phone, and information about why someone is making an important decision is not conducive to quick approaches like a poll.

**Additional family planning questions, subpopulations of interest, and/or non–family planning content.** When asked what additional questions, modules, and populations would be useful to consider adding to PMA2020, in-country stakeholders gave a range of suggestions, some of which are already being piloted in certain countries. These suggestions include questions or full modules on adolescents' access to family planning services, nutrition/malnutrition (several mentions), infant and maternal mortality (several mentions), maternal health in general (several mentions), female empowerment and financial independence, postpartum family planning, migration, religion, perimenopausal women who are "at the upper end of their reproductive cycle," abortion, and deliveries with skilled birth attendants, to name a few. Frequently cited populations to explore in more depth were adolescents (especially married adolescents), women at the upper range of eligibility for the PMA2020 survey, men (mentioned by PMA2020 staff and decisionmakers in DRC, Ethiopia, Ghana, and Niger), and such vulnerable and marginalized populations as women living with HIV (noted in Kenya) and women who are handicapped (from a respondent in Burkina Faso). Most respondents did not call for additional demographic data to be collected, and some worried that asking for more personal data would inhibit responses later in the survey. Acknowledging that adding questions or modules would necessarily lengthen the survey, stakeholders preferred less-frequent surveys with a larger number of questions instead of data collection on a smaller set of questions every six months.

There was strong support across several countries (Uganda, Nigeria, and India) and types of respondents (UNFPA, USAID, an NGO, and a government stakeholder) to explore qualitatively the behavioral factors that influence family planning, including the demand-side cultural factors affecting acceptance of family planning methods, such as the cultural preference among some populations in India for sons. As one Ugandan respondent noted, "We are missing out on the story behind the numbers," and a Nigerian respondent noted, "We keep seeing the same causes of non-use of family planning methods. . . . They keep coming up, the same reasons. . . . Why? And if we don't know the why, you can't do much about it."

Several PMA2020 staff members, including one in Nigeria, acknowledged that they have "so many requests during dissemination" to expand the survey beyond the existing questions and base modules, but they recognize that "every survey has its limitations . . . budgetary constraints . . . and technical reasons" to stay focused primarily on family planning. Several stakeholders noted that the survey cannot expand indefinitely, and, generally, there was support for rotating modules around a core of family planning questions, depending on the current needs of the country.

These perspectives make it clear that the length of the survey and respondent burden are important considerations.

### *Topics and Question Selection*

Currently, there are two PMA2020 survey instruments: a two-part household/female questionnaire and a service delivery point questionnaire. The household survey is asked of a representative in each household, and the female version is asked of eligible women in each household. The survey parts are linked so that the final data set includes both household and female information. The service delivery point survey is administered within service delivery points, as described in Chapter Two. The link between the service delivery points and the

surveyed households is at the enumeration area level, though Global Positioning System (GPS) data could connect the two on a more granular level, as long as deidentification is maintained.

#### *Option: Vary Content and Questions*

Reducing the frequency of the survey is discussed above; another option is to conduct the survey with the same frequency but to rotate the content such that different modules or even different questions are asked at different intervals.

Under the current design, all questions (with some skip patterns) are asked in either the household or the service delivery point survey. Except for small differences by country, mostly regarding local resources and housing for answer choices, the same survey instrument is administered in every round. PMA2020 has begun experimenting with adding new modules, such as menstrual hygiene, women's empowerment, and schistosomiasis. To our knowledge, these modules have not been implemented in a rotating format. Neither has PMA2020 varied specific questions according to region of the country, rural/urban designation, or subgroups of respondents. Using information gathered in previous rounds, such as statistical properties of certain indicators, questions could be administered to different subgroups and at different frequencies.

##### Strengths of This Approach

Varying questions could increase efficiency by asking questions only of the number of people needed for a statistically meaningful response—for indicators that are highly clustered, fewer people would need to be asked. Varying the frequency for particular questions (i.e., modules that are only asked every other survey round) could also address concerns about data being collected more frequently than might be useful, leaving more room for innovation and creativity. For example, PMA2020 could roll out different versions of questions to validate and measure the impact of question wording; PMA2020 could target certain questions to specific locations to evaluate local interventions; and questions about topics with higher coefficients of variance could be asked of more people or more frequently, while others that are more consistently answered could be asked less frequently.

This approach could be applied to any sampling scheme. It would have the benefit of keeping resident enumerators employed year-round, broadening the data platform, and, thereby, potentially facilitating greater country ownership.

PMA2020 is already using a digital platform, so technology can facilitate the administration of different questions to different populations or with different frequencies, using the same interviewer and hardware, without much additional training of the interviewer.

##### Potential Weaknesses

Varying questions complicates implementation, though not significantly, given the digital administration. It would require additional programming for mobile-phone data collection, additional training of the resident enumerators, and additional analyses to take advantage of the varied questioning.

##### Implications for PMA2020

Members of the Gates Institute have identified one way of structuring this potential approach by performing PMA2020 surveys quarterly with rotating content modules. Increased variation within survey modules and even at the question level is feasible and would set PMA2020 apart from other surveys, such as the DHS. The ability to quickly test multiple versions of

questions, to understand the relationship of a new variable to others, or to quickly gather data on the impact of a recent intervention were all suggested by stakeholders, and all are within the current capabilities of PMA2020. It would not be particularly expensive or add to required infrastructure. Additional needs would be programming, training, and analysis, with targeted administration to certain households or areas, but it would not require significant changes to the sampling design.

### Facility-Based Family Planning Data

As discussed in Chapter Five, there is enthusiasm for using service statistics to monitor progress toward family planning goals, and, indeed, a number of key FP2020 indicators rely on this information (couple-years of protection, for example). These indicators change quickly and frequently. Collecting data on them could provide rapid information about use and demand, as well as limitations to access, including stock-outs. PMA2020 could enhance how it complements service statistics in two ways.

#### *Option: Service Delivery Point–Based Selection*

One option currently under consideration by the Gates Foundation is for service delivery points to be the primary sampling unit. Then the survey would select households that use a specific service delivery point for care, rather than using a population-based or household-based approach. This would be an effective design for events that occur within health care facilities, such as complications of a sterilization procedure. However, there are many different ways of achieving contraception. For example, while IUDs and sterilization require medical intervention, condoms do not require any interaction with a health care facility. In fact, using a sampling unit that only uses visitors to a health care facility might misrepresent the population significantly and certainly would not provide any information about non-users.

#### *Option: Improve PMA2020's Contribution to Service Statistics*

PMA2020 gathers rich contextual information in its service delivery point surveys. This information adds knowledge about quality, services available, and more to the facility-based data it gathers and to routine service statistics in the HMIS. Service delivery point data can validate, triangulate, and contextualize other sources of such data, such as issues related to demand for family planning. This might require more frequent data collection, which would be enabled by reducing the frequency of the household survey.

One model for continuous service statistics comes from the DHS in Senegal, where the DHS is collecting data from health facilities. Using four different data collection methods (health facility inventory, interviews with health care providers in the facility, observations of health care consultations, and exit interviews with the clients of observed consultations), this continuous service provision assessment is in its fifth year in 2017. Each year, the sample and data collection topics have varied, but as with the continuous DHS, the data can be aggregated in different ways, because though the sample has changed each round, eventually the whole country is represented. By the end of the fifth phase, "the Continuous Service Provision Assessment will provide a complete picture of the health care system in Senegal, as all health facilities in Senegal will have been surveyed" (DHS, undated).

However, this continuous and ultimately national data collection seems an infeasible model for PMA2020 because it requires collecting from all parts of the country, and PMA2020 has focused on remaining in targeted geographic locations. However, within the context of

other designs, such as a panel of health facilities or collection of service statistics from service delivery points without the interview or observation portion, more-frequent service delivery point surveys providing contextual information about access, quality, and demand could enhance their usefulness.

## Summary

Stakeholders raised a number of points about the PMA2020 survey design. They raised concerns about resampling the same households, noted confusion around marketing itself as providing nationally representative estimates but only sampling in a few states in certain countries, wanted more subnational data, and wanted more-granular information about what mCPR is conveying within different subpopulations.

We have presented different options for potential changes to the design of PMA2020 surveys, some of which may be more feasible for the program, and some of which may be less feasible, particularly given the express goal of the Gates Foundation to contain costs. The benefits and drawbacks of changing one element of the survey design are highly dependent on decisions about others.

Ultimately, PMA2020's design should be responsive to the data needs of its users and should fill a gap in the family planning data landscape in the countries in which it operates. Most survey questions can be asked annually rather than every six months. Cost savings accruing from lengthening the survey intervals from semiannual to annual can be applied to other options that meet stakeholder needs. For example, PMA2020 is currently well positioned to provide frequent estimates for topics of a sensitive nature at a subnational level. Other promising options for PMA2020, depending on priorities of the program, include expanding the pilot of panel data collection, continuing to pursue collecting data by mobile phone rather than face to face (while also studying the trade-offs involved), implementing rolling samples, and rotating modules or administering them in a targeted fashion to particular subgroups of respondents.

We summarize our recommendations for design modifications in Chapter Fourteen, with a focus on a panel design and decreasing the frequency of data collection, while considering ways to increase use of service delivery point surveys and improve efficiency through optimal design.

CHAPTER NINE

# Track20 Goals, Accomplishments, and Challenges

## Overview

We sought stakeholder views of Track20's goals, its accomplishments, and the challenges it faces now. Interviews with more than 200 stakeholders from program countries, the Gates Foundation, Avenir, and other U.S.-based respondents reveal that the Track20 approach is well respected. It has achieved most of its objectives and is on track to achieve its original goal.

Stakeholders thought that Track20 has the potential to make significant, lasting impacts on family planning data collection, analysis, and use around the world. It has achieved a balance between standardizing a methodologically sound system for producing consensus estimates (using FPET) from available family planning data sources while accommodating flexibility and allowing countries to define their own priorities for improving family planning services.

In this chapter, we summarize stakeholder views, highlighting key messages that emerged from the interviews. We also discuss the duties, strengths, and limitations of the M&E officers. We conclude the chapter by presenting an empiric logic model reflecting revisions to the initial Track20 model presented in Chapter Four. The revised model reflects inputs from the stakeholder interviews.

## Goals of Track20

According to Gates Foundation staff**,** Track20 was intended to add accountability to the FP2020 agenda. Track20 reflects recognition that accomplishing that agenda will require more than new surveys. Rather, Gates Foundation family planning programs needed to obtain governmental buy-in to build capacity, making it possible to better leverage existing capacity on the ground to analyze and use data to advance family planning programming in countries.

**Track20's priority is to maximize use of all available family planning data.** Members of the Avenir team described Track20's purpose as improving monitoring of global progress for FP2020 in priority countries, or, put more simply, to "maximize the use of all data" and to "disseminate data and help governments use data." They described two parallel but mutually reinforcing parts of Track20's work:

- **global efforts**, which include compiling data for FP2020 from participating countries, analyzing them, measuring annual progress, and reporting the data to global partners in various forms so that they can push the FP2020 agenda forward

- **country-specific efforts**, which include building capacity within MOHs to use family planning data more effectively for decisionmaking, and, specifically, to "embed data capacity within countries" so that they improve their data generation, analysis, and use.

Avenir staff emphasized that, unlike PMA2020, it is not a cohesive, unified "program" but rather a global methodology and new system for monitoring family planning data. They emphasized its efforts to establish a country-owned process to implementing an agenda in a "country-first way." Avenir staff feel that the original Track20 goals and objectives are still largely pertinent—focusing on building countries' capacity to analyze and use data for monitoring family planning programs at different geographic levels. They distinguish between *dissemination* and *use* of data and prefer the term *data-informed decisionmaking* as a more precise alternative to *data for decisionmaking*.

Other U.S.-based stakeholders (outside of Avenir) held similar views of Track20's overall goal of measuring progress toward FP2020 goals using all available data inputs. One stakeholder called its work "an advocacy exercise within countries."

Track20 was originally designed with a fair amount of flexibility at both the global and country levels such that the implementation varies substantially across countries and contexts. This flexibility has been productive and necessary in many ways, but it also has a downside: One Track20 contractor commented that when they first started working on this initiative, they had a "hard time figuring out what we were supposed to be doing." Even with the flexibility to implement the program differently across the countries, most stakeholders seem to have converged on the need to focus more actively on improving the quality of service statistics.

**The work of Track20 has become more country-focused.** As with PMA2020, the goals of Track20 have evolved over time. One of the biggest changes was that Track20 was originally very focused on the FP2020 core indicators. While those indicators were very useful *across countries,* the indicators that are of the highest priority and most useful to *in-country decisionmakers* are more nuanced. So, over time, Track20 has become much more country-focused.

Initially, Avenir spent a lot of time socializing the FP2020 core indicators and training partners on how to calculate them. Now, there is more emphasis on supporting countries to examine their own family planning goals and to determine whether their program-level data show progress toward those goals. For example, Track20's relatively new FP Goals model aims to improve strategic planning at the country level by establishing a baseline description of what family planning efforts and programs are currently under way, developing and comparing multiple scenarios of future scale-up, and projecting mCPR growth for each scenario. This model has the potential to significantly change country-level decisionmaking and future donor investments.

Track20 has evolved to focus both on providing official national estimates for global reporting purposes *and* providing better data (at both the national and subnational levels) for planning and tracking progress against national family planning goals. Indeed, the Track20 website explicitly notes that "Track20 develops tools that help partners—at both the country and global level—engage in data analysis" (Track20, 2017). A Gates Foundation respondent noted that there is now less perceived value in national family planning estimates for global reporting purposes, compared with more-actionable subnational and local estimates; in essence, this view supports the prevailing sentiment in program countries about the importance of actionable data from national to subnational levels.

The views of in-country stakeholders were similar to those based in the United States. Not surprisingly, the views of Track20 staff in program countries aligned most closely with those of Avenir. In particular, they explained that Track20 helps monitor progress toward FP2020 goals; builds capacity; supports data use; and, in relevant countries, coordinates with PMA2020. M&E officers have been trained "to be able to run their own analyses using FPET" and participate in "any meeting that they feel is going to help inform decisions as to what programs we should be focusing on." Track20 staff noted that their role is to ensure that "FP2020 indicators become part of service statistics and are regularly compiled and collected" and to "give people a sense of data and have them use data in a scientific way."

Government respondents closely identified Track20 with helping countries track their progress and provide technical assistance to support achieving FP2020 goals. NGO and other stakeholders had varying levels of understanding of Track20's goals depending on their individual roles. Both groups associated Track20 with using FPET to track the 17 FP2020 core indicators. Most understood that Track20 provides secondary analysis of existing data sets and does not collect primary data.

## Accomplishments of Track20

**Track20 is helping countries define what a health information system should be.** Stakeholders at the Gates Foundation see Track20 as stimulating "valuable conversations about data" in countries where it operates and helping to foster cultures of using data to drive decisionmaking. Track20 is seen as being very "pragmatic, taking a user lens" by facilitating alignment within countries about indicators, establishing consensus on progress toward key goals, and identifying where to pull levers to reach those goals. In-country stakeholders described a strong working relationship with Avenir, whose staff members they consistently described as helpful, responsive, and always available when a country representative called.

Avenir staff also cited multiple accomplishments. They feel that the program's design and implementation are working well and believe that the program is on track to achieve its goals. They highlighted **several successful features of the program**. They have developed a data-driven and methodologically consistent process to establish consensus around key indicators from a variety of sometimes-conflicting data sources (though some countries indicated that they would like more-frequent consensus meetings to discuss their data), they have helped countries "get more value out of the data" that they have, and they have built trust over time with key stakeholders in MOHs while also "pushing governments forward" to take more ownership of the process beyond just annual reporting. In addition, they have introduced such innovations as incorporating service statistics into the FPET model to track country progress, prioritizing family planning investments using the FP Goals model, and displaying survey data in ways that "make it come alive."

Other U.S.-based stakeholders noted substantial contributions by Track20—in particular, cultivating collaborative relationships with in-country M&E officers, responding to country needs, taking advantage of existing data, and improving the use of data. Several respondents specifically praised Track20's efforts to help countries understand, improve, and ultimately use service statistics—that is, to define what a routine health information system should be.

**Track20 is partnering with countries to help them achieve their family planning goals.** In-country decisionmakers overwhelmingly viewed Track20 as a partner in their efforts

to improve family planning and achieve the FP2020 goals. They credited the program with its ability to work with local governments, provide reliable data, and encourage data use. For instance, "Before Track20 we didn't have any information on who wanted contraceptive prevalence [information]. It was like someone who was blind and they managed to open our eyes a little."

Track20 has elucidated where countries are falling behind in achieving commitments and helps to develop new goals. Particularly in countries closer to reaching their FP2020 targets, there was a strong interest in using Track20 tools for microanalyses. Examples requested included assessing the method mix, hot spot analysis, reaching subpopulations like the disabled, incorporating private-sector data, and treating the private sector as a partner in the country's family planning goals.

The Avenir team is proud that they are working in more than double the number of countries originally planned. While it has taken time to get countries to understand the somewhat-complicated methodology behind FPET, now countries are "thinking differently about estimation." Furthermore, "provinces have learned how to start thinking about data use in a scientific way. Instead of waiting for donors and their surveys, provinces can start thinking about how to use data for their own purposes."

Specifically, Track20 is beginning to help countries develop their costed implementation plans (i.e., planning and management tools used by governments and partners to achieve the goals of a family planning program), often using "Track20's presentation of their data." These plans are a relatively recent development in several countries, and they require significant time and resources to develop. While the Track20 team provides some support, the countries take primary responsibility for deciding their priorities and developing their plan. These plans are viewed as a significant accomplishment. However, the Avenir team believes that there was a strong need for Track20 to work with countries to improve the quality of the costed implementation plans. These are expected to be "correct and, therefore, resourced." It has proven challenging for Track20 to intervene and restructure what is sometimes a "laundry list" of activities into something that logically connects those activities with particular goals and ensures that goals are realistic, supported by data, and not entirely donor-driven.

A key element of Track20's success is its **credibility with key stakeholders**. Respondents commented on its ability to offer "definitive answers," citing it as "one of the most trusted sources of information at the national level." They viewed Track20 as providing "a very reliable measure of where we are at any time with a lot of scientific rigor" and recognized the team's expertise. Furthermore, because Track20's methodology is considered rigorous (albeit not completely understood by some), its estimates are considered valid, and it can "douse the tension . . . scientifically [and] objectively"—that is, it can serve as an arbiter between competing data sources. Track20's goal of promoting data use is discussed in greater detail in Chapter Eleven.

## M&E Officers

### Duties, Strengths, and Limitations

Respondents from Avenir described the ideal M&E officer as having multiple functions: (1) pulling together a composite picture of family planning data from various sources, including assessing the data's quality; (2) analyzing and interpreting these data, signaling how pro-

grams are performing; and (3) disseminating the data and making recommendations to decisionmakers, including flagging gaps that require further exploration through surveys or special analyses.

### *Assessing the Quality of Data*

Among in-country respondents, little was said about the M&E officer's role in assessing the quality of family planning data, but nearly every respondent who answered these questions praised Track20's estimates as accurate and reliable. Track20 staff and NGOs reported that in many countries there is very little quality checking at the local level, or the data are only checked once they have been aggregated to the provincial or national level. Track20 staff spoke highly of the training they received on how to audit data. M&E officers and government and NGO colleagues meet regularly; data discrepancies are observed and addressed. Errors are also reported to other domestic data collection agencies, such as national statistics offices. For those Track20 officers with the capacity, responsibility for **evaluating data quality is considered a core obligation**.

### *Modeling and Analyses*

**Track20 is working with M&E officers to help them become "data interpreters"** who can not only conduct rigorous modeling using FPET but also "take this data and give it meaning, package it in a way that leadership can understand and better make decisions with." Their tools include technical briefs; oral presentations; and graphics, figures, and other visuals. M&E officers are considered experts (or budding experts) in finding "the story" in their data, and government colleagues and NGO representatives reported feeling comfortable approaching them "to further analyze or decipher data."

Not all users of Track20 estimates (i.e., decisionmakers) were well versed in the nuances of FPET's modeling techniques, but they all spoke highly of it as a useful tool and appreciated the output. They particularly highlighted the M&E officers' role in making the FPET outputs easily understandable to them.

### *Dissemination of Track20 Work*

Track20 estimates are disseminated in a number of ways, including narrative summaries, graphical representations of the data, Excel spreadsheets, and PowerPoint presentations. They are shared directly with decisionmakers when requested and are presented at meetings of technical and other working groups and at national data consensus meetings.

Stakeholders concurred that the national data consensus meetings were the preferred venue in which to share data and, for some, the only time that they learned of the updated estimates or interacted with the Track20 M&E officers. Some appreciated the comparison between provinces and countries because it incentivized them to improve their decisionmaking and strategic planning. Instead of having to wait at least five years for updates to the DHS, program managers can more easily modify their projects annually. While the national data consensus meetings are held at least annually, some countries have begun to hold them twice a year to maintain momentum and keep program staff motivated.

Conversely, weaknesses of the national data consensus meetings include respondents' uneven level of comfort with and knowledge of data and a perception that some important stakeholders are not at the table. For example, M&E officers report having to spend a significant amount of time during their presentations pausing to define statistical terms, such as *confidence interval*. In addition, while the Track20 team, donors, NGOs, and governmental

representatives are typically present at these meetings, some felt that there needed to be more representatives from those in "public and private sectors."

### *Strengths of M&E Officers*

Track20 M&E officers were commended for their accessibility, analytic skills, and positive role in convening multiple stakeholder groups "to rally around a cause"—that is, advancing family planning. Government and NGO colleagues alike complimented officers on their performance. Reflecting Avenir's priority for M&E officers to be embedded within MOHs, respondents commented that those officers had good relationships with senior health officials and with statistics offices.

**M&E officers have broad access to decisionmakers—**in some cases noting "constant contact" with them. From their positions as the "focal point at the MOH for family planning," M&E officers can readily liaise with decisionmakers within their own MOHs, other MOHs, staff from HMISs, national family planning councils, and UNFPA, to name a few examples provided. NGO representatives explained that in many countries there may be highly technical and qualified individuals who can identify gaps, but unless they are part of the system and can "internally bend the books" (i.e., have the ability to motivate change), they cannot make improvements in data collection, quality, or management. They praised the M&E officers for setting up effective processes for data cleaning and review, then giving "feedback to the site when there is lousy data, strange data, et cetera."

**M&E officers are hailed as data advocates** and praised for their ability to "interrogate data," highlight data quality as a key issue, and support data use for decisionmaking. M&E officers must be "in the know" and frequently share their work to generate more demand, stimulate more questions, and continually push others to ask "what else can be known" about family planning this country. For instance, in Zimbabwe, the M&E officer identified a gap in understanding the private sector's role in providing family planning services, and the officer recognized what needed to be outsourced and who should pay for collecting necessary data to fill this gap. In addition to generating questions, M&E officers are seen as a resource to answer questions. In short, "their job is to stand up, sell it [family planning data], and tell others why it's important."

As noted earlier, respondents shared that Track20 officers have elevated the discussion of family planning in the government, and particularly stimulated conversations around the need for good data. As one donor explained, "[The M&E officer] changed the way we discuss family planning." Some governmental decisionmakers heavily rely on M&E officers and wish there were more of them, with one noting, "I wish the state can have more trained M&E officers because honestly . . . whenever she's on leave, anything like that, I panic."

### *Country-Specific Challenges to M&E Officer Effectiveness*

The power of Track20 M&E officers to implement change is dependent on both the context of the country and the individual skill sets of the officers. Stakeholders reported several constraints on M&E officers.

First, several Track20 M&E officers described being **stretched quite thin and busy to the point of being overwhelmed with MOH responsibilities outside of family planning**. They reported feeling incapable of taking on more tasks and, at times, viewing their family planning M&E responsibilities almost as part-time work to be tackled on evenings and weekends.

**Financing affects the scope of M&E officer functions.** With such a multifaceted position, with various financing models, backgrounds, and positions within the government hierarchy among officers, it is not surprising that only a minority of the M&E officers are performing all of the functions listed earlier of an ideal M&E officer: aggregating various data sources, translating them into actionable messages, and helping decisionmakers understand how to use them. Respondents noted that "financing matters": In countries in which Avenir does not pay the salary of M&E officers, it has been challenging to help the officers cover the variety of functions described above. In terms of general officer qualifications, individuals who have worked in HMIS and have conducted data analyses in the past are more able to think about how to integrate various data sources to generate usable information. M&E officers are typically placed one step away from people who direct family planning programs (but are not usually at a senior level themselves).

**Some country leaders lack the background to understand data in general, and FPET more specifically.** Access to FPET is limited to M&E officers in some countries. Even when it is available to a wider audience, users have complained that they can understand the output, but the tool is not intuitive. Non-Track20 staff frequently requested additional training in both FPET and data analysis and use. In many countries, skill levels vary between the national and provincial levels.

In some instances, **Track20 officers felt that they were impeded by internal politics that inhibited the dissemination of their work.** Some M&E officers and their colleagues felt that they play key roles in the MOHs, while others felt that the offices they have been assigned had limited influence and that in government "they want you to do what they say, and not to do what you think is best." Some NGO advocates complained that Track20 officers were a tool of the MOH and should be better integrated into the country's family planning infrastructure. This would include site visits to provincial and district offices, reaching out more frequently than data consensus meetings, and better coordination within MOHs. To be clear, these comments were not consistent in any one country and appeared to be dependent on the institution and individual interviewed.

Finally, **weak data cultures were felt to limit the M&E officers' abilities to accelerate data use.** This includes lack of demand for data, limited recognition of the value of data, and, as mentioned earlier, variable understanding by decisionmakers of the data and their meaning.

#### *M&E Officer Model Beyond Family Planning*

In reaction to the success of the Track20 M&E officer model, and in spite of the challenges described above, many stakeholders felt that similar officers could be used in other areas of the government, including antenatal care, HIV/AIDS, nutrition, education, noncommunicable diseases, civil registration, and generic data analysis for the reams of data collected by the MOH. If M&E officers were less "stovepiped"—that is, less confined to family planning—this might enable potential diversification of funding sources. Respondents from Avenir agreed that M&E officers do not necessarily have to be working in family planning 100 percent of their time—in fact, if working across broader fields like maternal-child health and HIV, there could be a cross-theme innovations and better integration of the M&E officer into the larger system. Several donors suggested that additional M&E officers would facilitate integration among multiple health and development programs. Taking these views together, **the role of M&E officers could usefully expand beyond family planning rather than being "compartmentalized**

**within the FP2020 program."** They could serve as a "bridge" among health sectors, as "M&E is important everywhere."

**Expansion beyond family planning has both advantages and disadvantages.** In current circumstances, using Track20 M&E officers for additional duties and activities has had both positive and negative effects. On the positive side, some countries have expanded the officer's role to encompass all family planning work in the MOH, and these officers represent the MOH domestically and internationally, including at the FP2020 meeting. On the other hand, as noted earlier, the enthusiasm of MOHs for the various potential applications of the M&E officer model has meant that the officers have been asked to support a wide variety of activities, sometimes leaving little time for Track20 work.

## Other Challenges Faced by Track20

In addition to the challenges specific to the M&E officer role, there are additional challenges for the program as a whole.

**FPET estimates are only as strong as the quality of the data inputs.** Several stakeholders expressed reservations about including service statistics in the FPET modeling. Service statistics are considered a key input for the model when deemed to be of sufficient quality, but many stakeholders acknowledged their limitations, specifically their completeness and accuracy. Some even went so far as to say that "service statistics [are] all approximations" and "one data we don't take seriously is the HMIS . . . . it's not to be always trusted." Still, others lauded their usefulness and potential, and views of Burkina Faso's DHIS2 platform were particularly positive: It is implemented nationally, is of high quality and "very well used," is collected every three months, and provides "instant data" on local communities that are available to decisionmakers right after validation.

**M&E officers report challenges with obtaining trust from decisionmakers in FPET's estimates.** Trusting the estimates is vital to their use. A Track20 officer described an example of FPET giving an estimate for one indicator of 38 percent while the mini-DHS's estimate was 40.8 percent, and when "we presented this result to senior management at the ministry, we explained that different data sources were used to come up with our estimate. But they had a hard time accepting our estimate because it was lower. So [this] type of challenge, where decisionmakers didn't accept the evidence, was common at times, and this was a problem."

**The insufficient number of trained staff in many countries highlights the reliance on a single M&E officer.** In several countries, only one individual has been trained in how to use FPET. If he or she leaves, a gap in reporting is likely because of the lack of institutional memory. Non–M&E officer respondents in every country requested training (from Track20) for themselves and provincial officers in FPET and other data analytic methods. Fulfilling this request would build a cadre of individuals within and outside the MOH with similar skills and therefore create redundancy, should the M&E officer move to a different position. Specifically, NGO respondents requested help using the tool to target family planning investments.

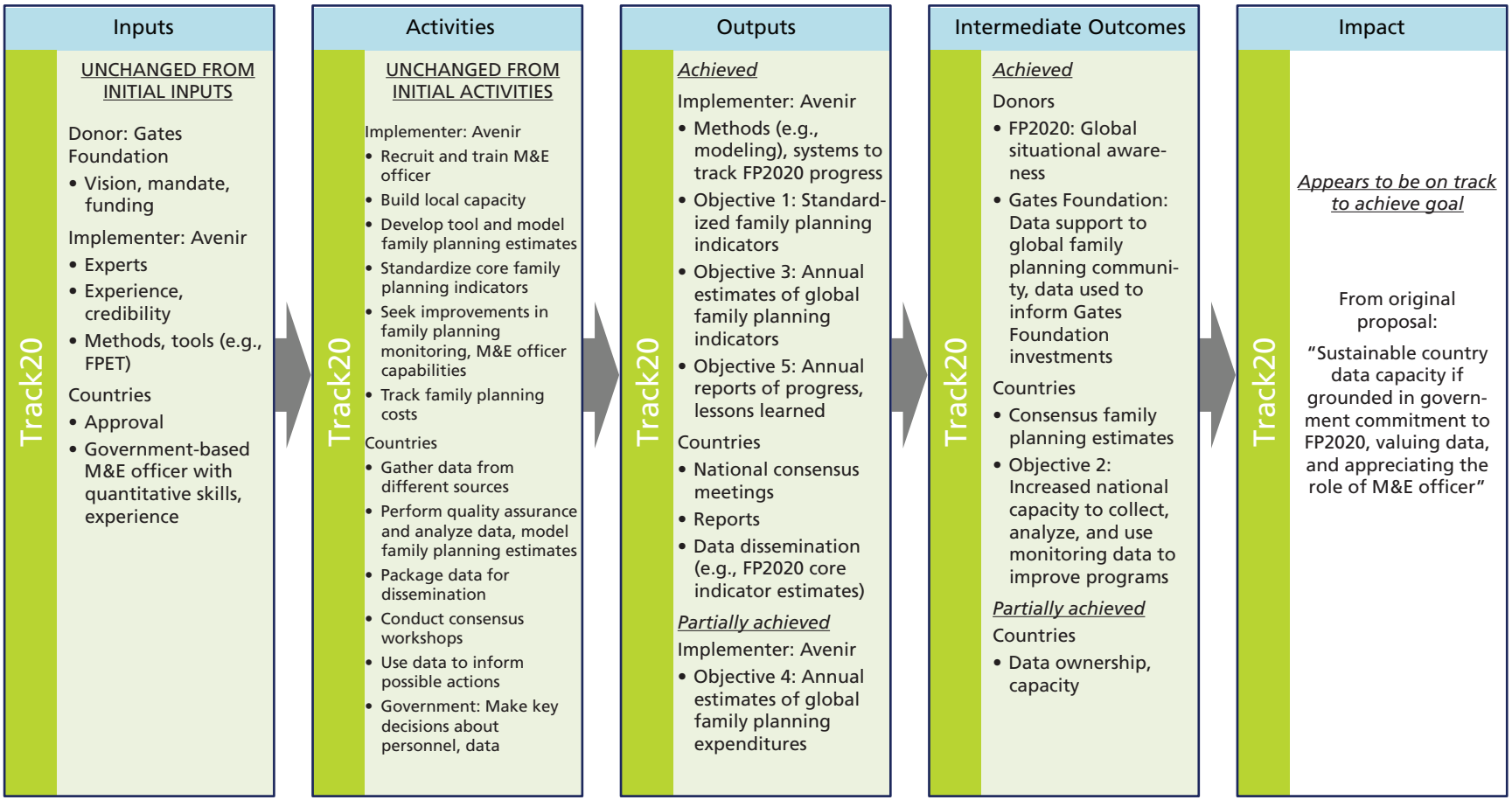
## Summary

Interviews with more than 200 stakeholders from program countries, the Gates Foundation, Avenir, and other U.S.-based respondents reveal that the Track20 approach is well respected and has achieved most of its intended objectives. It is on track to achieve its original goal, as shown in the empiric logic model based on stakeholder feedback (Figure 9.1).

Track20 is now moving on to promising new territory with the potential to make significant, lasting impacts on family planning data collection, analysis, and use around the world. Specifically, an evolving emphasis on country-owned agendas (through the FP Goals model), improving the quality of service statistics and improving the costed implementation plans to make them more realistic and data driven, and embedding M&E officers within MOHs to act as "interpreters" of data to decisionmakers have yielded important dividends. M&E officers face challenges with multiple competing demands, with helping decisionmakers understand their estimates and why data are important, and with navigating government bureaucracies.

Track20 seems to have achieved a strong balance between standardizing, across varying country contexts, a methodologically sound system for producing consensus estimates (using FPET) from available family planning data sources while allowing for a large degree of flexibility and country ownership. This decentralized model has allowed countries to determine what their specific needs are with respect to technical assistance, to decide how to finance the M&E officer(s) in their country, and to define their priorities for improving family planning in their country.

**Figure 9.1**
**Empiric Track20 Logic Model**




RAND *RR2112-9.1*

CHAPTER TEN

# Interactions Between PMA2020 and Track20

## Overview

As noted in Chapter Two, the PMA2020 and Track20 programs were originally intended to be "twinned." However, their activities have not been as well coordinated as intended in most countries, to the point where there is often very little interaction between the two. Our interviews with the Gates Foundation, U.S.-based stakeholders, and in-country stakeholders (in countries where both programs operate) included the following questions: (1) To what extent do PMA2020 and Track20 complement each other to achieve ultimate FP2020 goals? (2) How could the programs work together more effectively?

Respondents had few specific suggestions about how the programs could interact more effectively, but they generally agreed that more communication, shared activities, and contact between PMA2020 and Track20 would benefit each program, and they urged intensified efforts to find common ground.

In this chapter, we summarize respondents' opinions about program interactions.

## Need for a Shared Agenda

Different stakeholders commented on the need for stronger incentives for the two programs to collaborate on the ground. That is, they need a shared agenda to overcome some of the challenges inherent in having two separate grantee institutions in the United States and different institutional partners in program countries. In particular, the fact that PMA2020 principal investigators are based in universities may give rise to the view that "academics are running the show."

**PMA2020 has managed data more centrally than anticipated.** In-country partners had hoped that PMA2020 data would be available much more quickly to monitor their programs because of the smartphone-enabled data collection and rapid data uploading and cleaning. While PMA2020 provides more frequent information in countries where PMA2020 operates, Avenir and Track20 in-country partners thought that they would be able to easily access this information locally, and that is not happening. Both at the country level and at the Avenir headquarters level, they have to refer back to Baltimore to obtain data, "which has been more centrally focused than we expected."

**Data dissemination is not coordinated.** Multiple stakeholders commented on the variable, but typically weak or nonexistent, collaboration between the two programs with regard to data dissemination. Specifically, there was concern that PMA2020 disseminates its own

briefs on what the program called "core indicators," while Track20 looks at all available sources of family planning data to produce estimates of FP2020 indicators. Therefore, Track20 was challenged to make sure that decisionmakers understood that PMA2020's estimates were not the "official" FP2020 indicators, which had to be estimated using a different methodology. To address the need for more collaboration, in fall 2017, the two programs are planning to partner around an initiative to improve the use of service delivery point data. This partnership could yield useful insights into how to improve collaboration between PMA2020 and Track20.

## The Power of a Common Mission

PMA2020 and Track20 staff alike viewed their projects as complementary, with one stating, "PMA2020 helps provinces in which they collect data. Track20 helps the country as a whole." There was an overall call for greater cooperation and collaboration between projects and clearer messaging to decisionmakers around how they interact, such as PMA2020 "feeding into Track20 modeling." Staff from both programs generally agreed that "it would be good if the two programs work together" as "we are on the same team, just two different approaches . . . we are actually speaking the same language." Whether the respondent was a representative of PMA2020 or Track20 determined which program was deemed to need greater participation in the process.

## Highly Variable Interactions Between the Programs in Different Countries

The level of contact between programs differs by country. In one country, contact was so infrequent that the PMA2020 officer, when asked about interacting with Track20, responded, "To tell you the truth, I don't even know who the officer is." In that case, the only contact was during the national data consensus meeting. Some PMA2020 staff demonstrated a deficient understanding of Track20. Some described Track20 as "trying to access the milestone PMA2020 is achieving" or collecting facility-level data on family planning, contraceptive availability, and the quality of health facilities. Others explained that Track20 is collecting data from various sources about expenditures and family planning promotion to inform policymakers and measure impact. These differences vary by country; whereas some country staff were well informed of the other program's goals and worked closely together, others viewed it as a separate program with which they only interacted at the annual data consensus meeting. In Nigeria, a Track20 M&E officer commented that the relationship between the programs was rather weak: "We . . . at the ministry, we should be involved. Even at the planning stage, we should be carried along. If you come to the state, it's embarrassing for me to sit down and we are talking about [my] state, a survey in [my] state and I do not know it [was] held."

Conversely, in other countries, staff reported regular contact between the programs and said, "one can't work without the other." In particular, the programs intersect with respect to their connections with the MOH, with some M&E officers (specifically, in Uganda and Burkina Faso) feeling that they are the "link" between PMA2020 and the MOH. Countries that reported the highest levels of cooperation were those with unambiguous goals from the outset and in which both programs had good relationships with the MOH.

Attendance at the national data consensus meetings was also country-dependent. Some respondents described these meetings as an important opportunity to compare estimates, some reported them to be the only interaction between programs, and some PMA2020 staff members expressed an interest in participating but had not yet been invited.

Despite varied levels of interaction between programs, there was a clear interest in a stronger alliance. Officers from both projects commented that, at present, they only reach out to the team when they are directed to do so. In one country, the Track20 officer was keen to partner with PMA2020 on primary data collection so that it would be easier to incorporate into FPET.

#### Data Sharing Critical but Insufficient

Sharing data between programs was important to both PMA2020 and Track20. Track20 M&E officers in countries where PMA2020 operates noted that they would not have sufficient data for their modeling without PMA2020 and that it is highly beneficial when PMA2020's survey results align with the FPET estimates using service statistics.

Limitations of data sharing were often blamed on differing methodologies or surveys that were not sufficiently powered to be combined with other data sources. In some instances, Track20 officers expressed willingness to help verify and analyze PMA2020 data and open up the same opportunities for PMA2020 staff with respect to Track20 data. As noted earlier, communication is key. Program staff from both PMA2020 and Track20 expressed the need to be updated on the other program's data in a timely fashion and to not have to rely on visits to program websites to learn that data were available.

## Summary

From the beginning, PMA2020 and Track20 were intended to work very closely together. However, they diverged in their implementation in most of the countries we examined. The strength of the collaboration between the two varies by country, but, in general, there is an acknowledgement that more communication, shared activities, and contact between PMA2020 and Track20 would benefit each program.

Respondents had very few concrete suggestions for how the two could interact more effectively beyond involving the PMA2020 staff in the national data consensus meetings. Going forward, it would prove fruitful to intensify efforts to find common ground (as noted previously, one respondent commented, "We are on the same team") and to leverage efficiencies around disseminating family planning data to key decisionmakers.

CHAPTER ELEVEN

# Data Use

## Overview

The ultimate goal of collecting and analyzing data is *using* them to inform decisionmaking. Data use is the last major domain in RAND's data maturity model; we will explore the model in detail in the next chapter. In this chapter, we describe how each stakeholder group viewed the importance of data and how they assessed their country's ability to use data as a basis for decisionmaking.

We encountered a wide range of perspectives on the use of PMA2020 data and Track20 estimates. Respondents had diverse opinions about what facilitates or impedes data use. They were also candid about the capacity of decisionmakers in their own country to understand data well enough to use them effectively.

We begin by describing stakeholder opinions about the importance of data and the capacity to use them. We then cite examples provided by stakeholders of using PMA2020 data and Track20 estimates. In each case, we also describe what respondents viewed as facilitators of and barriers to using these resources.

## The Importance of Data and Capacity to Use Them

Gates Foundation stakeholders noted that "data use is a continuum—you want to be able to generate good data, improve the bad data, then start encouraging their use." Stakeholders felt that in an ideal world, national and subnational-level governments would drive the *demand* for data on which to make decisions, since it is their own programs that would benefit from better data. However, several interviewees felt that, to date, the demand for family planning data has come mainly from the global level (that is, from donors and others).

**The donor-driven demand for data has led to poor integration and use of data within countries**. One Gates Foundation representative noted that "we [at the Gates Foundation] have been both part of the problem and part of the solution" in terms of creating fragmented data systems. He continued, "Everyone wants to measure their own thing" and gave the example of six different partners in the Nigerian state of Kaduna who were all funded through the Gates Foundation and were each conducting a facility survey.

In contrast, in-country stakeholders consistently articulated how important data are for various purposes, even if some decisionmakers "tune out" when presented with data. Comments from in-country stakeholders cluster around several themes. First, stakeholders in most of the 15 countries noted how **data are important to use for program planning and budgeting.**

Respondents in several countries gave examples of developing their costed implementation plans, forecasting their commodity needs using data, and approving state or district budgets based on past performance.

**Respondents emphasized the importance of tracking progress against milestones to ensure accountability.** A government official in Ethiopia said, "We made a commitment at the London Summit, and now we have yearly data . . . so decisionmakers are forced to confront shortcomings in their specific regions . . . this has made decisionmakers take action." The PMA2020 data "puts decisionmakers on their toes" because "what is not tracked and measured doesn't always get done."

**Data are important for advocacy and showing a "pressing need to donors."** Stakeholders in several countries (Burkina Faso, Côte d'Ivoire, DRC, Niger, Nigeria, Tanzania, and Uganda) reflected this view. As a Track20 staff member in Côte d'Ivoire articulated, "In my mind, the estimations give us the power to demand money from decisionmakers, including donors. Without data, you don't have an argument." A stakeholder in Niger noted that family planning targets were revised and accelerated based on data. Finally, interviewees in government positions and NGOs in Côte d'Ivoire, Niger, Nigeria, and Uganda all noted the importance of data for developing "intelligent, informed policies."

**Some interviewees emphasized the legitimizing role of data.** As a government decisionmaker in India noted, for example, "You can tell the story with the data and people will believe. . . . without that data, it's just a story."

**A few stakeholders were quite critical about their country's capacity to use data.** For instance, one commented, "Laos, the country, is not number-driven." Interviewees from several countries (Indonesia, Pakistan, Uganda, and Zimbabwe) noted a shortage of "trained manpower" or "local know-how" around how to use data for effective decisionmaking. Specific needs included a staff member of Track20 in Pakistan who noted that there is no demographer on staff and a government representative from Indonesia who described how young researchers need more training on data analysis and dissemination.

## Examples of Using PMA2020 Data

**The Gates Institute has focused more on data generation than on data advocacy and use.** As discussed in Chapter Two, the fourth objective of the PMA2020 program in the original proposal was to promote the use of data to respond to family planning needs at the community level. But the July 2015 Results Framework developed by the Gates Institute, which identified the four primary outcomes against which they would be held accountable, listed "data generation" but not "data use." As noted earlier, the main focus of the Gates Institute in the initial years of the project has clearly been to quickly ramp up a complex, multi-country survey platform and successfully achieve a proof of concept.

However, Gates Foundation representatives expressed their hope that the PMA2020 grantee views marketing the value of the data as within their scope rather than handing off the data to others to promote their use. They felt that PMA2020 does not provide data that are directly relevant for use by health workers and more "on the ground" potential users but should at least help countries use data for program management. They further emphasized that PMA2020 is a valuable tool for decisionmaking within the Gates Foundation about where to target their resources—to specific countries, regions, or program elements. Interestingly, one

described the Gates Foundation as the largest consumer of PMA2020 data, noting that the Gates Foundation often asks the Gates Institute for custom analyses in order to meet their needs beyond the two-page reports summarizing the PMA2020 data by country.

Other U.S.-based stakeholders expressed similar views: "Data are only good if they are used"; otherwise, it is "all for naught." This statement underscores that **understanding how to use data is an integral part of the entire process.** Rather than expecting PMA2020 to simply collect the data and hand them off to someone else to encourage their use, it was felt that PMA2020 could be doing more to take advantage of opportunities to disseminate the data to local users and make sure that others can analyze key variables on their own.

Several stakeholders saw a natural division of labor between PMA2020 as the data generators and Track20 as the promoters of data use (or "data advocates," in one person's words). The Gates Institute shares the view that it is mainly responsible for generating data and observed that it has a standard process by which to disseminate them on a certain timeline. But it feels that its role stops short of advocating for or directly facilitating the use of PMA2020 data. However, the Gates Institute has begun to convene two kinds of workshops in countries: to disseminate data (e.g., to government and other decisionmakers) and to help in-country researchers analyze the PMA2020 data, with the goal of publishing their work; the Gates Institute regularly hosts the PMA2020 principal investigators in Baltimore to do the same.

Our interviews showed that while the two-page reports are disseminated widely to academics and non-academics, PMA2020 data sets themselves seem to be used mainly by academics, including substantial use of PMA2020 "micro–data sets" by people outside the country. Importantly, PMA2020 data are incorporated into the FPET model by Track20 staff as one of several data sources used in their Bayesian modeling to generate national estimates, and both PMA2020 and Track20 share the data with in-country decisionmakers who are involved in family planning (with this effort being a focus for Track20). As such, several stakeholders commented that PMA2020 is viewed as externally driven or imposed, with the data primarily managed and released by Johns Hopkins University or the local university, not a government ministry.

**"Data use" means different things to different people.** During in-country interviews, we asked respondents to provide examples of using PMA2020 data, initially keeping the question open-ended and then emphasizing data use *for decisionmaking.* They noted several different examples of data use.

PMA2020 data sets are analyzed for multiple purposes. Students and other researchers use them for **academic purposes**. Use of PMA2020 data was described as being "measured by the number of requests made, and by the number of data downloaded." When asked whether the researchers who are requesting and accessing the data are using them for purposes other than decisionmaking, the reply was "I have no idea about that." The surveys are also viewed as a key source of data for **developing costed implementation plans** and are used for the purpose of **advocacy**, such as for securing the inclusion of family planning in Ghana's National Health Insurance and requesting funding from donors. Furthermore, the survey data contribute to **designing and implementing programs, as well as assessing their effectiveness.** For instance, in Uganda, PMA2020 data were used to evaluate a program that aimed to increase uptake of Sayana Press (a subcutaneous injectable hormonal contraceptive). Similarly, in DRC, PMA2020 data were used to assess the effectiveness of a campaign to advertise the Sayana Press method on billboards throughout a particular region of the country.

Several respondents gave specific examples of using PMA2020 data to **track and address commodity stock-outs,** such as communicating to the MOH the need for improved supply chain management. While not a dominant theme (despite "Accountability" appearing in its name), a few stakeholders mentioned that PMA2020 data **ensure accountability** on the part of policymakers and allow countries to track their progress on an annual basis. Finally, PMA2020 surveys are used to **improve the quality of family planning services.** Respondents cited examples of finding solutions to the survey finding that poor-quality counseling was a problem in Ethiopia and, in Kenya, using the data to prompt an investigation into the reason behind the finding that "38 percent of our public health facilities do not provide IUDs, yet we only have 2-percent stock-out." The Kenyan Minister of Health developed an action plan to address this finding from the PMA2020 survey, which included interviewing the directors of the facilities that are not providing IUDs. The directors reported that it was a "skills issue," so the minister began working to provide resources in the current budget for skills training for the affected facilities. In summary, "The leaders ask themselves further question[s] and then make decisions based on data."

## Facilitators of and Barriers to Using PMA2020 Data

We asked in-country stakeholders to comment on facilitators of and barriers to using PMA2020 data. While some general themes emerged, it was challenging to elicit specific examples, despite targeted probes. In this section, we describe facilitators of and barriers to PMA2020 data use that emerged from in-country interviews.

One theme emerged clearly from several countries—Ethiopia, Ghana, India, and Nigeria: **A key facilitator of PMA2020 data use is the ability of PMA2020 data to meet the data needs of decisionmakers** (see Chapter Five). Stakeholders cited specific characteristics of PMA2020 that make it useful to decisionmakers: the annual frequency; the inclusion of service delivery point data; the ability to parse out different populations, such as adolescents, rural versus urban women, and married women; the subnational estimates (when available); and the inclusion of quality indicators. Conversely, **a barrier to PMA2020 data use is the perception expressed by several in-country stakeholders that the data could do a better job of meeting data users' needs**. This barrier was typically discussed in the context of decisionmakers needing subnational estimates. For instance, in Indonesia, family planning data are needed at the district level. In Nigeria, the PMA2020-affiliated respondent lamented that the government had repeatedly made requests for data at the local level, which "PMA2020 is not powered to do." The commissioner said that if PMA2020 "can give him the information at the local government level, he'll do something but he can't do it for the entire state because he doesn't have the funding to do so."

Another repeated theme was **the importance of having strong connections to policymakers and, relatedly, the importance of deliberate dissemination efforts**. Respondents from several countries (Burkina Faso, DRC, Nigeria, and Uganda) described how the PMA2020 program disseminates its work at key family planning–related committee meetings and conferences. Relationships with the Ministry of Health were another key facilitator, with a particular emphasis in Nigeria on close relationships at the state level rather than federal. In Niger, the PMA2020 team noted that the MOH and NGOs call them to ask for the latest results, facilitating their use, and PMA2020 staff in India described having free access to the

MOH in the state where it operates, Rajasthan. A stakeholder from Burkina Faso noted that the connection between PMA2020 and the MOH was strengthened by the presence of both on the technical working group for reproductive health, which meets regularly. Several stakeholders in Ethiopia, both within PMA2020 and the MOH, noted the "good working relationship" between PMA2020 and the MOH.

Similarly, **stakeholders viewed weak connections to policymakers as a barrier to PMA2020 data use**. As a respondent from USAID in Ghana noted, "I think the problem with PMA2020 is that they have not been able to necessarily break through to the health sector widely. I know they've been trying, they've been trying to present at the Ministry of Health level, at the health sector working group level." In Ghana, a PMA2020 staff member stated that they have a very good relationship with the Ghana Health Service, the autonomous agency under the MOH responsible for service delivery that uses the PMA2020 data to track family planning service provision and uptake. However, the relationship between PMA2020 and the higher levels of the MOH is not as strong. Respondents in Niger noted that the MOH was not involved in the planning and execution of PMA2020, there were disagreements about the desired sample size, and PMA2020 is still working on getting the MOH to "accept the data" and buy in to the benefits of its approach. (Of note, after the RAND team member's visit to Niger, we learned that the MOH did officially recognize the data in May 2017, indicating that it would be using the PMA2020 national estimates going forward as part of its work with Track20.)

Among in-country stakeholders, **additional facilitators of PMA2020 data use were the clear standards for who can use PMA2020 data and the clear explanation of procedures to access the data sets on the PMA2020 website**. In addition, the two-page reports prepared by the Baltimore team were praised for being "very nicely done" and **presenting complex data in a visually appealing way** so that they are easily referenced by government officials.

**Trust in the PMA2020 data, decisionmaker buy-in** (both from health and religious leaders), **and partnerships with other key stakeholders in program countries were other facilitators of PMA2020 data use.** Stakeholders in Uganda and Ethiopia mentioned **a lack of trust in the data quality of PMA2020 as hindering their use**. With respect to decisionmaker buy-in, a Kenyan government official observed that PMA2020 data showed the same trend in mCPR as other data sources with which he or she was more familiar, which led him or her to trust the PMA2020 data. The emir of Kano State (Nigeria) was present at a PMA2020 dissemination meeting and was wholeheartedly supportive of the survey. The Ugandan PMA2020 staff noted that their collaboration with the Ugandan Bureau of Statistics gives them legitimacy in the eyes of the MOH and other stakeholders, and if they had not formed that partnership, the impression would be "Oh, you guys are just independent researchers!" and decisionmakers would be less motivated to use PMA2020 data. This opinion aligns with comments from a Gates Foundation interviewee who noted that countries should work through their own bureaus of statistics so that PMA2020 is trusted and not seen as "top down."

In terms of partnerships with other stakeholders in country, a PMA2020 leader in Nigeria described efforts to build more rapport with other stakeholders, including NGOs, the Advance Family Planning program, decisionmakers, and Track20 partners: "We are now sharing more and connecting more with them." Again, specifics on how these partnerships have led to concrete examples of PMA2020 data use are elusive, but the sense of the respondents is that these partnerships are a key facilitator of PMA2020 data use or, at the very least, awareness.

On the other hand, **the most commonly cited barrier to using PMA2020 data—**noted by both PMA2020 staff and decisionmakers themselves**—was that decisionmakers, researchers, and the general public did not know about this relatively new data source**, a disconnect that underscored the need for more effective dissemination. PMA2020 program staff, particularly in Ethiopia, Ghana, and Nigeria, noted that their own dissemination efforts could be strengthened because while "access is free—anybody can use it at any time. The problem is people do not know" about the data. As the PMA2020 leadership expressed in Ghana, "We need to find better ways of packaging the data for different subpopulations, and different actors, and different levels of decisionmakers so that they'll take the data we have and use them." In DRC, a government representative wanted the program to better communicate its goals and objectives, not just its data.

The PMA2020 team in Nigeria provided the following assessment of how their dissemination efforts had evolved: "Initially, like our first two rounds, we're very busy trying to get to grips with the project itself so we were not doing much [dissemination to end users]. . . . honestly, by the time we'd finish the field work and analyses, there was no energy left. But now we are growing and there's more of this [dissemination] happening."

Some respondents said that being inundated with family planning data is a barrier to PMA2020 data use (DRC, Indonesia). However, more commonly, **respondents across several countries cited a lack of capacity and training in how to use data and an overall weak "data culture" as critical barriers.** For instance, an NGO representative from DRC described how "many decisionmakers can quote the data but may not actually use it to inform their decisions" because they have not "internalized" it. Decisionmakers in Ethiopia were perceived as "setting their own annual plans" that are "not actually [derived] from evidence," and in Nigeria, an NGO representative took it one step further: "The gap is in whether they even know . . . why they need the data."

While the RAND team asked the in-country stakeholders to comment primarily on facilitators of and barriers to using PMA2020 data, some U.S.-based stakeholders commented on **problems with the accessibility and usability of PMA2020 data files and other tools on the PMA2020 website by researchers for analytic purposes.** From the researcher perspective, several critiques emerged.

- In a few cases, variable names differ from country to country (e.g., the variable name for the enumeration areas in Nigeria differs from that used in Ghana), making cross-country comparisons somewhat more challenging. While these differences are noted in the codebook, it is more challenging to write code that can be applied across multiple countries.
- In contrast with efforts to maximize data comparability between the DHS and MICS for analytic purposes, it was felt that there was not the "same effort at consistency with other surveys occurring for PMA2020."
- The PMA2020 survey questionnaire does not contain variable names, which would enable researchers to easily identify the questions asked for each variable in the data set.
- Several stakeholders noted that documentation for (research) users could be clearer, particularly related to the codebook and around how weighting was accomplished.
- Another critique was related to the disclaimer that appears with data downloads in the user notes: "*Disclaimer: PMA2020 cannot provide in-depth support for data analysis or data related questions, however, to assist the end-user, explanation of some variables is provided below."* As one individual who uses PMA2020 for analysis said, "It sounds like they're

collecting [data] for internal use and they're sharing but not providing help . . . . But if they want others to get to use it, then this is insufficient."
- Finally, while several stakeholders compared PMA2020's Datalab (PMA2020, 2016b) favorably to the DHS's StatCompiler (DHS Program STATcompiler, undated), one respondent noted that the data download button was nonfunctional (which was confirmed by the RAND team when performing the statistical analyses described in Chapter Seven), a concrete barrier to using the data.

## Examples of Using Track20 Estimates

Gates Foundation and U.S.-based stakeholders did not comment on use of Track20 estimates, focusing instead on use of PMA2020 data. However, in-country respondents were asked to provide examples of using Track20 estimates for decisionmaking, and, as with questions about using PMA2020 data, they gave responses that varied in specificity and concreteness.

The **most specific example of data use for decisionmaking** came from a Track20 staff member in Nigeria. The Track20 FP Goals model was used to help Kaduna State **prioritize family planning activities**. As the Track20 member described it, they solicited baseline information from 2015, put into the model, and "the model now says, 'Okay, from all the activities that [have] been happening, this is the impact we think demand creation is going to have if they're going to get to their goal.' It came up with three scenarios, A, B, C." Scenario A was the status quo with no interventions. Scenario B suggested moderate interventions, and Scenario C is fully implementing the costed implementation plan: "All the resources are available, you're doing everything. This is where you're going to be in 2020. The state will now decide which of the scenarios they want to go with." Kaduna chose to implement Scenario B, and it was based on Track20 modeling.

There were numerous nonspecific mentions of using Track20 estimates for "activity planning" (Lao PDR); improving and prioritizing programming, such as for postpartum family planning and Sayana Press uptake (e.g., Nigeria, Uganda); targeting interventions to specific populations, such as adolescents and rural women (Tanzania) or low-performing regions (Ethiopia); revising and accelerating family planning targets (Burkina Faso, Kenya, and Niger); and incorporating the estimates into state and federal costed implementation plans (Nigeria, Pakistan, Tanzania, Uganda, and Zimbabwe).

Stakeholders gave several examples of Track20 estimates being used to **solicit funding and other advocacy purposes**, including cabinet members using Track20 estimates to solicit donor support for various family planning initiatives and convincing the Ministry of Budget to fund family planning initiatives (Côte d'Ivoire); a Nigerian legislator being so impressed by the Track20 estimates that he vowed to become a champion of family planning and protect it in the next budget period; and a governor in one county in Kenya allocating 5 million Kenyan shillings specifically for family planning, as compared with other years, when they "never had a single cent from the county government, and this is because of the data we have from the Track20."

As with PMA2020, in-country respondents in Ethiopia, Lao PDR, and Nigeria commented that they use Track20 estimates to **evaluate stock-outs and improve logistics so that women have access to necessary commodities**.

Finally, respondents observed that a clear role for Track20 estimates was that they served as a harmonized figure around which there was consensus, "because it is very important for the country to speak in one voice." Stakeholders appreciated that Track20's FPET took into account **disparate survey results and service statistics to produce more "realistic" and "high-quality" estimates than one single survey might produce** (Indonesia, Nigeria, Uganda). Decisionmakers described feeling flooded with data at times and "bogged down" by a lot of uncertainty around which source to use, so they found it immensely useful to have a rigorous model that incorporates all available data to produce a best estimate. However, there was little elaboration on how those harmonized consensus estimates were then put to use.

## Facilitators of and Barriers to Using Track20 Estimates

In-country respondents described facilitators of and barriers to using Track20 estimates. We summarize their comments below.

Respondents commented that **decisionmakers' use of Track20 estimates was facilitated by national data consensus meetings,** which raised awareness of the Track20 methodology, presented and debated the latest estimates, delved into the sources of the FPET-produced estimates, examined the quality of routine service statistics, and more. Respondents in DRC, Kenya, and Pakistan specifically mentioned these meetings as a critical catalyst of Track20 estimate use because they typically have a diverse list of stakeholders in attendance. As a Kenyan NGO representative commented, "the space that the consensus meeting creates is critical because it's one thing for a government department to sit down and chant data and say, 'All right, Kenya, this is your data. This year we shall only measure mCPR for married women. Sorry if you're interested in all women or sorry if you're interested in adolescents.' This creates an opportunity for informed consent and informed decisionmaking by a broader audience." Perhaps the strongest endorsement for the importance of these consensus meetings is the view that the process of establishing consensus after vigorous debate "creates buy-in because when those indicators are decided, when the data is decided, it's not for one person. It's not for these FP20[20] thing that has come from abroad. It's really a decision that's made jointly with the group."

On the other hand, while several stakeholders praised Track20's efforts to build awareness through dissemination efforts such as the national data consensus meetings, **in some countries, stakeholders felt that lack of awareness was a barrier to data use**. In addition to the need to increase awareness (noted in India, Niger, and Uganda), a related challenge is obtaining buy-in from decisionmakers. Specifically, a Track20 staff member in Indonesia described how it is a struggle to get decisionmakers, particularly in the National Development Planning Agency, to "believe" and "accept" the FPET estimates, that "we can forecast for one point from the many kinds of surveys."

Another identified facilitator is the **presence of a champion of the Track20 methodology and resultant estimates.** One example of a champion in Nigeria has already been mentioned—the legislator who vowed, after hearing Track20 present its data, to protect family planning in the next budget cycle. Another example is a Nigerian government official who described how "there's no way we can do anything, in Lagos state at least, that we don't refer to data. The honorable commissioner for health is somebody that loves data. Evidence-based activities, if

you write a memo or anything to him and you don't put data, you can't even do it, because you know you have to back it up with data."

Stakeholders in several countries, such as India and DRC, cited the **deliberate embedding of the Track20-supported M&E officers within the MOH as a key facilitator of promoting use of Track20 estimates** because of their access to decisionmakers.

**Another facilitator of Track20 use is a general increase in the demand for data** and, relatedly, a "clear transformation from the previous less evidence-based use of data." One Kenyan NGO representative again referenced the National Data Consensus Meeting, but this time for creating "a better understanding, especially of policymakers, on the implications of what this data is telling them. . . . it really has created that increased visibility of the data around family planning in Kenya." This individual elaborated, "I would say that the culture around data use has improved. . . . On a routine basis you will see referencing of data" in action plans, progress reports, and presentations by government officials. Respondents called for "indicators to measure the extent of the use of evidence for decisionmaking," as they certainly perceive that it is changing. In fact, the RAND data maturity framework was created to supply just that: a semi-quantitative measure of the extent to which evidence is used for decisionmaking. (The framework and analysis of ratings are discussed in the next chapter.)

Conversely, stakeholders in several countries (Burkina Faso, Indonesia, Lao PDR, Niger, Nigeria, and Uganda) were quite **critical about the strength of the "data culture" in their countries, observing that a lack of demand for data and appreciation for why data are needed** hinders the use of Track20 estimates. For instance, respondents noted that **decisionmakers may be aware of the data but did not always know to use them** and that **they need to get more accustomed to data-based decisions**. A Nigerian USAID representative asked, self-critically, "How often do we demand family planning information for either decisionmaking or for awareness creation or for budgeting?"

As with PMA2020, another facilitator of Track20 use is the **ability to communicate data to decisionmakers through charts and other diagrams that quickly convey key information**. Several Track20 staff noted that ministers are too busy to look at more than a page with a bar or pie chart; others noted that they provided diagrams to other ministries and then were gratified to see Track20's diagrams show up in their presentations.

Several stakeholders highlighted the **incorporation of service statistics into the FPET model and noted efforts to improve their quality as key facilitators of the use of, and utility of, Track20 estimates.** While Track20 staff acknowledged that service statistics are "not perfect," they are "hands-on data information for us, so that we can track our progress on a monthly basis." Others noted that service statistics clearly meet the needs of decisionmakers at a more local level and that there is a growing appreciation of their utility in tracking progress around "implementing our family planning program and making decisions."

Respondents commented on the **need for more capacity-building around the FPET tool's methodology** in order to increase decisionmakers' use of Track20 estimates, as it is viewed as somewhat of a "black box." A governmental representative from the Pakistani Statistics Bureau, a Track20 staff member from Indonesia, an Indonesian government official, and a Tanzanian NGO representative all stated that they had a basic understanding of the FPET tool, but they wanted more training on the tool to be more self-sufficient, and they wanted to understand more deeply the theoretical basis underpinning the Bayesian modeling. Even M&E officers expressed a desire for more information on how, exactly, the modeled estimates were calculated so that they would be able to defend the estimates to decisionmakers and donors.

A final barrier to the fully realized use of Track20 estimates, which echoes the non–program-specific data needs of decisionmakers described in Chapter Five, is that **decisionmakers want subnational Track20 estimates**. A Track20 staff member from Nigeria offered the following assessment about whether the estimates for the 17 core family planning indicators produced by the FPET tool are being used by decisionmakers: "For now it's not being used by anybody. . . . It's not being used by them in the sense that we're in a country where we're so large . . . and . . . their target audience for now is just on two states. . . . They don't represent the whole country. . . . There's a need for them to actually spread this information across the country. Not just limited to these states."

## Summary

We asked each stakeholder group how important they thought data were in decisionmaking and how prepared decisionmakers were in their respective countries to understand and use data. We also asked them to describe facilitators and impediments to using PMA2020 data and Track20 estimates and to give us examples of situations in which those resources had been used.

Stakeholder views about the importance of data use varied substantially by stakeholder group and by country. In general, Gates Foundation staff felt that the demand for data had come from the global level, making it more difficult to integrate data at the country level. However, in-country stakeholders consistently asserted that the data were invaluable for a variety of uses, including, for example, program planning, development of a costed implementation plan, and forecasting commodity needs. Data were also seen as essential to tracking progress against milestones and for demonstrating "pressing needs" to donors. In-country respondents were candid about the inadequate capacity of decisionmakers in their country to interpret data so that they could use them when making decisions, or even so that they could follow data presentations from PMA2020 or Track20 staff or from other experts.

Some respondents saw a natural division between PMA2020 as data generators and Track20 as promoters of data use. The Gates Institute shares the view that its responsibilities include collecting *and* disseminating its data but stop short of advocating for their use for in-country decisionmaking. However, it is convening workshops and hosting meetings to help in-country researchers conduct and publish data analyses, and PMA2020 and Track20 are piloting a collaboration to promote use of service delivery point data.

Respondents commented in general terms but provided few concrete examples of using either PMA2020 data or Track20 estimates for decisionmaking. Some of the more specific examples included activity planning, prioritizing family planning activities, developing operational plans and costed implementation plans, evaluating stock-outs and improving logistics, and soliciting funding. A clearly articulated role for Track20's estimates was that they harmonize disparate survey and service statistics to produce more realistic and better-quality estimates than any one survey could produce, but respondents often did not elaborate further on how these harmonized estimates were then *used*.

When asked to describe facilitators of and barriers to using PMA2020 data, in-country respondents mentioned meeting decisionmakers' needs as facilitating use; on the other hand, the perception that the data could be more responsive to their needs was a barrier. Other facilitators included having strong connections with policymakers; weak connections were seen as a

barrier. Trust in the PMA2020 data and partnerships with other key stakeholders in program countries were essential to data use. The most commonly cited barrier was a general lack of awareness that the PMA2020 data existed, highlighting the need for more effective dissemination. Having an in-country champion of the Track20 methodology helped to promote use of Track20 estimates.

Multiple respondents thought that receptivity to using data for decisionmaking had increased—what they referred to as an improved data culture. However, not all respondents shared that view. Ultimately, they thought that use of PMA2020 data and Track20 estimates was hindered by a lack of demand for data and a lack of appreciation for why data are needed.

CHAPTER TWELVE

# Data Maturity

## Overview

The Gates Foundation asked RAND to develop a maturity framework to help understand where countries are along the pathway toward mature data systems. Our framework is organized within three broad domains: **organizational readiness** (including staffing, leadership and staff buy-in, and infrastructure), **data systems** (including data collection, data management, data analytics, data governance, and institutionalization), and **data use**. We used the framework to assess data maturity for PMA2020 and Track20; it could be used in the future as a tool to periodically take stock of countries' further progress toward increasingly mature data systems.

Data use, as a critical part of data maturity, was described in the previous chapter. This chapter presents both qualitatively and quantitatively depicted perspectives of all three domains of data maturity—i.e. the components that lead to high-quality data, ripe for use for decision-making and action.

Here we first present key themes (albeit not an exhaustive list) impacting data maturity, as described by a wide range of stakeholders (i.e., staff affiliated with PMA2020 and Track20 programs, government officials, members of bilateral/multilateral organizations, and NGO representatives) who participated in semistructured interviews during in-country site visits. Because this analysis focuses on the on-the-ground processes of implementing PMA2020 and Track20, we present only responses from in-country stakeholders.

After summarizing qualitative findings from those interviews, we discuss our analysis of the data maturity ratings of specific program-level aspects of data maturity by stakeholder group and country. The complexity of data maturity lends itself to this complementary mixed-methods analysis: The interviews provide rich, open-ended contextual data, supplemented by more-structured ratings of specific factors. Together, both illustrate where these programs and the countries in which they are being implemented fall along the continuum of data maturity.

## Qualitative Analyses of Interviews with In-Country Stakeholders

### PMA2020

This section describes factors impacting data maturity of PMA2020, as identified through interviews with in-country stakeholders.

### *Organizational Readiness*

Organizational readiness encompasses staffing, buy-in and communication among staff, physical infrastructure and resources, and leadership buy-in. Here we consider mainly staffing because we did not ask specifically about infrastructure in our interviews (we captured that factor instead in the data maturity ratings, which are described later in this chapter); leadership buy-in is aligned more closely with the analyses of program sustainability, which are discussed in Chapter Thirteen.

**Staffing.** Stakeholders across countries described a partial reliance on well-respected "champions"—often professors, lecturers, and researchers within academic spheres—to enhance the institutionalization of PMA2020. For example, PMA2020 in Indonesia is housed within the government body, the National Family Planning Coordination Board, but its *de facto* home is in three geographically dispersed universities. Academics, who are officially apolitical, are seen as more trustworthy and accessible, as one stakeholder from a multilateral organization in Uganda explained.

But one champion cannot sustain an undertaking like PMA2020 alone, and respondents cited the difficulty of hiring adequate staff and the need for more comprehensive and skilled teams as barriers to data maturity. Inconsistency in hiring female resident enumerators was also cited as a barrier in DRC, Ethiopia, Ghana, India, Indonesia, and Nigeria. As one respondent in Nigeria explained, "I mean, life happens. Sometimes, we have resident enumerators who need to move on. They go to school. They get married, and so we have fewer and fewer people, but we don't have enough funding to do a fresh recruitment and training." In addition, respondents in DRC and India called out the need for a broader team, including administrative assistants, IT specialists, and data analysts.

**Partnerships.** For Indonesia, Kenya, Nigeria, and Uganda, the foundation of organizational readiness was building partnerships with well-established institutions, which enhanced the trustworthiness of PMA2020 to key stakeholders. For instance, PMA2020 in Kenya and Uganda is directly tied to the Bureau of Statistics. In addition, having a well-integrated, apolitical team of researchers was cited as a means of strengthening trust in its findings as well as PMA2020's agility, or ability to make necessary "course corrections," as one representative from Niger reported. However, poor dissemination of PMA2020 data was cited by respondents in DRC, Ghana, India, Indonesia, Nigeria, and Uganda as a significant barrier, which was attributed to both a lack of funding and limited skills to communicate survey results. Despite these barriers, PMA2020-affiliated staff, such as one respondent in Nigeria, reported being, "quite happy with where we are . . . we are going to add another 80 clusters in Oyo State. . . . You need a big, competent team to look at that." In other words, PMA2020 teams understand their limitations and are aware of how increased organizational capacity would help enhance overall data maturity.

### *Data Systems*

**Data Collection.** PMA2020 data collection through the use of female resident enumerators and smartphone technology was largely seen as a novel, exciting, efficient methodology, which echoes the finding from Chapter Six on the goals and advantages of PMA2020. Data collection was furthermore facilitated by concerted efforts to build rapport with community leaders in the enumeration areas. Stakeholders in Burkina Faso, Ethiopia, Ghana, India, Nigeria, and Uganda specifically referenced rapport-building in the enumeration areas as a means of conveying confidence in the survey, convincing residents that the survey reflects local dialects and

cultural customs, enhancing protection and security for the resident enumerators (as well as the expensive PMA2020 equipment), and increasing response rates.

The RAND team had the unique opportunity of witnessing the benefits of this rapport firsthand while accompanying a resident enumerator in Uganda through a peri-urban enumeration area. Although the enumerator resides close to the enumeration area, the slum-like conditions presented a challenge in terms of both data collection and her personal security; however, her ability to rely on and trust a local female elder who owned a popular shop contributed to her high response rates in the area and her overall job satisfaction. An additional facilitator of data collection was the active role of supervisors, who intervened when problems arose, particularly in relation to accessing data from service delivery points, which is done exclusively by supervisors in multiple countries. Furthermore, supervisors in Uganda communicate frequently via mobile chat applications to troubleshoot and share best practices.

Respondents also referenced a broad range of factors that stymied data collection, including geographical issues (e.g., limited mobility in mountainous regions and during rainy seasons, general lack of infrastructure such as electricity and roads) (Ethiopia, Ghana, Nigeria, and Uganda); language/dialect issues (mentioned in Chapter Six), which can present last-minute stoppages in the deployment of PMA2020 and lead to misunderstandings (Ethiopia, India, Indonesia, Nigeria, and Uganda); poorly delineated enumeration areas (Ethiopia, Ghana, and Nigeria); and overall suboptimal response rates due to the sensitivity of the questions (Niger), reluctance of adolescents to answer questions about family planning in front of their parents (Indonesia), low response rates at opposite ends of the socioeconomic-status spectrum (Ghana and Uganda), respondent fatigue (Nigeria), and resident enumerators altering the ages of potential respondents so as to avoid having to administer the survey to more members of the household (Nigeria). In addition, the reluctance of service delivery points to provide responses presented a particular challenge in Ghana, Indonesia, Nigeria, and Uganda. These issues were tied to a bureaucratic approvals process to access data, limited staff availability to respond, a hesitancy to share finances, and a lack of incentive, which one respondent from Uganda described as "The government won't provide the stocks after all [even if stock-outs are shown], so why should we give you the data?"

**Technology.** Although technology is crucial for both PMA2020 data collection and the data analysis, respondents described technology mainly as it pertains to data collection. As stated, the mobile phone technology is seen as an overall advantage of the PMA2020 platform. Stakeholders also reported enhancing their use of technology through geographic information system training, in addition to using backup data storage systems, such as Google Drive, and supplying extra power banks and local mobile SIM cards. Poor connectivity issues were cited in DRC, Ethiopia, Ghana, Nigeria, and Uganda, which often lead to shortcomings in data transfers from the field back to the central PMA2020 team.

**Data Management.** PMA2020 teams took great precautions to ensure high data quality. For instance, PMA2020 principal investigators and supervisors described monitoring the time each resident enumerator spent administering the survey; if the time spent was deemed too short, she was called in for an audit. Ghana, Ethiopia, India, Indonesia, Nigeria, and Uganda all instituted quality control checks, ranging from an independent evaluation team to having data managers in each region represented within the PMA2020 countries. Interestingly, only Nigeria cited PMA2020's DataLab as a facilitator of data maturity. The prominent barrier of data management was difficulty downloading data in real time and finding time to perform regular data cleaning (Ethiopia, Indonesia, and Nigeria).

**Data Analytics.** Across the board, the successful analysis of PMA2020 data was greatly facilitated by regular discussions with the Gates Institute at the Johns Hopkins University, which enhanced in-country data analytics capabilities and quick resolution of any discrepancies that arose. However, Ethiopia, Indonesia, Kenya, and Uganda also expressed concerns surrounding the capacity of their researchers to perform data analysis and the need for more training and time to perform higher-level analyses and really obtain "the wow factor" with the data, as one principal investigator expressed.

**Data Governance.** PMA2020 stakeholders discussed two facilitators to data governance. In Uganda, extra precautions were taken to implement automatic erasure of data from smartphones once they were uploaded. The Ugandan stakeholders also mentioned the straightforward process of accessing PMA2020 data sets, as did the Indonesian stakeholders—Indonesian students are highly encouraged to perform thesis work with PMA2020 data. Respondents also expressed concerns over the confidentiality and protection of the data due to the cloud-based server and fact that data can be readily accessed by the Gates Institute. Stakeholders in Côte d'Ivoire, Ethiopia, and Uganda stated this concern, and respondents in Indonesia further noted that they are mandated by law to house the data exclusively at the National Family Planning Coordinating Board.

**Data Institutionalization.** The process through which PMA2020 data become part and parcel of a country's family planning programming is complex and often impacted by stakeholders within the system. Involving key stakeholders in the process of interpreting and presenting PMA2020 helped to concretize PMA2020 data within a country, as stated by respondents in Niger, Nigeria, and Uganda. A respondent in Nigeria stated, "We visit every state, invite stakeholders, including state ministries of health, ministries of planning, ministries of education, ministries of water resources, ministries of rural and community development, the local government councils, the traditional rulers, the religious rulers, and, in fact, even the academics. . . . We invite also development partners working in reproductive health, and during these dissemination workshops, our findings are subjected to rigorous discussions, and we learn so much from such discussions."

On other end of the spectrum, poor or inconsistent dissemination to stakeholders hindered the institutionalization of PMA2020 data. For example, a respondent in Ethiopia cited that there was no way of knowing whether PMA2020 data sets had been downloaded. Another respondent from Ghana reported, "We've only had one national dissemination that was funded by PMA." Lastly, a respondent from Indonesia reported, "If the indicators look good, then they [decisionmakers] say, 'Good.' If it's not a good number, then they say, 'Oh, this [PMA2020] is foolish. There are not enough resources to do the survey.'"

### Track20

In-country program stakeholders offered their opinions about factors that impact the data maturity of Track20.

#### *Organizational Readiness*

Interestingly, no facilitators of organizational readiness were cited, although observations revealed support for the activities of the Track20 M&E officer in terms of space and resources, as well as in fostering connections to stakeholders. Barriers included a lack of basic resources for Track20 operations. For example, a respondent in Côte d'Ivoire stated, "We don't have an office to work in—we squat," and another in Nigeria said, "I'm talking about [needing] basic

things in the office." Other barriers included a lack of role clarity among Track20 personnel, which was a significant issue in Nigeria, as well as a poor understanding of Track20's overall goal and mission. This issue is exacerbated by turnover within MOHs and parliament, where Track20 stakeholders must routinely convince new members of their mission.

It is important to reiterate that countries are implementing a novel and sophisticated program in difficult organizational contexts and may find themselves still at the early stages of understanding and implementing it. One respondent from DRC stated poignantly, "What can I say I'm proud of? For now, I can't really say because, it's not good when you are trained and then you really still don't know [what to do] . . . . I don't know how to put it. I've not really seen the essence of the training. . . . Everything is just new."

**Staffing.** While user friendly, the sophisticated modeling techniques of Track20, coupled with the challenges of promoting family planning, require a particular blend of staffing and personnel support. Thus, one can commiserate with the respondent from DRC who laments having limitations with Track20. The variation in the skill sets, backgrounds, and expertise of Track20 M&E officers impacts the data maturity of Track20, a fact well understood by the Gates Foundation and Avenir, and is discussed in Chapter Nine of this report. In addition to interpersonal politics, respondents in Côte d'Ivoire, Indonesia, Nigeria, Tanzania, and Uganda cited a shortage of experts who can perform the tasks required of an M&E officer. The reasons for this were varied but included staff turnover due to limited incentives and compensation and a general lack of capacity and expertise to use FPET. On the whole, personnel issues appear to be tied to a lack of role clarity and disparities in M&E officer expertise and connections to influential stakeholders, two factors that may warrant consideration for future implementation of Track20. Despite the noted barriers, valuable assistance from in-country and U.S.-based Avenir representatives was cited as a facilitator of Track20 data maturity, particularly the work that was performed behind the scenes to ensure the professional growth of the M&E officers.

### *Data Systems*

**Data Collection.** Track20 performs secondary data analysis; thus, for Track20, data collection is secondary data *collation* that is dependent on the various forms of data feeding into Track20. The poor quality of service statistics afflicted *every* Track20 country. To counter this, respondents in Ethiopia, Kenya, Pakistan, Tanzania, and Zimbabwe implemented quality control mechanisms, such as quality scorecards or feedback to service centers providing data. The quality of service statistics is compromised by paper-based entry (Lao PDR), a reluctance of private-sector health facilities to share data (Indonesia and Nigeria), and accessing DHIS2/HMIS data in general (Tanzania and India). Thus, one potential mechanism to enhance the data maturity of Track20 would be increasing the capacity of reliable data entry at service delivery points in Track20 countries.

**Data Management.** Because Track20 performs secondary data analysis, facilitators and barriers of data management are linked to quality control mechanisms and a lack of reliable, high-quality data from service statistics, as described above in the "Data Collection" section. In addition, respondents in Indonesia referenced the difficulty of performing quality checks on the "massive amounts of service data" coming in from the country's more than 500 districts, as well as a lack of capacity to properly assess data quality, except in advance of the consensus meeting (Lao PDR).

**Data Analytics.** Only one respondent in Tanzania cited the successful pilot test of Track20's innovative translation of service statistics to the Estimated Modern Use tool, which

should enable countries to translate service statistics data into proxy estimates for mCPR. Most respondents honed in on the use of FPET. Respondents in India, Nigeria, Pakistan, and Uganda and reported receiving adequate training and understanding how to create predictive models, and, in general, they found FPET to be a very useful tool. Suboptimal use of FPET, reported in DRC and Tanzania, was impacted by limitations in understanding how to use FPET and, interestingly, skepticism of Bayesian models in Indonesia. Disparities in the optimization of FPET is part and parcel of the variation in skill sets and experience of the Track20 M&E officers, which Chapter Nine explores in detail.

**Technology.** The technology used within the Track20 initiative is tied to analysis and production of Track20 estimates. Thus, no facilitators related to technology were cited beyond the use of FPET (India, Nigeria, Pakistan, Tanzania, and Uganda); the Service Statistics to Estimated Modern Use tool (Tanzania); and statistical programs, such as R (India). Barriers were related to the need for more computers and hardware, which is expected, given the resource constraints in the sub-Saharan African sample. For instance, one respondent from Kenya stated, "Physical resources, the computers and even just the software and all that has always been a challenge within the ministry. That one is a challenge in all the projects we do. There's never enough resources to share around."

**Data Governance.** Data governance was not a prominently discussed aspect of data maturity, which may be attributed to clearly outlined and standardized practices of data analysis and outputs within the FPET tool.

**Data Institutionalization.** Interactions with Avenir personnel were critical to devising country-based strategies to improve the institutionalization of Track20 estimates. For instance, Avenir assisted Indonesia by noting that "branding" Track20 as an Avenir/Gates Foundation–funded project could hinder buy-in, trust, and consistent data use and advised that Track20-related work be marketed as a product of the Indonesian government rather than an international initiative. In addition, a respondent in Burkina Faso described how Avenir works with local NGOs to reach family planning goals. The center of data institutionalization was often cited as the national consensus meeting when progress on indicators was discussed among key national stakeholders. This process was also tied to the notion that Track20 served as an initial impetus to improve family planning in Track20 countries, which had the iterative effect of bolstering the institutionalization of Track20. This was particularly the case in India, Nigeria, Pakistan, Uganda, and Zimbabwe. One respondent in Nigeria stated,

> We as human beings, our journey is clearer if you know where you are coming from, and then you know the expectation, like when you start school you know, in 400 level, you are supposed to graduate in four years and you want to graduate maybe at first class. Track20 sort of does that for family planning. It shows you where you are, shows you where you ought to be. Just that presentation makes it all clear.

Just as active collaboration at the consensus meetings facilitated the institutionalization of Track20, poor communication of data outputs and limited participation in consensus meetings served as critical barriers. In Tanzania, there are few opportunities to interface with key decisionmakers outside of the consensus meeting. Despite the positive view of one respondent in Nigeria, another lamented, "Very few people know about Track20 in Nigeria, and not even people from PMA2020 know about Track20 yet." In addition, observations from the consensus meetings identified a lack of understanding of Track20 indicators, reluctance to share

information on finances and overhead costs, and key stakeholders who were either not present or not invited to the meetings. Hence, in all countries, institutionalization through the consensus meeting leaves room for improvement.

## Analyses of Data Maturity Ratings from the 15 Program Countries

Our data maturity framework served as a basis to collect ratings from all 15 study countries in order to gain a fuller appreciation of the state of data maturity and the process of improvement in these countries. As described in Chapter Four, we drew from several exemplars of best practices from a cross-section of industries to develop a composite data maturity framework that assesses the maturity of organizational readiness, data systems, and data use across countries while at the same time capturing the particularities of the data generation and management processes of PMA2020 and Track20. In total, 147 (65 percent) of the 225 individuals interviewed completed the ratings. (Others noted that they did not have the time to complete the ratings or did not respond to repeated follow-up requests.) Their quantitative feedback is the basis for the analyses described below.

To analyze the PMA2020 and Track20 ratings, we compiled average scores across domains and areas, countries, and country stakeholder group. We then used the quantitative data to assess domains and areas that respondents rated as relatively strong or weak, first by type of stakeholder (i.e., staff affiliated with PMA2020 or Track20, government officials, nongovernmental and other in-country representatives with family planning programming) and then by country. We also analyzed country-specific data by type of stakeholder. Examination of self-assessed ratings by PMA2020 and Track20 staff compared with ratings by nonprogram stakeholders helped to calibrate the self-assessments of the two programs.

The data maturity framework elements and associated questions are shown in Chapter Four (Figure 4.3). It is important to note before comparing scores across countries, programs, and even individuals that there is inherent subjectivity and variation in one country's or one person's ratings as compared with another's.

### Data Maturity Ratings by Theme and Respondent Group

On a scale from 1 to 10, government officials and representatives of bilateral, multilateral, and nongovernmental organizations (hereafter *government/NGO* respondents) rated the data maturity of both PMA2020 and Track20 programs consistently lower than these programs rated themselves (Table 12.1). This may be because they are less familiar with the programs and therefore with their actual levels of data maturity; on the other hand, their distance from these programs could also make them more objective and critical observers. Since they are the end-users of the data products, the perceptions of these decisionmakers certainly bear careful consideration. The ratings ranged quite widely across data maturity elements and respondent group, from a minimum of 5.2 (government/NGO rating of Track20 data use policy and accessibility) to a maximum of 9.4 (Track20 staff self-assessment of privacy protection), indicating variability in both data maturity elements and perceptions of different stakeholders.

**Table 12.1**
**Consolidated Data Maturity Ratings of PMA2020 and Track20, by Stakeholder Group**

| | | | PMA2020 | | Track20 | | |
|---|---|---|---|---|---|---|---|
| Domain | Subdomain | Area | PMA2020 Self-Assessment | Government/ NGO Assessment of PMA2020 | Track20 Self-Assessment | Government/ NGO Assessment of Track20 | Overall Average |
| Organization | Readiness | Staffing | 7.9 | | 6.4 | 5.3 | 6.5 |
| | | Staff buy-in | 8.0 | | 7.5 | 7.4 | 7.6 |
| | | Leadership buy-in | 8.5 | | 7.8 | 7.8 | 8.0 |
| | | Communication | 7.9 | | 7.7 | 6.2 | 7.3 |
| | | Infrastructure | 7.0 | | 6.1 | 7.5 | 6.9 |
| Data systems | Collection | Technology | 8.1 | | | | 8.1 |
| | | Response rates and completeness | 8.8 | | | | 8.8 |
| | | Data quality | 8.8 | 7.3 | 7.5 | 6.8 | 7.6 |
| | | Data sufficiency | 7.5 | 7.0 | 8.1 | 7.0 | 7.4 |
| | | Geographic granularity | 6.6 | 6.0 | 6.7 | 6.4 | 6.4 |
| | Manage-ment | Quality assurance processes/ practices | 8.4 | 6.6 | 6.7 | 6.6 | 7.1 |
| | | Quality improvement orientation | 8.7 | | 7.1 | 6.8 | 7.5 |
| | | Privacy protection | 9.1 | | 9.4 | 8.6 | 9.0 |
| | | Integration | 7.4 | | 8.1 | 7.8 | 7.8 |
| | | Documentation | 8.5 | | 7.8 | 6.6 | 7.6 |
| | Analytics | Technology | 8.6 | | 7.9 | 7.4 | 7.9 |
| | | Analytics capabilities | 8.3 | | 7.4 | 6.9 | 7.5 |
| | | Analytics quality | 9.0 | 7.8 | 8.1 | 7.8 | 8.2 |
| | | Analytics utility | 8.3 | 7.3 | 7.8 | 7.1 | 7.6 |
| | Governance | Data use policy and accessibility | 8.3 | 7.3 | 7.2 | 5.2 | 7.0 |
| | | Security policy/ protection | 9.0 | 7.0 | 8.2 | 6.9 | 7.8 |
| | | Standards development and adoption | 8.8 | | 8.3 | 7.5 | 8.2 |

**Table 12.1—continued**

| Domain | Subdomain | Area | PMA2020 | | Track20 | | Overall Average |
|---|---|---|---|---|---|---|---|
| | | | PMA2020 Self-Assessment | Government/NGO Assessment of PMA2020 | Track20 Self-Assessment | Government/NGO Assessment of Track20 | |
| | Institutional-ization | Policies | 7.7 | 7.0 | 7.9 | 7.0 | 7.4 |
| | | Data collection processes | 7.3 | 7.0 | 7.9 | 7.0 | 7.3 |
| | | Data management processes | 7.2 | 6.5 | 7.7 | 6.5 | 7.0 |
| | | Data analytics processes | 7.2 | 6.1 | 7.1 | 6.1 | 6.6 |
| | | Data-sharing processes | 6.5 | 6.3 | 7.1 | 6.3 | 6.5 |
| | | Data communications processes | 5.6 | 6.0 | 7.0 | 6.0 | 6.2 |
| Data use | Use | Valuation | 7.8 | 6.8 | 7.3 | 7.0 | 7.2 |
| | | Ownership/stewardship | 7.5 | 6.8 | 7.1 | 6.7 | 7.0 |
| | | Stage of change toward regular use of data for informed decisionmaking | 4.0 | 4.8 | 4.2 | 4.8 | 4.5 |
| Overall averages | | | 7.9 | 6.8 | 7.5 | 6.9 | 7.3 |

NOTES: All areas are scored on a scale from 1 (red; beginning) through 10 (green; advanced), except for the last item under data use, which is scored from 1 through 6. The overall averages include only the items scored on a 10-point scale.

## Data Maturity Ratings by Theme and Country

### *Overall Ratings by All Stakeholders*

Data maturity ratings varied significantly from country to country. The discussion below presents the ratings from all stakeholder groups from the 15 program countries included in the RAND evaluation (Table 12.2). The subsequent sections present country-specific ratings separately for PMA2020 staff, Track20 staff, and government/NGO representatives—who, as noted above, often had different perceptions about the level of specific data maturity elements.

Overall, the country with the highest ratings was Burkina Faso, followed by India, Ethiopia, and Ghana; the lowest-rated country was DRC.

### *Organizational Readiness*

The overall ratings for elements of organizational readiness were rated relatively low. Despite its overall high data maturity rating, Burkina Faso was rated low for organizational readiness, as were DRC and Lao PDR. Leadership buy-in was the highest-rated organizational readiness element; staffing and communications were the lowest-rated elements.

**Table 12.2**
**Ratings by All Stakeholders of Key Data Maturity Domains and Areas, by Country**

| Domain | Subdomain | Area | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Overall Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Organization | Readiness | Staffing | 5.0 | 8.3 | 5.0 | 7.0 | 7.0 | 7.8 | 8.2 | 8.5 | 4.7 | 9.3 | 5.4 | 6.1 | 5.0 | 8.5 | 8.0 | 6.5 |
| | | Staff buy-in | 8.0 | 8.1 | 5.5 | 7.9 | 8.4 | 8.6 | 7.7 | 8.5 | 6.8 | 9.0 | 7.5 | 8.1 | 7.5 | 7.5 | 8.0 | 7.6 |
| | | Leadership buy-in | 8.0 | 8.7 | 7.0 | 8.1 | 8.7 | 9.2 | 8.3 | 9.5 | 7.8 | 6.5 | 8.4 | 8.6 | 7.5 | 8.5 | 9.0 | 8.0 |
| | | Communication | 8.0 | 8.1 | 8.0 | 7.6 | 8.2 | 8.8 | 7.5 | 8.0 | 5.8 | 7.7 | 8.2 | 7.5 | 7.0 | 8.5 | 8.0 | 7.3 |
| | | Infrastructure | 1.0 | 6.3 | 5.5 | 5.6 | 6.2 | 9.1 | 7.7 | 5.0 | 6.5 | 8.5 | 6.9 | 8.7 | 5.0 | 8.0 | 6.0 | 6.9 |
| Data systems | Collection | Technology | | 7.5 | 8.0 | 8.4 | 8.0 | 8.5 | 8.0 | 8.0 | | 8.0 | 7.8 | | | | | 8.1 |
| | | Response rates, completeness | | 8.0 | 9.0 | 8.9 | 8.6 | 9.8 | 7.5 | 9.0 | | 9.3 | 8.3 | | | | | 8.8 |
| | | Data quality | 8.3 | 7.5 | 7.0 | 8.6 | 8.6 | 8.1 | 7.9 | 7.5 | 6.4 | 7.7 | 7.9 | 6.6 | 6.1 | 7.4 | 7.8 | 7.6 |
| | | Data sufficiency | 8.0 | 9.3 | 7.0 | 8.1 | 7.1 | 7.6 | 7.7 | 6.1 | 6.0 | 8.4 | 7.5 | 8.0 | 8.5 | 6.9 | 7.0 | 7.4 |
| | | Geographic granularity | 6.0 | 7.8 | 3.8 | 7.0 | 7.5 | 6.7 | 6.7 | 7.0 | 6.5 | 5.4 | 6.2 | 6.7 | 8.2 | 6.8 | 7.4 | 6.4 |
| | Management | Quality assurance processes/ practices | 8.0 | 8.1 | 4.7 | 7.3 | 7.5 | 7.8 | 7.7 | 7.8 | 6.8 | 5.9 | 7.8 | 6.5 | 6.7 | 7.2 | 7.4 | 7.1 |
| | | Quality improvement orientation | 7.0 | 9.3 | 7.3 | 8.3 | 8.8 | 8.2 | 7.5 | 8.0 | 4.5 | 7.8 | 8.6 | 7.8 | 7.8 | 8.5 | 7.5 | 7.5 |
| | | Privacy protection | 8.0 | 9.0 | 9.0 | 8.7 | 9.6 | 9.9 | 8.0 | 9.5 | 10.0 | 5.8 | 9.5 | 9.2 | 8.8 | 9.5 | 9.0 | 9.0 |
| | | Integration | 8.0 | 8.2 | 6.0 | 6.8 | 8.0 | 8.4 | 7.8 | 8.0 | 6.3 | 8.0 | 7.7 | 9.1 | 7.5 | 9.0 | 8.5 | 7.8 |
| | | Documentation | 8.0 | 7.8 | 6.8 | 8.7 | 8.3 | 8.9 | 8.7 | 8.5 | 7.0 | 7.7 | 7.9 | 7.3 | 7.5 | 7.5 | 10.0 | 7.6 |

**Table 12.2—continued**

| Domain | Subdomain | Area | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Overall Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Analytics | Technology | 8.0 | 8.5 | 8.3 | 7.3 | 8.8 | 9.3 | 8.3 | 9.0 | 8.0 | 9.0 | 7.8 | 7.6 | 6.5 | 7.0 | 8.5 | 7.9 |
| | | Analytics capabilities | 8.0 | 8.8 | 7.3 | 7.6 | 9.3 | 8.4 | 8.0 | 7.5 | 5.8 | 8.3 | 7.8 | 8.0 | 6.8 | 9.5 | 5.0 | 7.5 |
| | | Analytics quality | 8.5 | 9.5 | 7.0 | 8.2 | 8.1 | 8.2 | 8.0 | 7.5 | 8.0 | 7.1 | 7.9 | 8.0 | 8.3 | 8.6 | 8.5 | 8.2 |
| | | Analytics utility | 8.5 | 9.0 | 7.3 | 8.2 | 8.6 | 8.1 | 7.7 | 7.1 | 8.0 | 5.9 | 7.4 | 8.0 | 8.0 | 8.0 | 7.5 | 7.6 |
| | Governance | Data use policy and accessibility | 7.5 | 6.3 | 5.3 | 7.9 | 8.5 | 7.9 | 7.4 | 6.9 | 3.6 | 5.6 | 7.6 | 6.7 | 7.8 | 8.4 | 5.4 | 7.0 |
| | | Security policy/ protection | | 8.0 | 7.5 | 8.8 | 9.0 | 10.0 | 5.3 | 10.0 | 8.0 | 8.7 | 8.7 | 6.8 | 8.0 | 9.3 | | 7.8 |
| | | Standards development and adoption | 8.0 | 9.0 | 7.5 | 8.8 | | 8.5 | 6.3 | 8.5 | 8.5 | 7.0 | 9.3 | 7.6 | 9.0 | 9.0 | 9.0 | 8.2 |
| | Institutional-ization | Policies | 9.0 | 8.2 | 6.2 | 8.3 | 7.4 | 7.8 | 8.0 | 8.9 | 8.5 | 7.3 | 6.6 | 6.5 | 8.3 | 7.0 | 8.1 | 7.4 |
| | | Data collection processes | 9.5 | 8.7 | 4.8 | 7.9 | 7.2 | 8.1 | 7.6 | 7.5 | 7.8 | 8.3 | 6.6 | 6.4 | 7.6 | 7.2 | 8.0 | 7.3 |
| | | Data management processes | 9.5 | 8.5 | 4.7 | 7.5 | 6.7 | 8.1 | 7.7 | 6.8 | 7.2 | 8.8 | 6.1 | 6.6 | 6.8 | 7.5 | 7.3 | 7.0 |
| | | Data analytics processes | 8.5 | 7.7 | 4.7 | 7.7 | 6.2 | 7.9 | 7.3 | 6.3 | 6.6 | 7.3 | 6.1 | 6.5 | 5.3 | 7.7 | 6.6 | 6.6 |
| | | Data-sharing processes | 8.0 | 6.5 | 4.7 | 7.0 | 6.4 | 7.7 | 7.2 | 6.4 | 7.0 | 7.9 | 5.5 | 6.3 | 6.4 | 7.6 | 6.8 | 6.5 |
| | | Data communications processes | 7.5 | 6.3 | 3.8 | 7.5 | 6.2 | 7.8 | 7.5 | 6.2 | 6.9 | 6.8 | 5.8 | 6.4 | 5.8 | 6.4 | 6.7 | 6.2 |

**Table 12.2—continued**

| Domain | Subdomain | Area | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Overall Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data use | Use | Valuation | 8.5 | 8.5 | 6.3 | 7.5 | 8.1 | 7.5 | 7.6 | 7.5 | 7.0 | 7.0 | 6.9 | 6.4 | 7.5 | 7.1 | 7.5 | 7.2 |
| | | Ownership/ stewardship | 8.0 | 6.0 | 6.6 | 7.6 | 7.0 | 7.0 | 7.4 | 7.3 | 7.6 | 6.5 | 6.5 | 6.5 | 7.4 | 6.6 | 7.2 | 7.0 |
| | | Stage of change toward regular use of data for informed decisionmaking | 6.0 | 2.7 | 3.7 | 4.8 | 5.3 | 4.4 | 5.3 | 5.9 | 3.4 | 3.6 | 4.2 | 4.8 | 3.7 | 4.0 | 4.6 | 4.5 |
| Average scores per country | | | 8.0 | 7.7 | 5.9 | 7.7 | 7.7 | 7.8 | 7.6 | 7.4 | 6.8 | 7.1 | 7.5 | 7.2 | 6.9 | 7.2 | 7.2 | 7.3 |

NOTES: All areas are scored on a scale from 1 (red; beginning) through 10 (green; advanced), except for the last item under data use, which is scored from 1 through 6. The overall averages include only the items scored on a 10-point scale.

#### *Data Systems*

The highest-rated data systems elements fell under collection (response rates and completeness), management (privacy protection), and analytics (quality, standards development, and adoption). These elements were rated relatively highly across all countries. The lowest-rated data system elements, across most countries, included geographic granularity under data collection and data sharing and communications under institutionalization. Notable low ratings also included quality improvement orientation, data use policy, and analytics capabilities in Lao PDR; analytics capabilities in Uganda; and quality assurance practices in DRC.

#### *Data Use*

Data use elements scored in the mid-range of average ratings across all data maturity elements, with the highest ratings in Burkina Faso and Kenya and the lowest ratings in Côte d'Ivoire, followed by Lao PDR, Niger, DRC, and Tanzania.

### PMA2020 Personnel Ratings by Country

The following discussion presents the ratings from all stakeholder groups from the ten countries for which data were available (there are 11 PMA2020 countries in the study, but there were no completed matrixes from Burkina Faso). Overall, PMA2020 staff in India had the most favorable perception of the state of PMA2020 data maturity, with Niger and Nigeria not far behind; Uganda gave the most critical assessment of its data maturity (Table 12.3). Personnel affiliated with PMA2020 tended to rate more maturity areas as relatively strong and fewer as relatively weak.

#### *Organizational Readiness*

PMA2020 personnel considered leadership buy-in to be the strongest element of their organizational readiness, particularly in India and Kenya. As noted in the overall ratings described above, they rated data infrastructure as relatively weaker, especially in DRC, Ethiopia, and Ghana.

#### *Data Systems*

PMA2020 personnel rated several data systems domains and areas as stronger than others. These included data collection (response rates and completeness and data quality, especially in India and Niger); data management (quality improvement orientation, privacy protection, and documentation, especially in Kenya); analytics (technology in India and Kenya, analytics quality in Côte d'Ivoire and India); and governance (security policy/protection and standards development and adoption in India and Kenya). The commonality among the areas of data systems and data systems strengths appears to be that they are internal to the process or deal with rigor/quality issues that can be controlled by PMA2020 staff (e.g., response rates, data quality, privacy protection, analytics quality).

PMA2020 personnel tended to rate only a few domains as weak, including data collection (geographic granularity, rated low across all countries) and institutionalization (data-sharing processes and data communications processes, especially in DRC and Côte d'Ivoire). These areas are largely out of the control of PMA2020 country programs. Ratings may reflect a lack of resources, underdevelopment of the country, or communications issues within the country. Whether this indicates that PMA2020 programs are doing all they can in difficult operating environments or seeking to attribute difficulties to external factors beyond their control (or both) is open to interpretation. The level of geographic granularity is necessarily constricted by

**Table 12.3**
**PMA2020 Personnel Ratings of Key Data Maturity Domains and Areas, by Country**

| Domain | Subdomain | Area | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Niger | Nigeria | Uganda | Average Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Organization | Readiness | Staff buy-in | 7.5 | 6.5 | 7.9 | 8.4 | 8.5 | 7.3 | 8.0 | 8.0 | 8.0 | | 8.0 |
| | | Leadership buy-in | 7.5 | 8.5 | 8.1 | 8.7 | 10.0 | 8.5 | 10.0 | 7.0 | 8.7 | | 8.5 |
| | | Infrastructure | 8.5 | 6.0 | 5.6 | 6.2 | 9.8 | 7.3 | 7.0 | 8.0 | 6.8 | | 7.0 |
| Data systems | Collection | Response rates and completeness | 8.0 | 9.0 | 8.9 | 8.6 | 9.8 | 7.5 | 9.0 | 9.3 | 8.3 | | 8.8 |
| | | Data quality | | 8.5 | 8.9 | 9.0 | 9.1 | 8.0 | 8.0 | 9.0 | 8.7 | 8.0 | 8.8 |
| | | Geographic granularity | 6.5 | 2.5 | 7.5 | 7.4 | 7.4 | 7.7 | 7.0 | 4.7 | 6.5 | 6.0 | 6.6 |
| | Management | Quality improvement orientation | 10.0 | 7.5 | 8.3 | 8.8 | 9.4 | 8.0 | 10.0 | 8.7 | 8.8 | | 8.7 |
| | | Privacy protection | 9.0 | 9.0 | 8.7 | 9.6 | 10.0 | 8.0 | 10.0 | 9.7 | 9.0 | | 9.1 |
| | | Documentation | 6.5 | 6.5 | 8.7 | 8.3 | 9.5 | 8.3 | 9.0 | 8.3 | 9.3 | | 8.5 |
| | Analytics | Technology | 8.5 | 8.5 | 7.3 | 8.8 | 10.0 | 8.7 | 10.0 | 9.0 | 8.0 | | 8.6 |
| | | Analytics quality | 10.0 | 8.0 | 8.3 | 7.5 | 9.3 | 8.3 | 7.0 | 6.7 | 8.0 | | 9.0 |
| | Governance | Security policy/protection | 7.0 | 9.0 | 8.8 | 9.0 | 10.0 | 5.3 | 10.0 | 9.3 | 9.0 | | 9.0 |
| | | Standards development and adoption | | | 8.8 | | 9.7 | 4.7 | 9.0 | | | | 8.8 |
| | Institutionalization | Data-sharing processes | 5.5 | 4.5 | 6.4 | 6.2 | 9.2 | 7.0 | 6.0 | 8.7 | | 7.0 | 6.5 |
| | | Data communications processes | 6.0 | 3.0 | | 5.8 | 10.0 | 8.0 | 5.0 | | | 5.0 | 5.6 |
| Data use | Use | Ownership/stewardship | 5.0 | 7.0 | 7.4 | 7.2 | 8.0 | 8.0 | 7.0 | 8.7 | | 6.0 | 7.5 |
| | | Stage of change toward regular use of data for informed decisionmaking | 2.0 | 4.0 | 3.8 | 5.0 | 3.5 | 5.7 | 6.0 | 3.3 | | 2.0 | 4.0 |
| Average scores per country | | | 7.6 | 6.9 | 8.0 | 7.8 | 9.3 | 7.7 | 8.0 | 8.4 | 8.2 | 6.3 | 8.0 |

NOTES: All areas are scored on a scale from 1 (red; beginning) through 10 (green; advanced), except for the last item under data use, which is scored from 1 through 6. The overall averages include only the items scored on a 10-point scale. Burkina Faso is not included because none of its interviewees submitted a data maturity rating sheet.

funding realities, while the data infrastructure in a given country is certainly beyond the scope of PMA2020 alone to address. Both factors somewhat limit the effectiveness of PMA2020, but for different reasons that are not within their control.

#### *Data Use*

PMA2020 respondents in all countries rated their stage of data use as relatively low, especially in Côte d'Ivoire and Uganda. One particularly interesting finding to emerge was the divergence between two groups of countries in regard to the stage of change area within the data use domain. Respondents in one group of countries (Ghana, Indonesia, and Kenya) tended to rate their country as further along the progression to regular use of data compared with those in a smaller group of countries (Côte d'Ivoire and Uganda) who thought that their country was at early stages in this progression, suggesting that they would like to use data but do not yet fully understand them or the ways in which they might be useful. The countries in the first group, which reportedly use data frequently or regularly, have traditionally higher per capita incomes or governance scores than the two countries in the second group.

One interpretation of this dichotomy might be a potential correlation of government economic level and governance with the basic understanding of data and the perception of willingness or ability to use PMA2020 data. Those countries with greater capacity or governance may be better able to use the data and may also be further along in the stages of government usage. Meanwhile, those countries that are not as developed or do not have the same governance capacity may feel that they are still in the nascent stages of incorporating PMA2020 data into their decisionmaking.

### Track20 Personnel Ratings by Country

Track20 program respondents were less positive in their overall assessments of Track20 data maturity compared with their PMA2020 counterparts, albeit still generally positive about the program. Indeed, their ratings fell squarely in the middle ranges, apparently reflecting a general satisfaction with the program's level of data maturity across the different countries. There was less variance among the scores in general, with Track20 personnel in Côte d'Ivoire and Uganda rating their country's data maturity highest and those in DRC and Nigeria rating it lowest (Table 12.4). As with PMA2020, there were fewer areas rated as relatively weak by Track20 personnel.

A comparison of areas rated as stronger versus weaker reveals patterns similar to those described above from PMA2020 staff (Table 12.3). The main unifying theme of many of the highly rated areas appears again to be those factors that the country programs can directly control or influence, including data sufficiency, privacy protection, analytics quality, and security policy. Yet again, infrastructure and geographic granularity are cited as factors hampering the efficacy of family planning programs, this time in regard to Track20 efforts.

#### *Organizational Readiness*

Track20 respondents rated staffing highest in Niger, Côte d'Ivoire, and Uganda and lowest in DRC. They rated data infrastructure strongest in Pakistan and Niger and lowest in Burkina Faso and Kenya.

#### *Data Systems*

Track20 personnel rated several areas as relatively strong: data collection (Track20 data sufficiency), especially in Burkina Faso, Niger, Côte d'Ivoire, and Indonesia; data management

**Table 12.4**
**Track20 Personnel Ratings of Key Data Maturity Domains and Areas, by Country**

| Domain | Subdomain | Area | Burkina Faso | Côte d'Ivoire | DRC | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Average Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Organization | Readiness | Staffing | 5.0 | 8.7 | 4.5 | 6.6 | 8.0 | 8.0 | 5.3 | 9.0 | 4.6 | 7.0 | 5.0 | 8.5 | 8.0 | 6.4 |
| | | Infrastructure | 1.0 | 4.0 | 5.0 | 8.4 | 8.0 | 3.0 | 6.0 | 9.0 | 5.0 | 10.0 | 5.0 | 8.0 | 6.0 | 6.1 |
| Data systems | Collection | Data sufficiency (Track20) | 10.0 | 9.3 | 8.0 | 8.6 | 9.0 | 6.0 | 6.0 | 10.0 | 7.8 | 8.0 | 8.5 | 8.5 | 7.0 | 8.1 |
| | | Geographic granularity | 5.0 | 9.0 | 4.5 | 7.8 | 5.0 | 7.0 | 7.3 | 5.0 | 5.4 | 7.0 | 8.3 | 7.5 | 7.0 | 6.7 |
| | Management | Quality assurance processes/ practices | 8.0 | 6.3 | 4.0 | 7.8 | 8.0 | 8.0 | 6.3 | 5.0 | 6.4 | 7.0 | 6.3 | 7.5 | | 6.7 |
| | | Privacy protection | 8.0 | 9.0 | | 9.8 | 8.0 | 9.0 | | 2.0 | 9.5 | 10.0 | 8.8 | 9.5 | 9.0 | 9.4 |
| | | Integration | 8.0 | 9.3 | 7.5 | 7.8 | 8.0 | 9.0 | 7.5 | 8.0 | 7.4 | 10.0 | 7.5 | 9.0 | 8.5 | 8.1 |
| | Analytics | Analytics quality | 8.0 | 8.5 | 7.0 | 8.2 | 8.0 | 8.0 | 8.0 | 9.0 | 7.4 | 8.0 | 8.3 | 10.0 | 8.5 | 8.1 |
| | Governance | Security policy/ protection | | 9.0 | 6.0 | | | | 9.0 | 8.0 | 8.0 | 7.0 | 8.0 | 9.3 | | 8.2 |
| | | Standards development and adoption | 8.0 | 9.0 | 7.5 | 7.3 | 8.0 | 8.0 | 9.0 | 7.0 | 8.7 | 8.0 | 9.0 | 9.0 | 9.0 | 8.3 |
| Data use | Use | Ownership/ stewardship | 8.0 | 9.0 | 6.5 | 6.8 | 7.0 | 7.0 | 10.0 | 5.0 | 6.2 | 7.0 | 7.8 | 7.0 | 7.5 | 7.1 |
| | | Stage of change toward regular use of data for informed decisionmaking | 6.0 | 5.0 | 3.5 | 4.6 | 5.0 | 6.0 | | 3.0 | 3.6 | 5.0 | 2.0 | 5.5 | 4.0 | 4.2 |
| Average score per country | | | 7.6 | 8.5 | 6.1 | 7.7 | 7.6 | 7.4 | 7.3 | 7.4 | 6.9 | 7.8 | 7.1 | 8.1 | 7.6 | 7.4 |

NOTES: All areas are scored on a scale from 1 (red; beginning) through 10 (green; advanced), except for the last item under data use, which is scored from 1 through 6. The overall averages include only the items scored on a 10-point scale.

(data protection, integration), especially in Côte d'Ivoire, Pakistan, and Uganda; and data analytics and governance (security policy/protection, standards development, and adoption), especially in Côte d'Ivoire and Uganda. Across most countries, except Côte d'Ivoire and Tanzania, respondents rated geographic granularity low. They rated data management (quality assurance processes/practices) as relatively low across most countries and data privacy very low in Niger.

#### *Data Use*

The ratings of Track20 personnel reflected a wide range of perceived achievement in data use, from well established (in Burkina Faso, Kenya, and Uganda) to very low (in Tanzania and Niger).

### Comparison of Ratings for PMA2020 and Track20 by Their Respective Staff

Comparison of responses from respondents affiliated with PMA2020 and Track20 reveals some overlap in terms of the areas that personnel of both programs perceive as strong. Those domains fall largely under data systems: management (privacy protection), analytics (analytics quality), and governance (security policy/protection, standards development and adoption). However, geographic granularity was rated relatively lower by both groups. Again, organizational readiness (data infrastructure) was rated as generally weak by both PMA2020 and Track20 staff in several countries, though there was not unanimous agreement. For example, PMA2020 and Track20 respondents in India and Niger felt that infrastructure was not an issue.

There were two areas in which PMA2020 and Track20 ratings diverged considerably: organizational readiness (staffing) and data systems (governance—data use policy and accessibility). PMA2020 staff gave much higher ratings in both of these areas than did Track20 staff. However, it is worth noting that only PMA2020 respondents answered the question about PMA2020 staffing, while both Track20 and government/NGO respondents rated the Track20 staffing. However, government/NGO respondents did validate the data use policy and acceptability area, which PMA2020 rated much higher than Track20 (see the next section).

In general, the Track20 self-assessment scores tended to be lower than those of PMA2020 staff. There was, however, greater across-country agreement from Track20 respondents and slightly less variation (measured by standard deviation) in the self-assessment scores than for PMA2020 staff. Within countries, there was general agreement among stakeholders across all organizational affiliations (e.g., PMA2020 or Track20 program staff; government officials; or bilateral, multilateral or nongovernmental organizations) with regard to their rating of questions about the effect and effectiveness of PMA2020 or Track20 efforts more broadly (e.g., analytics capabilities, quality, or utility).

### Government and Nonprogram Partner Ratings

We used the ratings of government, bilateral, multilateral, and nongovernmental organization respondents (hereafter *government/NGO* respondents) to validate the ratings of PMA2020 and Track20 staff. These respondents rated only data systems and data use and were not asked about PMA2020 and Track20 organizational readiness (Table 12.5).

Government/NGO respondents in Burkina Faso rated the data maturity of their programs more highly than in any other country; those in DRC and India rated the maturity of their programs the lowest. But the government/NGO ratings were consistently lower than the self-assessments of both PMA2020 or Track20 respondents across the board, although, in gen-

**Table 12.5**
**Government/NGO Ratings of Key Data Maturity Domains and Areas, by Country**

| Domain | Subdomain | Area | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Average Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data systems | Collection | Data quality | 8.0 | | 5.0 | 8.3 | 8.2 | 6.8 | 8.0 | 7.8 | | 4.0 | 7.8 | 6.1 | | 8.0 | | 7.3 |
| | | | 9.0 | | 5.0 | 8.5 | | 7.0 | 7.5 | 7.2 | 6.2 | 4.0 | 7.4 | 6.0 | 6.7 | 7.3 | 7.1 | 6.8 |
| | Analytics | Analytics quality | 9.0 | | 7.0 | 8.0 | 8.7 | 7.3 | 8.0 | 7.8 | | 6.0 | 8.5 | | | 8.5 | | 7.8 |
| | | | 9.0 | 10.0 | 5.0 | 8.0 | | 7.0 | 7.5 | 8.0 | 7.8 | 9.0 | 8.5 | 7.3 | 8.8 | 7.0 | 7.6 | 7.8 |
| | | Analytics utility | 10.0 | | 7.5 | 8.0 | 8.6 | 6.4 | 7.6 | 7.3 | | 3.5 | 8.1 | | | 6.0 | | 7.3 |
| | | | 10.0 | 9.0 | 5.0 | 7.5 | | 6.0 | 7.3 | 7.8 | 7.8 | 6.0 | 8.0 | 7.0 | 7.2 | 5.7 | 6.0 | 7.1 |
| | Governance | Data use policy and accessibility | 8.0 | | 7.5 | 7.0 | 8.5 | 6.0 | 8.0 | 6.0 | | 2.0 | 8.2 | | | 9.5 | | 7.3 |
| | | | 8.0 | 5.0 | 4.5 | 8.0 | | 6.5 | 5.6 | 3.7 | 6.2 | 1.0 | 4.3 | 6.4 | 7.0 | 9.0 | 3.7 | 5.2 |
| | Institutionalization | Policies | 9.0 | 8.0 | 4.5 | 7.7 | 8.4 | 7.1 | 7.8 | 7.7 | 7.5 | 7.0 | 5.4 | 6.0 | 8.0 | 6.4 | 8.2 | 7.0 |
| | | Data management processes | 9.0 | 8.0 | 4.5 | 7.5 | 7.2 | 6.7 | 8.4 | 6.3 | 6.8 | 8.0 | 5.3 | 6.1 | 6.0 | 6.9 | 7.0 | 6.5 |
| | | Data analytics processes | 9.0 | 6.0 | 4.5 | 7.5 | 6.4 | 6.0 | 7.3 | 6.9 | 5.6 | 7.0 | 5.4 | 6.0 | 5.2 | 6.8 | 6.1 | 6.1 |
| | | Data-sharing processes | 8.0 | 5.0 | 4.5 | 7.5 | 6.6 | 6.3 | 7.5 | 6.3 | 6.5 | 8.0 | 5.1 | 5.7 | 5.3 | 7.7 | 6.6 | 6.3 |
| | | Data communications processes | 8.0 | 4.0 | 4.0 | 7.5 | 6.6 | 6.3 | 7.5 | 5.6 | 5.3 | 7.5 | 5.3 | 5.9 | 4.5 | 7.1 | 6.4 | 6.0 |
| Data use | Use | Ownership/ stewardship | 8.0 | | 6.0 | 8.0 | 6.8 | 6.0 | 7.5 | 7.4 | | 5.0 | 6.3 | | | 7.0 | | 6.8 |
| | | | 8.0 | 4.0 | 6.5 | 7.5 | | 6.4 | 6.8 | 8.4 | 5.2 | 6.5 | 7.0 | 5.9 | 7.1 | 6.7 | 6.8 | 6.7 |
| | | Stage of change toward regular use of data for informed decisionmaking | 6.0 | 1.0 | 4.0 | 6.0 | 5.6 | 4.4 | 5.5 | 5.5 | | 4.0 | 4.6 | | | 4.3 | | 4.8 |
| | | | 6.0 | 1.0 | 3.0 | 5.5 | | 5.6 | 4.8 | 5.8 | 3.4 | 5.0 | 5.0 | 4.6 | 5.3 | 4.7 | 5.2 | 4.8 |
| Average score per country | | | 8.4 | 7.4 | 5.1 | 7.5 | 7.7 | 6.3 | 7.5 | 7.0 | 6.5 | 5.8 | 7.4 | 6.5 | 6.9 | 7.1 | 6.9 | 6.9 |

NOTES: Where there are two rows of scores for an area, the first row indicates average ratings for PMA2020 data maturity, and the second row indicates average ratings for Track20 data maturity. All areas are scored on a scale from 1 (red; beginning) through 10 (green; advanced), except for the last item under data use, which is scored from 1 through 6. The overall averages include only the items scored on a 10-point scale.

eral, both groups identified roughly the same group of stronger and weaker data maturity elements. The fact that personnel outside of these programs did not view their level of data maturity as favorably as internal staff is perhaps not surprising, given that they have less knowledge of and investment in them. Another possible explanation is that they are potentially less biased toward favorable ratings than are the program staff about their own program.

Government/NGO ratings also demonstrated a much larger variance within countries. The fact that there was much less consistency to these scores may also not be surprising, given the breadth of the organizations involved, from family planning officials within the respective MOHs, to the heads of NGOs, to government ministers only tangentially involved in these programs.

#### *Data Systems*

Government and NGO personnel seemed to be satisfied overall with the data products produced by PMA2020 and Track20, specifically rating the following domains and areas highly (though, interestingly, not as highly as their PMA2020 and Track20 counterparts): collection (data quality—PMA2020 only), analytics (analytics quality, analytics utility), and governance (data use and accessibility—PMA2020 only). Broadly, they seemed to believe in the rigor of the process used by these programs, evidenced by the high scores for data quality, analytics quality, analytics utility, and data use and accessibility.

Those domains and areas perceived as weak were collection (data quality—Track20 only), governance (data use and accessibility—Track20 only), and institutionalization (data management processes, data analytics processes, data sharing processes, and data communications processes). Institutionalization of processes by PMA2020 and Track20 is a clear area for improvement.

These respondents rated PMA2020 and Track20 differently for several important domains. In most instances, they rated PMA2020 higher than Track20: data quality (in DRC, Ethiopia, Kenya, Nigeria, and Uganda), analytics quality (in DRC, Indonesia, and Uganda; Track20 rated higher than PMA2020 in Niger), analytics utility (in DRC and Ethiopia; Track20 rated higher in Niger), and governance—data use policy and accessibility (in DRC, Indonesia, Kenya, and Nigeria; Track20 rated higher in Ethiopia).

#### *Data Use*

Government/NGO respondents rated data use higher than did PMA2020 and Track20 personnel, with the highest ratings for Burkina Faso, Ethiopia, and Kenya and the lowest rating for Côte d'Ivoire.

### Country-Specific Conclusions and Within-Country Variations

Several patterns emerge from comparisons of ratings within and across countries. With the caveat of subjectivity of ratings in mind, as well as the major caveat that the two programs launched at different times in different countries, certain groupings of countries emerge with data maturity consistently rated more highly than other countries. Five countries in particular—Burkina Faso, Côte d'Ivoire, Ethiopia, Ghana, and India—were more consistently rated highly across the elements of data maturity compared with other countries included in this evaluation. Whether the perceived high level of data maturity is attributable to outstanding programs, to the country's own data system maturity, or to both is open to interpretation.

There were also two countries for which respondents consistently scored the level of data maturity below other countries: DRC and Lao PDR. While Lao PDR has only Track20 programming (i.e., no PMA2020), respondents from both countries rated the various elements of data maturity lower than respondents in other countries. In DRC, government/NGO respondents rated the data maturity of these programs lower on average than any set of stakeholders assessing either program. It would seem that PMA2020 respondents, Track20 respondents, government officials, and other stakeholders in these two countries are less satisfied than those in other countries with their progress toward data maturity. However, this also could be interpreted to mean that the level of data maturity in DRC and Lao PDR are merely fine or middle-of-the-road, if unspectacular, or that the raters in these countries are more stringent with their ratings, but we have offered a relativistic analysis here to present an illustrative comparison.

Of particular interest was the large variability of ratings across stakeholder groups within countries. In several countries, different groups had somewhat different perceptions about the degree of data maturity in the country. For example, in several countries, PMA2020 and Track20 respondents rated data maturity consistently higher than did government officials and their bilateral, multilateral, and nongovernmental partners. In four countries (DRC, India, Kenya, and Niger), the average ratings of PMA2020 respondents were at least one point above the average scores of their government/NGO counterparts. In seven countries (Côte d'Ivoire, DRC, India, Lao PDR, Niger, Pakistan, and Uganda), Track20 respondents rated their programs at least one point above their government/NGO counterparts.

This misalignment suggests that people within the respective programs perceive their data maturity and methods to be more highly effective and rigorous than do end users or those outside the programs. This could represent an issue of data quality or could indicate a possible breakdown of communication or poor dissemination of results. In only a few cases did PMA2020 or Track20 staff rate the data maturity of their programs below that of their government/NGO counterparts. There were four instances of large discrepancies between PMA2020 and Track20 ratings; in three of these cases (India, Niger, and Nigeria), respondents rated PMA2020 significantly higher, while in one case (Uganda), Track20 was rated significantly higher. It would be interesting to know whether low levels of integration or communication were behind these discrepancies or whether they are in fact due to low levels of data maturity.

### Synthesis of the Data Maturity Ratings

Despite the different perceptions of data maturity within and across countries and the few areas that were rated as relatively low, the clear majority of respondents were positive in their assessment of the state of data maturity in the surveyed countries. As the color-coded tables indicate, there is widespread satisfaction in the support for and progress of these programs across the board.

Regarding PMA2020 specifically, respondents felt that organizational readiness, most aspects of data systems (collection, analysis, management, and governance), and data use are proceeding relatively well, with some notable exceptions for some countries. One aspect of data systems—institutionalization—stands out as an area of concern to be addressed by program staff. Turning to Track20, some aspects of data systems (data management and analytics) and data use emerged as strong domains. Other areas within data systems (governance and institutionalization) and organizational readiness did not score well. These results indicate that both PMA2020 and Track20 may want to focus on improving the ways in which they share and

**Table 12.6**
**Overall Data Maturity Ratings, by Respondent Group and Country**

| Respondent | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PMA2020 | | 7.6 | 6.9 | 8.0 | 7.8 | 9.3 | 7.7 | 8.0 | | 8.4 | 8.2 | | | 6.3 | |
| Track20 | 7.5 | 8.5 | 6.1 | | | 7.7 | 7.6 | 7.4 | 7.3 | 7.4 | 6.9 | 7.8 | 7.1 | 8.1 | 7.6 |
| Government/NGO | 8.4 | 7.4 | 5.1 | 7.5 | 7.7 | 6.3 | 7.5 | 7 | 6.5 | 5.8 | 7.4 | 6.5 | 6.9 | 7.1 | 6.9 |
| Country average | 8.0 | 7.8 | 6.0 | 7.7 | 7.8 | 7.7 | 7.6 | 7.5 | 6.9 | 7.2 | 7.5 | 7.2 | 7.0 | 7.2 | 7.3 |

NOTE: All areas are scored on a scale from 1 (red; beginning) through 10 (green; advanced).

communicate data, while Track20 in particular should focus on staffing, data use, and data accessibility issues.

Notwithstanding the diversity within and across countries of the stakeholders who rated data maturity elements and issues of overall averages masking important element- and country-specific ratings, we can draw a few conclusions from these ratings. Nearly all countries fell into the high end of developing data maturity or the low end of advanced maturity (average scores in the 7 to 8 range on the 10-point scale), with DRC at a lower level of developing data maturity (average score 6.0). PMA2020 and Track20 program staff tended to rate (self-assess) data maturity in their country slightly higher than did nonprogram staff in nearly all countries (Table 12.6). The explanations for this slight discrepancy are unclear. Perhaps program staff have more positive views of the various elements of data maturity in their countries due to their direct experience with many of them through their work. Alternatively, the difference could reflect a lack of sensitivity of program staff to their data end users and/or more untarnished views of reality on the part of nonprogram staff. Also, the overall average for Burkina Faso stands out, largely because of high ratings from government and nongovernment respondents, who otherwise tended to rate data maturity lower than did PMA2020 or Track20 staff.

The responses from these 145 country stakeholders suggest that they view PMA2020 and Track20 as having considerable quality and rigor even if there remains room for improvement. We hope that PMA2020 and Track20 programs, as well as individual country programs, will find this quantitative feedback and its implications useful in their efforts to improve data maturity.

## Summary

PMA2020 and Track20, while well established and highly informed by best practices, are still in development, and they are not yet well known by all key stakeholders in some program countries. It is important to note that the data maturity of these programs encompasses a broad spectrum, and the barriers and facilitators, as well as the successes and opportunities we have described, exist on a continuum. The factors impacting data maturity in each country and for each program are highly variable. Some factors can be attributed to the operationalization of

either or both programs, while other factors, such as longstanding government bureaucracy, are outside the purview of the Gates Foundation and are not discussed here in detail.

Furthermore, the domains of data maturity are not mutually exclusive; for instance, strong teamwork among PMA2020-affiliated staff is both part of organizational readiness and a key factor for successful data collection. The themes described in this chapter depict those that were most salient to key stakeholders and/or that presented clear opportunities for overcoming limitations to data maturity.

Stakeholder perceptions reveal enthusiasm for the full optimization of these programs. They also highlight some of the workarounds that stakeholders pursue to overcome shortcomings in the complex systems required for advanced data maturity. Experiences and perspectives with the data maturity of the programs were not uniform across different types of stakeholders, thus highlighting the importance of conducting interviews across multiple stakeholder groups. Stakeholders conveyed to us that even with optimal data generation, the prevailing view is that data management, data analysis, and organizational readiness in terms of knowing how to use the data for decisionmaking all need further attention.

We developed and applied the data maturity framework as a proof of concept. The ratings were consistent with what we learned in our interviews. Therefore, our framework may be a promising tool for further development and testing as a complementary and more structured way to assess data maturity. Further development could include, for example, the addition of user-defined weights for specific data maturity elements. We believe that this framework is perhaps more appropriate for monitoring progress within countries than for comparisons across countries.

CHAPTER THIRTEEN

# Sustainability

## Overview

Sustainability is a nearly universal priority in the development community, yet it is a concept that is often difficult to precisely define and actually measure. The RAND team drew from a published definition of sustainability related to health information systems and data usage, which is relevant to consideration of the sustainability of PMA2020 and Track20:

> Sustainability [of a data system] refers to the tendency of the system to endure over time and space and is directly concerned with the system becoming institutionalized in the workings of the health department. Institutionalization can thus be described as the process by which [health information] systems can be sustained over time (Kimaro and Nhampossa, 2007).

For a system to be truly sustainable, the international development literature consistently emphasizes that it must be *locally driven, owned, and operated*—in other words, *institutionalized*. With recent evidence highlighting family planning investments as a "best buy" for development and consensus around FP2020, the Gates Foundation has supported PMA2020 and Track20 as two data programs to help governments track their continued progress toward their FP2020 targets. (Starbird, Norton, and Marcus, 2016). *Strong internal (government) and external (donor) support* are critical along the way, as is *collaboration* in program design and implementation across key actors, including all levels of government within a country (from national to local), donors, and civil society (Jamison et al., 2006). Consistent, predictable *funding support* is needed, alongside programmatic approaches that support the *technical and operational capability* of the system in terms of human and technological capacity and operational processes. Sustainability in the strengthening of country-level health systems and data management thus depends on the degree to which local systems and communities internalize and incorporate related principles into their everyday operations. The RAND sustainability framework described in Chapter Four includes domains related to financial sustainability, technical sustainability, operational sustainability, and data culture. It provides a checklist of sustainability-enabling factors considered key to the sustainability of PMA2020 and Track20 and a means to measure progress over time.

This chapter describes the perspectives on sustainability offered by more than 200 in-country stakeholders and more than 30 U.S.-based stakeholders, as reflected through semistructured interviews and the structured sustainability ratings completed by approximately 60 percent of in-country interviewees. It begins with stakeholders' perspectives on key themes that arose from the interviews, followed by more quantitative analysis drawn from the structured sustainability ratings. As with the discussion of data maturity (Chapter Twelve), the

qualitative and quantitative analyses discussed here complement one another and provide a rich understanding of the perspectives gleaned from these many stakeholders.

## Qualitative Analyses from In-Country Stakeholder Interviews

Interviews with all stakeholder groups (Gates Foundation, Gates Institute, Avenir, other U.S.-based experts, and program countries) reflected interest in the sustainability of PMA2020 and Track20, but the specific financial, technical, and operational dimensions appear to lack clarity and alignment. The following sections describe stakeholders' perspectives on the potential sustainability of each program. Because of the volume and richness of perspectives from the in-country stakeholders, their views are organized by theme, aligned with the sustainability framework.

It is important to note that some facilitators and barriers to sustainable data systems are more linked to the country than to either PMA2020 or Track20. Political will to support family planning broadly and data systems specifically is vital but may not be linked specifically to the two programs, though it may be subject to advocacy from them or on their behalf. Political instability and even routine elections (especially if there is government turnover) can facilitate or impede family planning efforts. Competing priorities may also limit the amount of support and the sustainability of family planning programs and associated data systems. Comments from several countries now classified as lower-middle-income countries indicated that they have seen a drop in development assistance since transitioning from low-income status. Finally, cultural resistance to family planning in some countries (often religious—whether Catholic, evangelical Christian, or Muslim) may also stand in the way. To address this, several respondents suggested framing family planning within a broader context of women's and children's health.

### Sustainability of PMA2020

From the Gates Foundation's perspective, **sustainability would mean that the surveys are sufficiently affordable to attract co-financing, since the foundation does not envisage fully financing them indefinitely**, and that countries would recognize the value of the PMA2020 data platform, integrate it into their M&E architecture, and **use the data to inform country programming**. **Countries would become progressively less dependent on both the technical support provided by the Gates Institute and financial support from the Gates Foundation**. Both the Gates Foundation and other U.S.-based stakeholders see other donors (more than the countries themselves) as the main source of the future financial sustainability of PMA2020—whether these donors are FP2020 partners or other bilateral or multilateral organizations supporting family planning or other development programs. **Reducing survey costs** and providing itemized budget projections would be essential for potential co-financing partners, but clear business plans along those lines remain pending from the Baltimore team. One interviewee wondered aloud whether incentives for the Gates Institute were properly aligned to help achieve increasing country independence and sustainability of PMA2020.

The views of other U.S.-based stakeholders came mainly from members of the PMA2020 External Consultative Group, all of whom are very familiar with the program and have participated in discussions about its achievements and possible future directions. Several expressed doubt about the program's financial sustainability and concerns about PMA2020 competing

with the DHS or MICS for donor funding. One considered PMA2020 as still a proof of concept rather than a sustainable platform. Several felt that PMA2020 should fill an identifiable void in countries, responding to their data needs. Countries are more likely to share in the cost of PMA2020 surveys if they value data and understand how PMA2020 surveys complement other data sources and address specific country needs. At least one of these stakeholders noted that the increasing use of service statistics and modeled estimates raises questions about the necessity of frequent family planning surveys like PMA2020.

The in-country stakeholders—including both PMA2020 staff and nonstaff—provided interesting insights about PMA2020 sustainability from the field perspective. The discussion that follows is organized by theme.

#### *Financial Sustainability*

Health infrastructure lacks resources around the globe, and, unfortunately, family planning is rarely positioned as a top priority. Thus, not surprisingly, **financing was a commonly cited barrier to PMA2020 sustainability**. Barriers related to financing included a lack of reliability in terms of the delivery of funds (as was particularly the case in Côte d'Ivoire), debates over funding (Indonesia), a lack of government support for funding PMA2020 (India), and, across countries, a need to show the cost-benefit analysis of funding.

Nearly all stakeholders recognized the **significant costs associated with the current PMA2020 surveys,** and many (including government officials) felt that, **in principle, governments should share in the costs** because they value and ultimately benefit from the data. However, very few expressed confidence in the willingness or ability of the respective governments to assume full costs in the near term (or even over the longer term). While one UNFPA respondent noted that the electronic collection and processing of survey data reduce labor costs, an NGO respondent from the same country suggested cost savings through leaner surveys (fewer questions, while retaining critical content).

The first ten PMA2020 countries have been supported financially by the program. Beginning with Côte d'Ivoire, the program is requiring some degree of co-financing. A respondent from that country commented on the importance of co-financing from the outset. The Gates Institute noted that this is the direction it hopes to follow into the future, but the surveys first needed to be tested, validated, and legitimized.

Several respondents suggested interesting approaches or opportunities for co-financing. One respondent in Ethiopia noted that the MOH has provided consultants at the national and regional coordinator levels, which could be considered in-kind co-financing of sorts. Another Ethiopian interviewee suggested co-financing from the private sector because of the appealing features of the business model (e.g., good management, profit incentive). A respondent in India described a unique situation in that country—a law requiring corporations to invest a certain percentage of their profits (a large amount overall) in society. Perhaps most promising, and more broadly applicable, was the suggestion about approaching the Global Financing Facility (a funding mechanism in many countries, supported by the World Bank but owned by the government) for co-financing. For example, a respondent in Nigeria suggested that this entity night be approached to consider the PMA2020 platform to meet their annual data needs. However, this would likely mean expansion beyond the current focus of most PMA2020 surveys on family planning. Indeed, several interviewees (from Ethiopia, Ghana, India, and Nigeria) commented that judicious inclusion of **additional survey modules (i.e., beyond family**

**planning, and perhaps on a rotating basis) could be a good approach and opportunity to broaden the base of potential co-financers**.

A PMA2020 staff member summed up one key to future funding: "Funding is important, but also understanding the usefulness and importance of data . . . . With this understanding, funding will follow."

### *Technical Sustainability*

Technical sustainability is a function of *capacity* already on hand in countries and *capacity-building* to help achieve sustainable, mature data systems. Respondents in five countries (Ethiopia, Ghana, India, Niger, and Nigeria) specifically noted that they felt that their country already has good technical capacity—most commented specifically on the capacity within the universities with which PMA2020 partners directly. Respondents from nearly all countries commented on capacity-building. Some specifically noted that **technical capacity (trained people) is essential to sustainability**. While some respondents (e.g., from Ethiopia and Uganda) felt that PMA2020 had built their capacity and increased their sense of empowerment, others (e.g., from Nigeria) felt that **capacity-building should be part of the program's plan, and much more is needed**, especially at subnational levels (e.g., at the county level in Kenya). PMA2020 staff have also provided training to others. For example, PMA2020 staff in Ghana have trained staff from the Ghana Statistical Services in the use of mobile technology for data collection.

Of particular interest, some respondents offered detailed ideas for the kind of data capacity needed. For example, a government official in Burkina Faso noted that local users need both data access and knowledge—**local actors need training on data usage and interpretation**. A respondent in Ghana noted, "If you want someone to use your data, **you must tell the person in a language that the person understands**, case closed." Another respondent from Uganda echoed the importance of people actually understanding what the data mean. These comments are a foundation for one of the recommendations stemming from this evaluation (Chapter Fourteen).

### *Operational Sustainability*

**Leadership buy-in** is a first prerequisite for operational sustainability. Such buy-in is manifested in different ways and at different levels, from a country's president to its parliament (e.g., Uganda), MOH (e.g., in Ethiopia), and state-level governments (e.g., in India and Nigeria). Once such leaders understand and value data, they will lend their vital support. Stakeholders in Uganda and Ghana noted the importance of buy-in from the national statistical office; the endorsement of PMA2020 by the Uganda Bureau of Statistics has lent credibility to the program, and PMA2020 staff in Ghana are making inroads with the Ghana Statistical Service.

Leadership buy-in opens the door to an increasing sense of **country ownership** of the data. Stakeholders expressed a wide range of views regarding the degree to which they felt that their country owns the PMA2020 data. Respondents from Ethiopia, Kenya, and Nigeria felt a strong sense of ownership by their country (or at the state level, in the case of Nigeria), while respondents from Ghana, India, and Indonesia perceive PMA2020 to be owned by third parties, such as a private-sector university (Ghana) or a U.S. university (India). Different respondents in Uganda—none from the government—expressed strong views in both directions. Some respondents offered simple ways to increase a sense of government ownership of

PMA2020 data: Government authorities should be the ones to issue invitations, convene the dissemination meetings, and present the data.

Respondents in at least three countries talked about **planned transition of PMA2020 to the government**—gradual transfer of responsibilities from the Gates Institute to the local PMA2020 team (already under way in DRC) or a well-crafted **exit strategy for gradual withdrawal of technical support** (Ghana, Nigeria). Two respondents reflected a contrasting view of country ownership and integration into their country's health or statistics structure. One PMA2020 staff member felt that the university should continue to own the PMA2020 process and data; another, from Nigeria, argued against the program being incorporated into the government statistics structure.

**Meeting policymakers' needs** is also a linchpin for sustainability. One way this is manifested is in generation of demand for data. Several respondents commented on this specifically, some in an aspirational sense (e.g., Burkina Faso, Ghana, Niger) and others as something that had been achieved (e.g., Nigeria, Uganda).

**Accountability**, or data use to inform policy, is another sustainability enabler. The use of PMA2020 data is described in more depth in Chapter Eleven. One respondent from Ghana noted that the more frequent PMA2020 data have "literally more or less determined the direction and pace of [family planning] activities," and another from Nigeria echoed this sentiment, noting that PMA2020 data have been used specifically for state-level planning.

**Stakeholder buy-in** is also important for sustainability, and respondents from all countries commented on key partners and stakeholders. Nearly all cited government entities (especially the MOH, statistics office, and subnational governments; some also mentioned schools of public health and a broader range of ministries, including planning and finance), as well as bilateral partners (e.g., USAID, DfID), multilateral partners (especially UNFPA, also WHO), and NGOs. Interestingly, respondents from many countries (e.g., Burkina Faso, Côte d'Ivoire, Ghana, Kenya, Niger, and Nigeria) also noted the importance of civil society. Perhaps of particular relevance, several countries have established multistakeholder working groups for planning and coordination (e.g., Burkina Faso and Nigeria, as well as Pakistan [not a PMA2020 country]). One respondent from Côte d'Ivoire summed up the indomitable spirit of collective action thusly: **"If the three types of groups work together—government, civil society, and other NGO partners—nothing is impossible."**

Another theme that emerged from the country interviews was not included in the original RAND sustainability-enabler framework but potentially should be: **visibility, dissemination of data, and advocacy**. As discussed in Chapter Eleven, advocacy is one dimension of data use; in principle, visibility, dissemination of data, and advocacy can also help enable the sustainability of data programs like PMA2020. These themes were mentioned by respondents in several countries, reflecting a spectrum from achievement (e.g., Kenya, Niger, Nigeria) to aspirational (e.g., Ghana, India). More than one respondent alluded to the visibility of PMA2020 at the top level of government (e.g., Niger, Ethiopia). Of particular interest is the leadership of Ethiopia's first lady in advocating for a major new child care center in one of the country's regions. This serves as a reminder of the potential role of first ladies in promulgating important social causes. The RAND African First Ladies' Initiative from 2010–2013 might offer further ideas for harnessing the visibility and credibility of first ladies for family planning broadly and related data programs more specifically (Brand, 2011). Others commented on the need to raise awareness and buy-in from other high-level authorities, such as commissioners in Nigeria or relevant institutes in India.

#### *Data Culture*

**A "data culture" is the ultimate aim of a mature, sustainable data system.** Such a culture involves normative change, as noted by astute interviewees in Indonesia and Niger. Respondents in most countries commented on the institutionalization of data use and integration of PMA2020 into their countries' own infrastructure. Such integration takes the form of integration into national planning (e.g., DRC), relevant government organizations—especially those responsible for population (e.g., India, Nigeria) or statistics (e.g., Ghana, Kenya)—and schools of public health (e.g., DRC, Uganda).

#### *Barriers to PMA2020 Sustainability*

As noted above, some barriers to PMA2020 sustainability are linked to the countries themselves, but others are linked directly to the program. **Cost was the most important barrier**, mentioned by PMA2020 staff and/or government officials in at least eight PMA2020 countries and probably a function of both the program and the country (e.g., Côte d'Ivoire, DRC, Ethiopia, Ghana, India, Kenya, Niger, Nigeria). At least two countries (Ethiopia, India) considered the **centralization of programming in Baltimore as a barrier to PMA2020 sustainability**. Reflecting the importance of the sustainability enablers discussed previously, some respondents in Ghana noted **inadequate technical capacity and lack of government ownership** as further barriers. Finally, several respondents (from Ethiopia, Ghana, and Indonesia) expressed concerns about the security or **privacy of data in the cloud**.

### Sustainability of Track20

In contrast with plentiful U.S. stakeholder comments on the PMA2020 program, neither Gates Foundation staff nor most of the other U.S.-based stakeholders commented on the sustainability of Track20. However, the Avenir team has given considerable thought to the program's sustainability, with one Avenir representative commenting that **"sustainability means that the M&E officer is there after [Track20] is gone."** Further, the **position would be integrated into the government's personnel system** (if not already established) and paid for by the government and/or other donors. Aside from that insightful comment, most of the perspectives about the sustainability of Track20 come from the in-country stakeholders, nearly all of whom were supportive of extending Track20's work beyond 2020. The most common request was for additional financing to support either additional M&E officers (especially at subnational levels) or capital costs, such as equipment, electricity, or travel.

#### *Financial Sustainability*

Compared with PMA2020, fewer respondents commented on the financial sustainability of Track20. In only one country (India) was there some degree of confidence that the government will continue to want the type of data provided through Track20 and potentially consider supporting ongoing efforts after Track20 support ends. Most other countries (e.g., Burkina Faso, Côte d'Ivoire, Kenya, Lao PDR, Uganda) would turn to bilateral or multilateral partners for financial support after Track20 programming ends.

#### *Technical Sustainability*

The current technical capacity of countries to carry out Track20 activities appears to range from relatively high in Nigeria ("For the first time, we can run annual estimates on our own") to relatively low, in particular a very shallow bench of experts beyond the M&E officer (e.g., Tanzania, Zimbabwe). The Track20 M&E officers appear to be well qualified and respected, but

respondents in several countries felt that **ongoing training and mentoring of these officers is vital to their sustainability**. At least one specifically mentioned the value of an on-site person to help train and mentor people at subnational level. Several also suggested that **more people within the MOH** (e.g., Kenya, Lao PDR, Uganda, Zimbabwe) and at **subnational levels** should be trained to use the FPET tool to develop their own estimates (e.g., Côte d'Ivoire, Lao PDR, Nigeria, Pakistan, Tanzania).

### *Operational Sustainability*

**Leadership buy-in** for Track20 is manifest in different ways across the program countries, from wanting routine, quality service statistics data (DRC), to using Track20 estimates in national planning (e.g., Pakistan, Tanzania), to the program responding to government needs and not vice versa (e.g., India). A multilateral agency representative in one country stated, "I believe Track20 is the backbone for the new architecture that we have built as part of FP2020."

Respondents from some countries already feel a strong sense of country **ownership** of Track20 (Ethiopia, India, Nigeria), while others did not (e.g., Indonesia, Zimbabwe), with one respondent in Zimbabwe noting that the government is not leading or directing the allocation of resources. Several respondents noted the **desire to build ownership of and trust in Track20 from the national level to the local level** (e.g., Burkina Faso, Pakistan, Tanzania). In one country (Lao PDR), the organizational location of Track20 staff in a relatively low-profile part of the MOH limited ownership and sustainability; in contrast, one respondent in Zimbabwe (not Track20 or government staff) commented that the M&E officer is placed too high in the MOH and should be at a lower level to be seen as better ingrained in the program. In an interesting twist on the concept of ownership, one respondent in Nigeria stated, **"I believe we should take ownership for sustainability" and called for explicit planning for sustainability**. Other respondents in the country recommended more government engagement at the consensus meetings, including more presentations and handling of questions and answers by government officials.

Respondents from most countries commented on key partners and **stakeholder buy-in**. They mentioned most of the same government, bilateral, multilateral, nongovernment, civil society, and working group partners noted previously for PMA2020.

Stakeholders also mentioned both achievements and opportunities for Track20 **visibility**. For example, the Community Engagement Working Group in Pakistan was an excellent venue in which to raise awareness of the technical expertise and enhance the reputation of Track20 for the country's family planning efforts; however, at least one government respondent noted that awareness of Track20 should be further raised by expanding involvement to more provinces. Others, such as in Tanzania, noted the need to involve high-level officials in consensus meetings to help ensure that they see the relevance of the information and of M&E more broadly.

### *Data Culture*

While very few respondents commented specifically on a data culture in the context of Track20, it was particularly interesting to note the institutionalization of FPET in India (by tweaking the country's software related to its HMIS database so that HMIS staff can directly generate the relevant core indicators) and Indonesia (FPET being used in government institutions). The fact that **nearly all Track20 M&E officers are already government employees already**

**embeds Track20 within government, an important anchor for sustainability** if the positions and salaries can be maintained after the formal Track20 program ends.

#### *Barriers to Track20 Sustainability*

While several respondents (e.g., from Côte d'Ivoire, Lao PDR, Zimbabwe) mentioned funding as a barrier to sustainability, others mentioned **inadequate capacity, including too thin a bench of experts** (e.g., Côte d'Ivoire, India, Lao PDR).

## Analyses of Sustainability Ratings from the 15 Program Countries

The RAND research team developed and tested a new framework of sustainability-enabling factors for PMA2020 and Track20. We considered the initial framework as a proof of concept for assessing progress toward sustainable data systems and data use. We tested it during our field interviews to complement the qualitative information from those interviews and assess its utility as a tool that relevant stakeholders can continue to use periodically to assess further progress. Because the proposed sustainability-enabling factors are not inherently limited to family planning, they are potentially applicable to other health development programming that involves the generation and use of data for informed decisionmaking—for example, across different Gates Foundation programs, as well as programs sponsored by bilateral and multilateral development organizations.

To assess the sustainability ratings, we employed a process similar to that used to evaluate data maturity. We analyzed the 136 completed ratings from the personnel we interviewed to look at overall results across the different categories, both by responder group (PMA2020, Track20, and government/NGO officials) and by country. Again, certain categories, and sustainability-enabling factors within those categories, emerged as rated relatively strong or weak, with considerable variation by factor, country, and respondent group. These findings are presented in the following sections.

### Ratings by Theme and Stakeholder Group

Assessing the sustainability data from a highly aggregated level, several broad trends appear (Table 13.1; most factors were rated on a scale from 1 to 5, where 1 [red] is the lowest and 5 [green] is the highest). As with data maturity, PMA2020 respondents tended to rate PMA2020 sustainability more highly than did Track20 respondents or government and NGO officials. The overall average Track20 self-assessed ratings and ratings by government/NGO officials for both programs were more comparable to one another.

Financing of PMA2020 and Track20 is overwhelmingly by donors. The most relevant ratings are in the categories of technical, operational, and data culture–related sustainability-enabling factors. In citing the specific areas of strength and weakness highlighted by the respondents, the category is listed first, with the specific sustainability-enabling factors following in parentheses. The overall categories rated highly across all stakeholder groups include several data culture factors (data impact on outcomes, trust in PMA2020 data, accuracy of PMA2020 and Track20 analytics, and data accessibility) and a few operational factors (donor buy-in, leadership buy-in, national government buy-in, effectiveness of multistakeholder body, and cultural acceptability). The first group of factors points to the quality, availability, and impact of the data being generated by PMA2020 and Track20. The second group of factors

**Table 13.1**
**Sustainability Ratings of PMA2020 and Track20, by Factor and Stakeholder Group**

| | | PMA2020 | | Track20 | | |
|---|---|---|---|---|---|---|
| Category | Sustainability-Enabling Factor | PMA2020 Self-Assessment | Government/ NGO Assessment | Track20 Self-Assessment | Government/ NGO Assessment | Overall Average |
| Financial | Donor financing (%) | 96.4 | 91.6 | 93.8 | 91.6 | 93.4 |
| | Domestic resourcing (%) | 5.0 | 10.5 | 10.1 | 10.5 | 9.0 |
| Technical | Data system and analytics capabilities | 4.3 | 3.5 | 3.6 | 3.5 | 3.7 |
| | Hardware maintenance | 4.3 | 3.2 | 3.3 | 3.1 | 3.5 |
| | Software maintenance | 4.2 | 2.8 | 2.9 | 2.9 | 3.2 |
| Operational/ programmatic | Leadership buy-in | 4.3 | 3.7 | 4.0 | 3.7 | 3.9 |
| | Country ownership of data | 3.9 | 3.7 | 3.6 | 3.6 | 3.7 |
| | Satisfaction of policymakers' data needs (PMA2020) | 3.5 | 3.2 | 3.4 | 3.6 | 3.4 |
| | Satisfaction of policymakers' data needs (Track20) | 3.7 | 3.4 | 3.7 | 3.7 | 3.6 |
| | Accountability | 3.6 | 3.3 | 3.4 | 3.8 | 3.5 |
| | Stakeholder buy-in (planning): | | | | | |
| | (a) local communities | 3.1 | 2.6 | 3.3 | 2.6 | 2.9 |
| | (b) subnational government | 3.5 | 3.4 | 3.5 | 3.4 | 3.5 |
| | (c) national government | 3.9 | 3.8 | 4.0 | 3.8 | 3.9 |
| | (d) civil society (community organizations) | 3.1 | 3.4 | 3.4 | 3.4 | 3.3 |
| | (e) donors | 4.3 | 4.1 | 4.1 | 4.1 | 4.2 |
| | Multistakeholder planning body | 20 yes<br>2 no | 65 yes<br>4 no | 20 yes<br>0 no | 65 yes<br>4 no | |
| | Effectiveness | 3.6 | 3.9 | 3.7 | 3.9 | 3.8 |
| | Collaboration | 3.9 | 3.5 | 3.6 | 3.5 | 3.6 |
| | Cultural acceptability | 4.2 | 3.9 | 3.8 | 3.9 | 3.9 |
| | Use of local expertise | 4.4 | 3.1 | 3.0 | 3.1 | 3.4 |
| Data culture | Data accessibility | 4.6 | 4.2 | 4.2 | 3.4 | 4.1 |
| | Trust in data (PMA2020) | 4.7 | 4.2 | 3.7 | 4.2 | 4.2 |
| | Analytics accuracy | 4.8 | 3.9 | 4.4 | 3.9 | 4.2 |
| | Data impact on outcomes | 4.7 | 4.2 | 4.7 | 4.3 | 4.5 |
| | Institutionalization of data use | 3.5 | 3.0 | 3.8 | 3.4 | 3.4 |
| Overall averages | | 4.0 | 3.5 | 3.7 | 3.6 | 3.7 |

NOTE: Factors were rated on a scale from 1 to 5, where 1 (red) is the lowest and 5 (green) is the highest.

indicates broad consensus at the national level and belief in the program and its goals. Having a respected, solid product that has the faith of high-level personnel is critical to achieving lasting change in family planning programs and targets.

However, a few weak areas also emerged from the data. Respondents rated the following factors as particularly weak: technical (hardware and software maintenance), operational (local community and civil society buy-in, use of local expertise), and data culture (institutionalization of data use [for PMA2020]). These weaknesses demonstrate that this **national-level buy-in has not yet translated to buy-in at the local level** and that **institutionalization of data use has not yet fully matured**. Local communities and civil society appear to not have been consulted or implicated in these programs to the degree that they should be. This lack of program diffusion may impair family planning program results and sustainability at the local level or away from the capital, which is where the most critical progress remains to be made in most developing countries. Overall, the average ratings for country ownership of PMA2020 data were lower than the average for sustainability-enabling factors rated as relatively strong and higher than the average for factors rated as relatively weak.

Using government/NGO ratings to validate self-assessed ratings by PMA2020 and Track20 program staff, government/NGO respondents rated 20 of 22 sustainability-enabling factors lower than did PMA2020 staff, but only 11 of 22 factors lower than did Track20 staff. The more-comparable ratings by government/NGO respondents and Track20 staff are perhaps not surprising, since Track20 staff work directly with and within government and, indeed, are typically government employees themselves. The lower ratings by government/NGO respondents for accessibility and analytics accuracy could indicate that the reports and family planning statistics that PMA2020 and Track20 staff members believe are being communicated to government officials or shared with NGOs are not reaching their desired audiences. It could also mean that these same officials have some lingering doubts as to the accuracy of these reports and statistics. There seems to be a good degree of faith in Track20 on the part of the government/NGO community, but these representatives felt that a clearer explanation of methodology and data processes would be helpful.

### Ratings by Theme and Country

#### *Overall Ratings by All Stakeholder Groups*

Overall ratings—by all stakeholder groups combined—about the sustainability of PMA2020 and Track20, by country from the ten countries with completed information, suggest overall higher ratings for certain countries—in particular, Uganda, Ghana, and India—and the lowest ratings for DRC (Table 13.2). Notable low-rated sustainability-enabling factors included (poor) use of local expertise in Lao PDR and Pakistan and (low) trust of data analytics in Niger.

#### *Technical Sustainability*

Technical sustainability elements received relatively **low ratings, especially for hardware and software maintenance**. In most countries, ratings for the sustainability of data systems and analytics were higher for PMA2020 than for Track20. The highest ratings were for Burkina Faso and Uganda, and the lowest ratings were for Tanzania.

#### *Operational Sustainability*

**The strongest operational sustainability-enabling factors were buy-in from donors, leaders, and national governments, and the lowest ratings were for buy-in from local**

**Table 13.2**
**Ratings from All Stakeholders of Sustainability-Enabling Factors, by Country**

| Category | Sustainability-Enabling Factor | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Overall Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Data system and analytics capabilities: | | | | | | | | | | | | | | | | |
| Technical | PMA2020 | 4.0 | 4.5 | 3.3 | 3.9 | 4.0 | 4.2 | 4.0 | 4.5 | 3.0 | 4.3 | 4.0 | 3.1 | 3.3 | 4.5 | 3.9 | 3.7 |
| | Track20 | 2.0 | 4.0 | 2.5 | | 4.4 | 4.1 | 3.7 | 3.5 | 2.8 | 4.5 | 3.2 | | 3.5 | 4.0 | | |
| | IT maintenance: | | | | | | | | | | | | | | | | |
| | Hardware vendor available | 5.0 | 3.5 | 3.0 | 3.7 | 5.0 | 4.1 | 3.0 | 4.0 | 3.5 | 2.7 | 4.2 | 2.9 | 2.0 | 5.0 | 3.0 | 3.5 |
| | Software vendor available | 3.0 | 3.5 | 3.3 | 4.0 | 4.3 | 4.0 | 3.0 | 4.0 | 5.0 | 4.5 | 4.3 | 2.9 | 2.0 | 3.5 | 3.0 | |
| | Mechanism in place: | | | | | | | | | | | | | | | | |
| | Hardware | 3.0 | 3.2 | 3.3 | 3.3 | 4.5 | 4.3 | 3.3 | 4.0 | 3.0 | 2.7 | 3.9 | 2.8 | 2.5 | 5.0 | 3.0 | 3.2 |
| | Software | 5.0 | 3.3 | 3.3 | 4.0 | 4.2 | 4.1 | 3.3 | 4.0 | 3.0 | 4.3 | 4.1 | 2.8 | 2.0 | 5.0 | 3.0 | |
| Operational/ programmatic | Leadership buy-in | 5.0 | 4.1 | 2.7 | 4.5 | 3.7 | 4.3 | 4.4 | 4.7 | 3.8 | 4.1 | 3.8 | 3.1 | 3.6 | 4.3 | 4.4 | 3.9 |
| | Country ownership of data: | | | | | | | | | | | | | | | | |
| | PMA2020 | 4.0 | 4.5 | 3.0 | 4.1 | 3.9 | 4.3 | 4.2 | 3.3 | | 3.7 | 3.2 | 3.7 | | 3.5 | | 3.7 |
| | Track20 | 3.0 | 3.0 | 2.5 | | | 3.7 | | 4.2 | 3.2 | 5.0 | 3.4 | | 3.7 | | 4.4 | |
| | Policymakers' use of data: | | | | | | | | | | | | | | | | |
| | PMA2020 | 3.5 | 3.0 | 2.3 | 3.3 | 3.7 | 3.6 | 3.8 | 3.8 | | 2.7 | 3.5 | 3.7 | | 3.3 | | 3.4 |
| | Track20 | 4.0 | 3.0 | 2.8 | | | 3.5 | 3.0 | 3.8 | 3.3 | 3.8 | 3.5 | 4.0 | 3.0 | | 4.0 | |

**Table 13.2—continued**

| Category | Sustainability-Enabling Factor | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Overall Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Policymakers' satisfaction with data: | | | | | | | | | | | | | | | | |
| | PMA2020 | 3.0 | 3.5 | 3.0 | 3.6 | 4.2 | 4.0 | 3.7 | 3.3 | | 2.8 | 3.6 | 3.3 | 3.5 | 3.5 | 4.0 | 3.6 |
| | Track20 | 3.0 | 3.0 | 3.5 | | | 3.5 | | 4.0 | 3.3 | 3.5 | 4.0 | | 4.3 | | 2.9 | |
| | Accountability: | | | | | | | | | | | | | | | | |
| | PMA2020 | 4.5 | 4.5 | 2.5 | 3.3 | 3.7 | 4.2 | 3.8 | 4.2 | | 3.1 | 2.5 | 3.7 | | 3.8 | | 3.5 |
| | Track20 | 4.0 | 3.8 | 2.0 | | | 4.2 | | 4.3 | 3.3 | 5.0 | 3.1 | | 3.3 | | 4.1 | |
| | Stakeholder buy-in (planning): | | | | | | | | | | | | | | | | |
| | (a) local community | 2.5 | 2.5 | 1.8 | 2.6 | 3.7 | 3.6 | 2.4 | 3.3 | 2.7 | 2.8 | 3.0 | 1.3 | 2.8 | 3.3 | 2.8 | 2.9 |
| | (b) subnational government | 2.5 | 2.1 | 2.2 | 2.9 | 4.0 | 4.0 | 2.9 | 3.7 | 3.5 | 2.7 | 3.8 | 3.6 | 2.9 | 4.0 | 4.0 | 3.5 |
| | (c) national government | 4.0 | 4.6 | 2.7 | 3.7 | 3.7 | 4.4 | 3.7 | 4.1 | 3.5 | 4.0 | 3.4 | 3.9 | 3.7 | 4.3 | 4.4 | 3.9 |
| | (d) civil society (community organizations) | 3.0 | 4.3 | 2.0 | 3.1 | 3.9 | 3.3 | 3.0 | 2.6 | 2.8 | 2.9 | 3.5 | 3.6 | 2.9 | 4.3 | 4.1 | 3.3 |
| | (e) donors | 3.5 | 4.8 | 3.8 | 4.0 | 4.3 | 4.3 | 4.1 | 3.9 | 3.8 | 4.1 | 4.2 | 4.4 | 4.3 | 3.8 | 4.5 | 4.2 |
| | Effectiveness | 3.5 | 4.3 | 3.7 | 3.5 | 4.0 | 3.5 | 3.3 | 4.2 | 3.5 | 4.0 | 3.3 | 3.9 | 4.1 | 3.8 | 4.6 | 3.8 |
| | Collaboration | 4.0 | 3.7 | 2.7 | 3.7 | 3.6 | 3.7 | 3.9 | 3.3 | 3.6 | 3.5 | 3.4 | 2.9 | 3.8 | 4.2 | 4.2 | 3.6 |
| | Cultural acceptability | 3.5 | 4.3 | 3.0 | 3.9 | 4.5 | 4.2 | 4.2 | 4.4 | 4.0 | 3.2 | 3.5 | 3.2 | 4.4 | 3.8 | 4.3 | 3.9 |
| | Use of local expertise | 3.0 | 3.8 | 3.3 | 4.0 | 4.1 | 4.4 | 4.4 | 3.4 | 2.0 | 3.1 | 2.9 | 1.9 | 3.5 | 4.5 | 3.0 | 3.4 |

**Table 13.2—continued**

| Category | Sustainability-Enabling Factor | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Overall Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data culture | Data accessibility: | | | | | | | | | | | | | | | | |
| | PMA2020 | 4.0 | * | 3.7 | 4.8 | 4.5 | 4.3 | 3.3 | 5.0 | | 4.5 | 4.7 | | | 5.0 | | 4.1 |
| | Track20 | 3.5 | 4.3 | 3.5 | 3.0 | | 4.2 | 3.8 | 3.8 | 3.0 | 3.5 | 3.9 | 3.7 | 4.0 | 4.2 | 3.6 | |
| | Trust in data accuracy: PMA2020 | 3.5 | * | 3.8 | 4.8 | 4.7 | 4.0 | 4.3 | 4.5 | | 3.7 | 4.5 | | | 5.0 | | 4.2 |
| | Trust in analytics accuracy: | | | | | | | | | | | | | | | | |
| | PMA2020 | 4.0 | * | 4.0 | 4.6 | 4.8 | 4.1 | 4.3 | 4.5 | | 2.8 | 4.5 | 3.6 | | 5.0 | | 4.2 |
| | Track20 | 4.5 | 4.3 | 3.8 | 4.0 | | 4.4 | 3.9 | 4.5 | 3.4 | 3.2 | 4.5 | 3.8 | 4.3 | 4.0 | 4.1 | |
| | Data impact on outcomes: | | | | | | | | | | | | | | | | |
| | PMA2020 | 4.0 | * | 3.7 | 4.7 | 4.6 | 4.2 | 4.2 | 4.5 | | 4.3 | 4.4 | | | 4.5 | | 4.5 |
| | Track20 | 4.0 | 4.3 | 4.3 | 4.8 | | 4.3 | 4.3 | 4.7 | 3.6 | 4.5 | 4.6 | 4.7 | 4.5 | 4.0 | 4.8 | |
| | Institutionalization of data use: | | | | | | | | | | | | | | | | |
| | PMA2020 | 3.0 | * | 2.7 | 3.0 | 3.6 | 3.2 | 3.3 | 3.5 | | 2.3 | 3.3 | | | 4.5 | | 3.4 |
| | Track20 | 3.5 | 4.2 | 3.5 | 2.7 | | 3.7 | 3.4 | 3.6 | 2.8 | 4.0 | 3.0 | 3.7 | 3.3 | 3.2 | 4.1 | |
| Average score by country | | 3.6 | 3.9 | 3.1 | 3.7 | 4.1 | 4.0 | 3.7 | 3.9 | 3.4 | 3.6 | 3.7 | 3.4 | 3.5 | 4.1 | 3.9 | 3.7 |

NOTE: Factors were rated on a scale from 1 to 5, where 1 (red) is the lowest and 5 (green) is the highest.

* Data use factors for PMA2020 in Côte d'Ivoire are not included because the first survey had not been conducted at the time of the RAND team's interviews.

**communities and civil society**. Cultural acceptability and the effectiveness of multistakeholder coordination bodies (where present) were also relatively high-rated sustainability-enabling factors. Operational sustainability was strongest in Uganda and weakest in DRC.

#### *Data Culture*

The **highest-rated sustainability-enabling factor in this domain was impact of data on outcomes, and the lowest-rated was institutionalization of data use**. Sustainability-enabling factors related to data culture were strongest in Uganda, Nigeria, Côte d'Ivoire, and Kenya and weakest in DRC.

### PMA2020 Personnel Ratings, by Country

Table 13.3 provides more granular detail of sustainability-enabling factors by country, as perceived by PMA2020 country staff. Across countries, these staff rated some categories and factors relatively higher than others: technical factors (data system and analytics capabilities, hardware and software maintenance), operational/programmatic factors (leadership buy-in, donor buy-in, cultural acceptability, use of local expertise), and data culture factors (data accessibility, trust in data analytics and accuracy, data impact on outcomes). PMA2020 respondents believe that technical program features, including the actual maintenance of software and hardware and the usage of these systems, are working well. PMA2020 respondents rated fewer factors as relatively weak. Some of these weak factors—buy-in by different stakeholders—are rated especially low in certain countries (Côte d'Ivoire, DRC, Kenya). Ratings from DRC PMA2020 respondents for nearly all factors were well below those from their counterparts in other countries, with leadership and stakeholder buy-in and satisfaction of policymakers' data needs being particular areas of concern.

### Track20 Personnel Ratings, by Country

Country-specific data from the 13 Track20 countries from which ratings were completed indicates that Track20 staff in four countries rated the sustainability of Track20 higher than did their counterparts in other countries: Uganda, Burkina Faso, Côte d'Ivoire, and Zimbabwe (Table 13.4). DRC, Niger, and Tanzania scored below their peers for Track20 sustainability. Again, this table provides more granular detail of sustainability-enabling factors by country—for example, it highlights particularly problematic factors in Niger, Nigeria, Pakistan, and Uganda.

### Country-Specific Perceptions of Government/NGO Personnel About PMA2020 and Track20 Sustainability

As noted above, government/NGO personnel effectively validated the majority of PMA2020 and Track20 self-assessed sustainability ratings, but their ratings were consistently lower, with more weak categories across the board. Their country-specific ratings (Table 13.5) indicate particularly problematic sustainability-enabling factors in certain countries, some of which highlight problems different from those identified by PMA2020 or Track20 staff (e.g., satisfaction of policymakers' needs by PMA2020 in Côte d'Ivoire [with the caveat that the first round of PMA2020 data collection was currently under way in this country at the time of our interviews], stakeholder buy-in in Uganda). Interestingly, government/NGO stakeholders rated a different set of countries as higher than others: Ghana, Kenya, and Zimbabwe. The three coun-

**Table 13.3**
**PMA2020 Personnel Ratings of Sustainability-Enabling Factors, by Country**

| Category | Sustainability-Enabling Factor | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Niger | Nigeria | Uganda | Overall Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Technical | Data system and analytics capabilities | 5.0 | 3.5 | 4.3 | 4.0 | 4.5 | 4.0 | 5.0 | 5.0 | 4.3 | 4.5 | 4.3 |
| | IT maintenance: | | | | | | | | | | | |
| | Hardware vendor available | 5.0 | 3.5 | 3.7 | 5.0 | 5.0 | 3.0 | 5.0 | 4.3 | 4.2 | 5.0 | 4.3 |
| | Software vendor available | 5.0 | 3.5 | 4.0 | 4.3 | 4.8 | 3.0 | 5.0 | 4.0 | 4.0 | 5.0 | 4.1 |
| | Hardware mechanism in place | 5.0 | 3.5 | 3.3 | 4.5 | 5.0 | 3.7 | 5.0 | 4.3 | 4.3 | 5.0 | 4.2 |
| | Software mechanism in place | 5.0 | 3.5 | 4.0 | 4.2 | 4.8 | 3.3 | 5.0 | 3.7 | 4.2 | 5.0 | 4.1 |
| Operational/ programmatic | Leadership buy-in | 5.0 | 2.5 | 4.4 | 4.0 | 4.9 | 5.0 | 5.0 | 4.3 | 4.0 | 5.0 | 4.3 |
| | Country ownership of data | 5.0 | 2.5 | 3.9 | 3.8 | 4.7 | 4.0 | 3.0 | 3.7 | 3.4 | 5.0 | 3.9 |
| | Satisfaction of policymakers' data needs | 5.0 | 2.0 | 2.7 | 3.8 | 4.5 | 3.7 | 4.0 | 3.3 | 3.4 | 3.0 | 3.5 |
| | Stakeholder buy-in (planning): | | | | | | | | | | | |
| | (a) local community | 1.0 | 1.0 | 3.5 | 4.0 | 4.5 | 3.0 | 2.0 | 3.3 | 2.0 | 5.0 | 3.1 |
| | (b) subnational government | 1.0 | 2.0 | 3.8 | 4.0 | 4.8 | 2.3 | 3.0 | 4.0 | 3.4 | 5.0 | 3.5 |
| | (c) national government | 5.0 | 3.0 | 4.0 | 4.0 | 5.0 | 4.3 | 4.0 | 4.0 | 2.0 | 5.0 | 3.9 |
| | (d) civil society (community organizations) | 4.0 | 2.0 | 2.5 | 4.0 | 4.5 | 4.0 | 1.0 | 2.3 | 2.4 | 5.0 | 3.1 |
| | (e) donors | 5.0 | 4.5 | 4.4 | 4.7 | 4.5 | 4.3 | 3.0 | 4.7 | 3.5 | 4.0 | 4.3 |

**Table 13.3—continued**

| Category | Sustainability-Enabling Factor | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Niger | Nigeria | Uganda | Overall Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cultural acceptability | 4.0 | 2.5 | 4.5 | 4.8 | 4.8 | 4.7 | 5.0 | 2.7 | 3.7 | 4.0 | 4.2 |
| | Use of local expertise | 5.0 | 4.5 | 4.3 | 4.3 | 4.8 | 5.0 | 4.0 | 3.7 | 4.2 | 5.0 | 4.4 |
| Data culture | Data accessibility | 4.0 | 3.5 | 4.5 | 4.7 | 5.0 | 4.7 | 5.0 | 5.0 | 4.8 | 5.0 | 4.6 |
| | Trust in data accuracy: PMA2020 | 3.0 | 4.5 | 4.8 | 4.7 | 4.8 | 4.5 | 5.0 | 4.3 | 5.0 | 5.0 | 4.7 |
| | Trust in analytics accuracy: | | | | | | | | | | | |
| | PMA2020 | 5.0 | 5.0 | 4.6 | 4.8 | 4.9 | 4.5 | 5.0 | 4.3 | 4.8 | 5.0 | 4.8 |
| | Track20 | | | 4.0 | | 4.5 | 4.5 | 5.0 | 2.0 | 5.0 | | 4.3 |
| | Data impact on outcomes: | | | | | | | | | | | |
| | PMA2020 | 5.0 | 4.5 | 4.7 | 4.8 | 4.6 | 4.3 | 5.0 | 4.0 | 4.8 | 5.0 | 4.7 |
| | Track20 | | | 5.0 | | 4.8 | 5.0 | 5.0 | 4.0 | 5.0 | 4.0 | 4.4 |
| | Institutionalization of data use | 3.0 | 2.5 | 3.0 | 4.0 | 3.8 | 3.7 | 4.0 | 4.0 | 3.0 | 4.0 | 3.5 |
| Average score by country | | 4.3 | 3.1 | 4.0 | 4.3 | 4.6 | 3.9 | 4.1 | 3.8 | 3.8 | 4.5 | 4.0 |

NOTE: Factors were rated on a scale from 1 to 5, where 1 (red) is the lowest and 5 (green) is the highest.

**Table 13.4**
**Track20 Personnel Ratings of Sustainability-Enabling Factors, by Country**

| Category | Sustainability-Enabling Factor | Burkina Faso | Côte d'Ivoire | DRC | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Average Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Technical | IT maintenance: | | | | | | | | | | | | | | |
| | Hardware vendor available | 5.0 | 2.5 | 2.5 | 3.3 | 3.0 | 3.0 | 4.0 | 1.0 | 3.5 | 3.0 | 2.0 | 2.0 | 3.0 | 3.0 |
| | Software vendor available | 3.0 | 2.5 | 3.0 | 3.3 | 3.0 | 3.0 | | 5.0 | 4.0 | 3.0 | 2.0 | 5.0 | 3.0 | 3.3 |
| | Hardware mechanism in place | 3.0 | 1.5 | 3.0 | 3.5 | 3.0 | 3.0 | | 1.0 | 2.3 | 3.0 | 2.5 | 2.0 | 3.0 | 2.6 |
| | Software mechanism in place | 5.0 | 2.0 | 3.0 | 3.5 | | 3.0 | | 5.0 | 3.0 | 3.0 | 2.0 | 1.0 | 3.0 | 2.9 |
| Operational/ programmatic | Leadership buy-in | 5.0 | 3.5 | 3.5 | 4.0 | 4.0 | 5.0 | | 4.0 | 4.2 | 3.0 | 3.5 | 5.0 | 4.5 | 4.0 |
| | Country ownership of data | 5.0 | | 4.0 | | 4.0 | 3.0 | | 4.0 | | 4.0 | | | | 4.0 |
| | Satisfaction of policymakers' data needs | 4.0 | | 2.5 | | 3.0 | 3.0 | | 4.0 | | 4.0 | | | | 3.3 |
| | Stakeholder buy-in: | | | | | | | | | | | | | | |
| | (a) local communities | 4.0 | 3.5 | 3.0 | 3.4 | 2.0 | 5.0 | | 1.0 | 4.0 | 1.0 | 2.5 | 5.0 | 3.0 | 3.3 |
| | (b) subnational government | 4.0 | 2.0 | 2.5 | 3.8 | 3.0 | 4.0 | | 1.0 | 4.2 | 4.0 | 2.5 | 5.0 | 4.0 | 3.5 |
| | (c) national government | 4.0 | 5.0 | 3.0 | 4.0 | 4.0 | 4.0 | | 4.0 | 4.2 | 4.0 | 3.5 | 5.0 | 4.5 | 4.0 |
| | (d) civil society | 3.0 | 5.0 | 2.0 | 2.4 | 3.0 | 4.0 | | 3.0 | 4.2 | 4.0 | 2.0 | 5.0 | 4.0 | 3.4 |
| | (e) donors | 4.0 | 5.0 | 4.0 | 4.0 | 5.0 | 4.0 | | 4.0 | 4.4 | 5.0 | 4.3 | | 4.5 | 4.1 |

**Table 13.4—continued**

| Category | Sustainability-Enabling Factor | Burkina Faso | Côte d'Ivoire | DRC | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Average Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Use of local expertise | 3.0 | 4.0 | 3.0 | 4.0 | 4.0 | 3.0 | | 3.0 | 1.3 | 2.0 | 3.0 | 5.0 | 4.0 | 3.0 |
| Data culture | Data accessibility | 4.0 | 5.0 | 3.0 | 4.6 | 4.0 | 4.0 | | 4.0 | 4.2 | 4.0 | 4.0 | 5.0 | 4.0 | 4.2 |
| | Accuracy: | | | | | | | | | | | | | | |
| | PMA2020 | 4.0 | 5.0 | 4.0 | 3.3 | 4.0 | 4.0 | | 3.0 | 4.3 | 4.0 | | 5.0 | | 4.0 |
| | Track20 | 4.0 | 5.0 | 4.0 | 4.8 | 4.0 | 4.0 | | 4.0 | 4.4 | 4.0 | 4.5 | 5.0 | 4.5 | 4.4 |
| | Data impact on outcomes: | | | | | | | | | | | | | | |
| | PMA2020 | 4.0 | 4.0 | 4.0 | 3.7 | 4.0 | 4.0 | | 4.0 | 3.8 | | | 5.0 | | 4.2 |
| | Track20 | 4.0 | 5.0 | 4.0 | 5.0 | 4.0 | 4.0 | | 5.0 | 4.2 | 5.0 | 4.5 | 5.0 | 4.5 | 4.7 |
| Average score per country | | 4.0 | 4.1 | 3.3 | 3.7 | 3.6 | 3.7 | 3.5 | 3.4 | 3.5 | 3.5 | 3.3 | 4.6 | 3.9 | 3.7 |

NOTE: Factors were rated on a scale from 1 to 5, where 1 (red) is the lowest and 5 (green) is the highest.

**Table 13.5**
**Government/NGO Personnel Ratings of Sustainability-Enabling Factors, by Country**

| Category | Sustainability-Enabling Factor | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Average Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Technical | Data system and analytics capabilities | 2.0 | 4.0 | 2.5 | | 4.4 | 4.1 | 3.7 | 3.5 | 2.8 | 4.5 | 3.2 | | 3.5 | 4.0 | | 3.5 |
| Operational/ programmatic | Country ownership of data: | | | | | | | | | | | | | | | | |
| | PMA2020 | 3.0 | 4.0 | 2.5 | 4.3 | 4.0 | 3.9 | 4.5 | 4.0 | | 3.5 | 3.0 | 3.3 | | 2.0 | | 3.7 |
| | Track20 | 3.0 | 2.0 | 2.5 | | | 3.8 | | 4.2 | 3.2 | 5.0 | 3.7 | | 3.4 | | 4.3 | 3.6 |
| | Satisfaction of policymakers' data needs: | | | | | | | | | | | | | | | | |
| | PMA2020 | 3.0 | 1.0 | 2.0 | 4.0 | 3.6 | 2.6 | 3.7 | 3.5 | | 2.0 | 3.5 | 3.3 | | 3.5 | | 3.2 |
| | Track20 | 3.0 | 3.0 | 3.5 | 3.5 | 4.0 | 3.4 | 4.0 | 3.7 | | 2.0 | 4.1 | 3.3 | | 3.0 | | 3.4 |
| | Accountability | 4.0 | 4.0 | 2.0 | 2.8 | 3.6 | 3.8 | 3.5 | 4.5 | | 2.0 | 2.5 | 3.3 | | 3.5 | | 3.3 |
| | Stakeholder buy-in (planning): | | | | | | | | | | | | | | | | |
| | (a) local communities | 1.0 | 3.0 | 1.5 | 1.7 | 3.4 | 2.9 | 2.2 | 3.0 | 2.7 | 4.0 | 3.0 | 1.7 | 3.0 | 1.7 | 2.6 | 2.6 |
| | (b) subnational government | 1.0 | 3.3 | 2.0 | 2.0 | 4.0 | 3.5 | 3.3 | 4.0 | 3.5 | 3.0 | 3.7 | 3.2 | 3.3 | 3.0 | 4.0 | 3.4 |
| | (c) national government | 4.0 | 3.7 | 2.0 | 3.3 | 3.4 | 4.1 | 2.8 | 4.3 | 3.5 | 4.0 | 3.9 | 3.9 | 3.8 | 3.7 | 4.3 | 3.8 |
| | (d) civil society (community organizations) | 3.0 | 4.0 | 2.0 | 3.7 | 3.8 | 3.1 | 2.0 | 2.8 | 2.8 | 3.5 | 3.9 | 3.1 | 3.8 | 3.7 | 4.3 | 3.4 |
| | (e) donors | 3.0 | 4.3 | 3.0 | 3.7 | 4.0 | 4.4 | 3.0 | 4.7 | 3.8 | 3.5 | 4.6 | 3.9 | 4.4 | 3.7 | 4.5 | 4.1 |

**Table 13.5—continued**

| Category | Sustainability-Enabling Factor | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Average Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Collaboration | 3.0 | 4.0 | 2.0 | 2.7 | 3.2 | 3.6 | 4.0 | 4.0 | 3.6 | 3.5 | 3.3 | 2.8 | 3.6 | 4.3 | 4.4 | 3.5 |
| | Use of local expertise | 3.0 | 2.5 | 2.5 | 3.8 | 3.8 | 4.5 | 4.3 | 3.3 | 2.0 | 2.5 | 3.1 | 1.8 | 4.0 | 4.0 | 2.0 | 3.1 |
| Data culture | Data accessibility: | | | | | | | | | | | | | | | | |
| | PMA2020 | 4.0 | | 4.5 | 5.0 | 4.4 | 3.6 | | | | 4.0 | 4.5 | | | 5.0 | | 4.2 |
| | Track20 | 3.0 | 3.5 | 4.0 | 3.0 | | 3.8 | 3.0 | 3.7 | 3.0 | 3.0 | 3.6 | 3.3 | 3.9 | 3.3 | 3.3 | 3.4 |
| | Trust in data/analytics accuracy | 4.0 | 5.0 | 3.0 | | | | | | | | 5.0 | | | 5.0 | | 4.2 |
| | Data impact on outcomes: | | | | | | | | | | | | | | | | |
| | PMA2020 | 4.0 | 4.0 | 2.5 | | 4.4 | 4.3 | | | | 5.0 | 4.7 | | | 4.0 | | 4.2 |
| | Track20 | 4.0 | 3.7 | 4.5 | 4.5 | | 3.3 | 4.0 | 5.0 | 3.6 | 4.5 | 4.6 | 4.3 | 4.4 | 4.0 | 5.0 | 4.3 |
| | Institutionalization of data use: | | | | | | | | | | | | | | | | |
| | PMA2020 | 3.0 | 4.0 | 2.0 | | 3.2 | 2.9 | | | | 1.0 | 3.0 | | | 5.0 | | 3.0 |
| | Track20 | 3.0 | 3.3 | 4.0 | 2.7 | | 3.4 | 3.7 | 3.7 | 2.8 | 4.0 | 3.1 | 3.4 | 3.3 | 3.3 | 4.3 | 3.4 |
| Average score per country | | 3.3 | 3.5 | 2.8 | 3.5 | 3.9 | 3.7 | 3.6 | 4.0 | 3.3 | 3.4 | 3.9 | 3.2 | 3.8 | 3.7 | 3.9 | 3.6 |

NOTE: Factors were rated on a scale from 1 to 5, where 1 (red) is the lowest and 5 (green) is the highest.

tries they rated as weakest (Burkina Faso, DRC, and Pakistan) overlapped more with countries rated as weaker by PMA2020 and Track20 program staff.

### Conclusions from Analyses of Sustainability Ratings

With the notable exception of the PMA2020 self-assessment of its technical capabilities, the sustainability ratings across the three different sets of respondents were largely comparable. The **sustainability-enabling factors considered strongest across countries and respondent groups have to do with leadership at the national level (leadership buy-in, national government buy-in, and donor buy-in**). Conversely, the **factors considered weakest across the board involve local engagement (local community buy-in, civil society buy-in, and use of local expertise) and local systems support (hardware and software maintenance)**. Country ownership of data and institutionalization of data use have considerable room for improvement. These factors weigh into the recommendations offered in Chapter Fourteen. It is undoubtedly important to have the support of the national government and key individuals within the MOHs, ministries of planning, and other key departments, as well as the donor community. However, if the PMA2020 and Track20 are to be truly sustainable, subnational and local authorities must understand and value their products and expertise as well, and they must feel a sense of ownership and use data for their own leadership and management purposes. PMA2020 and Track20 staff would do well to extend the scope (or at least knowledge and understanding) of their programs and products and involve subnational and local authorities and community leaders more directly.

In spite of the diversity of stakeholders and the relative subjectivity of these ratings, we can draw a few conclusions. The overall average rating for 12 of the 15 countries was 3.5 or greater (on a scale of 5), indicating a substantial foundation for sustainability but yet room for improvement (Table 13.6); the three countries with lower overall ratings (DRC, Lao PDR, and Pakistan) warrant greater attention to strengthening sustainability-enabling factors. The country- and factor-specific ratings presented in Table 13.2 suggest focused areas for attention in these and other countries. Given that DRC was also at the bottom of the data maturity ratings, it may require special attention or greater resources than other PMA2020 and Track20 program countries. However, an important point to consider about these proof-of-concept rating scales (both for data maturity, as discussed in Chapter Twelve, and for sustainability

**Table 13.6**
**Ratings of Sustainability-Enabling Factors, by Respondent Group and Country**

| Respondent | Burkina Faso | Côte d'Ivoire | DRC | Ethiopia | Ghana | India | Indonesia | Kenya | Lao PDR | Niger | Nigeria | Pakistan | Tanzania | Uganda | Zimbabwe | Overall Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PMA2020 | | 4.3 | 3.1 | 4 | 4.3 | 4.6 | 3.9 | 4.1 | | 3.8 | 3.8 | | | 4.5 | | 4.0 |
| Track20 | 4 | 4.1 | 3.3 | | | 3.7 | 3.6 | 3.7 | 3.5 | 3.4 | 3.5 | 3.5 | 3.3 | 4.6 | 3.9 | 3.7 |
| Government/ NGO | 3.3 | 3.5 | 2.8 | 3.5 | 3.9 | 3.7 | 3.6 | 4 | 3.3 | 3.4 | 3.9 | 3.2 | 3.8 | 3.7 | 3.9 | 3.6 |
| Country average | 3.7 | 4.0 | 3.1 | 3.8 | 4.1 | 4.0 | 3.7 | 3.9 | 3.4 | 3.5 | 3.7 | 3.4 | 3.6 | 4.3 | 3.9 | 3.8 |

NOTE: Factors were rated on a scale from 1 to 5, where 1 (red) is the lowest and 5 (green) is the highest.

here) is that comparing country to country may not be as fruitful as using these frameworks as assessment tools to be used longitudinally, over time, *within* a country, once a baseline has been established. Particularly because both PMA2020 and Track20 launched at different times in different countries, and because these ratings were specifically asking about the sustainability domains *in relation to* the two programs, this variation in how well established the programs are must be taken into account when interpreting these ratings, both within countries and across countries.

## Summary

The qualitative interview data and quantitative data from the sustainability matrixes provide complementary views on enablers of and barriers to PMA2020 and Track20 sustainability in these countries. While the quantitative ratings point to specific areas of achievement and concern in specific countries, the interview data add richness and important detail. For example, the ratings indicate only a mid-range level of analytic capabilities (Figure 14.1), and the interviews suggest that most stakeholders would like to see capacity-building efforts to create a larger number of personnel with M&E skills; the ratings indicate substantial leadership and other stakeholder buy-in for both programs, and the interview data provide examples from specific countries; the ratings suggest relatively little progress toward an institutionalized data culture, and the interview data point to specific opportunities to help improve this.

*Financial sustainability* (framed as co-financing) is of critical concern to the Gates Foundation, grantee organizations, and program countries and is particularly salient for PMA2020 because of the higher costs for PMA2020 surveys compared to Track20 programming. There is currently very little domestic resourcing of these programs, with the notable exception of in-kind resourcing—e.g., paid government personnel who serve as Track20 M&E officers and university personnel who serve as PMA2020 principal investigators. *Technical sustainability* is a concern, with ratings pointing to the need to ensure hardware and software maintenance and interview data consistently pointing to the need for more trained M&E personnel.

The most substantial feedback concerned *operational sustainability* and *data culture.* The strongest sustainability-enabling factors in the former were leadership buy-in and cultural acceptability, and the weakest were satisfaction of policymakers' needs, accountability in using data to inform policy, engagement of local communities and civil society, and use of local expertise. Nearly all sustainability-enabling factors related to data culture were rated as relatively strong, with the notable exception of institutionalization of data use. Interview data signaled that some key stakeholders do not understand the data or know how to use them effectively. This complementary information points to opportunities for specific action as PMA2020 and Track20 move into their next grant cycle. Moreover, the data point to countries that are doing well with regard to these factors (e.g., India, Uganda) and those that are not (e.g., DRC, Tanzania); they also identify specific barriers in a given country, again pointing to areas for focused attention.

One theme that emerged clearly from these analyses was the need to plan for sustainability and take deliberate actions to help enable it. This will include explicit attention to all four dimensions of sustainability—financial sustainability, technical sustainability, operational sustainability, and data culture. The findings from these analyses serve as the basis for some of the recommendations presented in Chapter Fourteen.

CHAPTER FOURTEEN

# Conclusions and Recommendations

## Overview

This evaluation included interviews with more than 200 stakeholders in the United States and 15 program countries, with qualitative analyses from those interviews and quantitative analyses related to PMA2020 statistical properties and the data maturity and sustainability ratings. In this chapter, we first highlight the key themes that emerged from our analyses. We then offer recommendations, both overarching and specific to each program, to address the important opportunities we identified.

## Conclusions Underpinning RAND's Recommendations

### Modifying PMA2020's Design to Enhance Usability by Decisionmakers

As described in Chapter Seven, our analyses identified a number of important statistical findings about the PMA2020 survey data. Focusing on five PMA2020 countries with different populations and demographics, as well as multiple rounds of PMA2020 surveys, we analyzed existing PMA2020 data to inform potential improvements to the survey design.

The populations surveyed by PMA2020 differed in many cases from the populations of another similarly focused survey, the DHS, but across the rounds of the PMA2020 survey, the population's known characteristics mostly changed only in ways that would be expected over time. We examined the effective sample size, which takes into account the correlation within clusters in the survey sample. In many cases, design effect was high, which reduced effective sample size. We also looked at variance within each cluster and identified opportunities for optimizing survey design by reducing the number of respondents in clusters with low variance. Finally, by examining round-to-round differences for key indicators, we found that the six-month frequency of PMA2020 surveys adds little value, as indicators in most countries did not change this quickly. Our findings inform recommendations about potentially spacing out the household PMA2020 surveys or asking certain questions less frequently.

These findings also inform our recommendations for other possible survey approaches; without further data, the statistical implications of new designs cannot fully be known, but within the framework of PMA2020 and taking into account stakeholder views as well, we offer design choices that could enhance PMA2020's ability to produce data that will be most useful to decisionmakers.

### Strengthening Country Ownership

Feedback from some U.S-based stakeholders and especially from those in the 15 program countries suggested some degree of country ownership of the data generated by PMA2020 and appreciation for the responsiveness of the Baltimore team to meet evolving country needs (e.g., in Ethiopia). But the preponderance of feedback suggested that countries have unmet needs that could be addressed by a strategic reorientation of PMA2020 toward tailoring data collection to each country's needs—i.e., a transition from a mostly donor-driven agenda toward an owner-driven agenda (Moore et al., 2012). Echoing many similar comments throughout this report, a stakeholder in Nigeria noted, "I want a situation where I don't present my data. The commissioner for health should present my data because it belongs to him. It belongs to the state. That's what I'm working on." Again reflecting a key theme that emerged from our interviews, a stakeholder in Tanzania summarized country ownership nicely:

> We need to find ways of making sure this becomes sustainable and probably develop some mechanisms of transition towards taking the responsibility, more responsibility being taken by the government than by the donors or the project. And then, probably phasing out in a smooth and [sustainable] manner, such that when the program or project ends, this team is totally integrated into the government program.

The program could be reconfigured into a more responsive mode, sacrificing the rigid standardization of some survey elements across countries. This could entail, for example, tailoring the sampling frame, geographic granularity, survey frequency, and survey content (family planning questions, additional modules) to meet countries' needs, within the bounds of available funding. Countries would have to weigh trade-offs and decide how to best allocate program resources to meet their needs. One could envisage, for example, that countries might wish to rotate annual or semiannual subnational surveys across different jurisdictions and leave national surveys to the DHS.

Effective partnerships and connections to relevant government organizations are another key element in country ownership. In-country stakeholders consistently cited government entities (e.g., MOH, statistics office, other ministries [such as finance and planning], subnational governments, schools of public health), bilateral (e.g., USAID, DfID) and multilateral (e.g., UNFPA, WHO) organizations, NGOs, civil society, and multistakeholder coordinating groups as necessary partners in the dissemination and use of data. Especially important is a better connection between the typically university-based PMA2020 team and government.

Stakeholders offered different ideas and examples for how to achieve more country ownership: explicit endorsement of PMA2020 by the MOH or statistics bureau (e.g., mentioned by a stakeholder in Uganda), placing personnel from such an agency onto the PMA2020 team, training government statistical personnel (e.g., mentioned in Ghana), or placing the PMA2020 team within an appropriate government agency. A Gates Foundation staff member noted that PMA2020 should be more thoughtful about partnerships and the implications of locating country staff within universities or "appended somewhere" outside of governments. The program can learn a lot from the DHS model—the DHS is vested with bureaus of statistics at various points, and the U.S. implementing partner is a technical adviser to such bureaus.

Finally, country ownership will not inherently evolve organically. Program and country partners should plan for it, whether from the outset (e.g., initial PMA2020 co-financing required in Côte d'Ivoire, stakeholder suggestion for an initial memorandum of agreement to clarify expectations toward this end) or planning explicitly for transition to full country ownership in those countries that do not currently have such agreements.

### Promoting Data Use

The original objectives of both PMA2020 and Track20 and comments from stakeholders in the United States and across the 15 program countries explicitly called for the effective use of data. As described in Chapter Eleven, different stakeholders have different ideas about what *data use* means—whether for research and publication purposes, for advocacy, or for programmatic purposes (including, but not limited to, governmental, NGO, and donor program planning, resource allocation, program implementation, and monitoring).

There was strong consensus among the various stakeholders interviewed for this evaluation that the use of PMA2020 data and Track20 estimates for planning, resourcing, and managing family planning programs is considered a higher priority than use for research purposes. Most stakeholders opined that both programs should focus more systematically and specifically on data use; this gap was especially apparent for PMA2020. For example, PMA2020 External Consultative Group interviewees reflected views that sound more academic than focused on country needs (e.g., strict focus on family planning, national-level estimates clearly prioritized over subnational estimates, postponement of capacity-building and sustainability until later, narrow concept of data use as pristine data sets for sophisticated analysts for global consumption rather than country program management). These dominating perspectives flavor their views on PMA2020 goals, objectives, and expectations. However, views of the External Consultative Group evolved by 2017 to reflect greater focus on countries as the main PMA2020 audience (and less exclusive focus on global FP2020). This is both a healthy evolution and also probably a sustainability enabler.

USAID's Data for Decision Making initiative is an informative model for the future directions of PMA2020 and Track20. The Centers for Disease Control and Prevention (CDC) implemented the project from 1991 through 1996 under an interagency funding agreement with USAID. The program aimed to help public health authorities in participating countries use data effectively in public health practice. Specifically, its goals were to

> (a) strengthen the capacity of decision makers to identify data needs for solving problems and to interpret and use data appropriately for public health decisions; (b) enhance the capacity of technical advisors to provide valid, essential, and timely data to decision makers clearly and effectively; and (c) strengthen health information systems (HISs) to facilitate the collection, analysis, reporting, presentation, and use of data at local, district, regional, and national levels (Pappaioanou et al., 2003).

The Data for Decision Making program provided on-the-job training for midlevel government personnel responsible for public health policy and programs, aiming for data-based solutions to identified problems. CDC implemented the program in four countries with decentralizing health systems that required increased data capacity at subnational levels: Bolivia, Cameroon, Mexico, and the Philippines. The data-driven and capacity-building focus in the

context of meeting subnational public health data needs in the 1990s is relevant to PMA2020 and Track20 meeting family planning data needs today.

### Enabling Mature, Sustainable Data Systems

The PMA2020 and Track20 programs have aimed specifically for high-quality data, but they have addressed less clearly—and to differing degrees—data system maturity and sustainability in program countries.

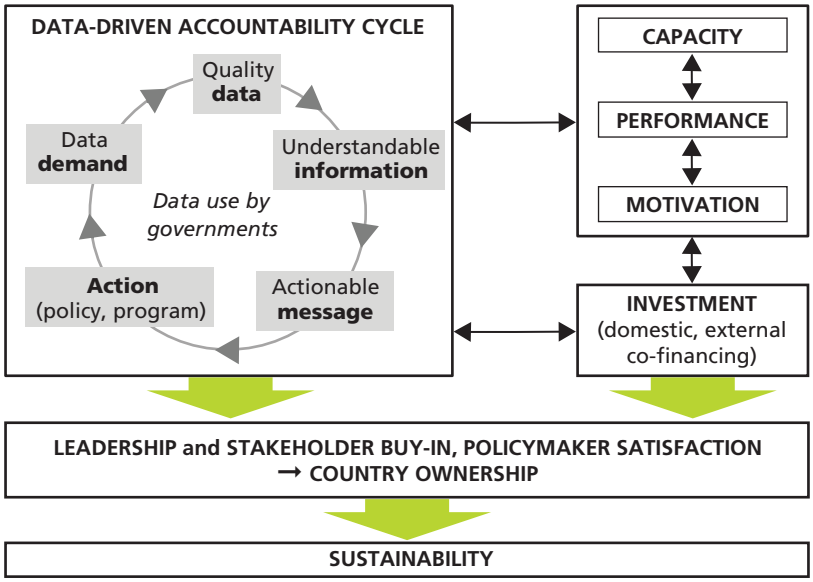
Despite the wide range of qualitative and quantitative responses to questions of data maturity, a few overarching factors have crystallized out of the analyses: First, collecting reliable data on service statistics is an Achilles heel of both programs and nearly every country. This fundamental issue demonstrates how intertwined the data maturity of PMA2020 and Track20 are with existing data systems, such as DHIS2. Second, communication of results to a range of stakeholders could be improved, especially in terms of highlighting why these programs should be considered as essential to the health and well-being of each country's population. Last, role clarity, particularly for Track20-affiliated staff, would greatly enhance the effectiveness of Track20 training, program visibility, and use of FPET and Service Statistics to Estimated Modern Use tools. Ultimately, planning for and measuring data maturity at the country level through frequent self-assessments (e.g., using the RAND data maturity framework), while giving attention to how data maturity ratings change across time and differ across stakeholder groups, would help engender in-country discussions of data maturity domains, in turn promoting ownership and directing program leaders to actionable tasks.

Countries and funding partners alike are interested in the sustainability of their investments in terms of funding, personnel, and operations. However, past decades of development assistance have considered sustainability mainly retrospectively—looking back and hoping that efforts have become sustainable. Based on insights from previous work (Moore et al., 2012; Young et al., 2014), the RAND team is strongly convinced that data maturity and sustainability should be explicit aims to pursue proactively and measure along the way. As an interviewee in Ethiopia noted, "I think ownership starts from the beginning, having a common goal." Or, as an NGO representative in Nigeria also noted, "[sustainability will depend on] a deliberate effort from both Track20 and PMA2020 right at the beginning to define what are those key sustainability instruments that they will put in there. It can come in the form of having simplified tools and aligning those tools with existing systems."

Mature, sustainable data systems are a backbone of health system development, efficiency, and effectiveness. The Gates Foundation brings a strong and unique comparative advantage to supporting data systems related to any of its program areas, including family planning. Perspectives from RAND's stakeholder interviews bore out several key concepts related to data maturity and sustainability.

Figure 14.1 presents a modified version of a figure first presented in Moore et al. (2012), based on what the RAND team learned from the present evaluation. Sustainability is grounded in a data-driven accountability cycle in which high-quality *data* are transformed into understandable *information* and packaged into actionable *messages* about what evidence-based *actions* to take, which then generates and eventually institutionalizes data *demand*. When this concept is applied specifically to data systems, it encompasses both data maturity and sustainability of data systems. Dissemination of data-driven actionable information and the actions taken will enhance data capacity, program performance, and personnel motivation. Collectively, these will signal accountability and responsibility and will attract investment. All these, in turn,

**Figure 14.1**
**Data-Driven Accountability Cycle as a Foundation for Data Maturity and Sustainability**




SOURCE: Adapted from Figure 5 in Moore et al., 2012.
RAND *RR2112-14.1*

enhance leadership and stakeholder buy-in, satisfaction of policymakers' data needs, and country ownership. All these factors enable and culminate in sustainability.

The Monitoring and Evaluation to Assess and Use Results (MEASURE) Evaluation program supported by USAID and implemented by a consortium of U.S. institutional partners led by the Carolina Population Center at the University of North Carolina—Chapel Hill is another example of how capacity-building around data demand and data use can be accomplished and is quite relevant for the Gates Foundation's consideration of future directions for PMA2020 and Track20 (MEASURE Evaluation, 2017). The MEASURE Evaluation, in its fourth phase during 2017, focused on increasing data demand and use during its second phase from 2003 through 2008. The MEASURE Evaluation's data demand and use strategy begins with an assessment that helps stakeholders, policymakers, and M&E practitioners identify where interventions are needed to increase demand and use. Once specific needs are identified, the program employs core tools to stimulate data demand, to build capacity to use data, and, thereby, to enhance data-informed decisions in health. However, despite over 20 years of support for health information systems, the sustainability of its investment is unknown.

### Institutionalizing a Strong Data Culture

A strong data culture embodies an evolutionary process of normative change involving data personnel, processes, and systems to achieve mature data systems: sustainable data capacity and effective data use by country stakeholders at all levels (from local to state/provincial and national), particularly for policy/program planning and performance monitoring. The stages of behavior change originally espoused by Prochaska and DiClemente are relevant to such efforts: raising awareness (of PMA2020 and Track20), enhancing understanding of data (e.g., by adding interpretative notes and actionable messages to all forms of data presentation),

intending action (leadership buy-in for surveys and other data that meet country needs), initial action (Gates Foundation–supported data programs), and sustained action (country-owned data programs and systems) (Prochaska and DiClemente, 1983). All countries have their own data culture, which is not inherently limited to family planning. This evaluation limits discussion primarily to family planning, but strengthening data cultures is potentially also relevant across multiple Gates Foundation programs: seeking and operationalizing opportunities to enhance data maturity and sustainability of data systems.

Feedback from the 15 program countries indicated a broad need for further capacity-building. This includes not only ensuring that the small handful of PMA2020 and Track20 country staff members have the needed skills, but also developing a larger cadre of trained M&E personnel to be placed at relevant subnational levels. Nearly 40 years ago, there was a similar need to build the world's cadre of epidemiologists. In response, CDC developed the Field Epidemiology Training Program (FETP), an on-the-job training program modeled after CDC's domestic Epidemic Intelligence Service training program. Since 1980, FETP programming has reached more than 70 countries around the world (CDC, Division of Global Health Protection, 2017). CDC provides a permanent on-site CDC epidemiologist to a country for three to five years to work alongside a designated country director to train a small handful (e.g., six to ten) new trainees per year over the course of a two-year apprenticeship training period. During their training, FETP participants function as government employees (which most already are); with mentorship, they evaluate public health surveillance information, investigate disease outbreaks, and carry out planned epidemiologic studies. As in the United States, the growing number of FETP graduates has increased the number of well-trained epidemiologists who populate health departments from the national to the local level. Such a capacity-development program strengthens the country's infrastructure, capabilities, and effectiveness. The RAND research team proposes a similar model for building the cadre of M&E professionals in PMA2020 and Track20 program countries, in response to the clear need expressed by country stakeholders.

## Recommendations

### Overarching Recommendations

#### *1. Promote country-driven agendas.*

Orienting both PMA2020 and Track20 toward country-driven (in other words, "owner-driven") agendas will help to **strengthen country ownership** of the processes that these programs are aiming to establish, including data collection, data use, and family planning provision. Country ownership entails active engagement with key stakeholders from national and local governments; bilateral, multilateral, and nongovernmental organizations; civil society; and multistakeholder working groups. Such groups (e.g., as in Pakistan and other countries) are often overseen by government, which enhances government ownership of multistakeholder programming. Country ownership also entails **planning for transition** to full government ownership/institutionalization of data systems (e.g., through mutually agreed upon expectations and exit strategy for donors).

***2. Intensify focus on data use.***

The RAND team's second overarching recommendation stems from one of the clearest messages that emerged from stakeholder interviews: the need to strengthen data use for decisionmaking. Building on the achievements of the first few years of implementation, which focused on generation of high-quality data and estimates, both programs should now incorporate a stronger—and explicit—focus on data use. As noted throughout this report, this challenge appears to be more salient for PMA2020 than for Track20, but there are indications of room for improvement in data use for decisionmaking at all levels across both programs.

We do not consider the first few years of PMA2020 and Track20 implementation as a failure in achievement; indeed, it is very reasonable to have used the first several years of both programs to establish, test, and garner buy-in for (a) PMA2020's innovative methods for data collection and processing, its data platform, and its rapid-turnaround data reports and (b) Track20's tools and M&E officer model. Nonetheless, it is time for PMA2020 in particular to now pivot more explicitly to better meeting country data needs—even at the expense of directly comparable, nationally representative survey data across countries—to both generate and facilitate the use of data needed by country decisionmakers.

Facilitation of data use must be more than a passive process; data should not just be made accessible in the hope that someone will use them. Drawing from the data-driven accountability cycle in Figure 14.1, high-quality *data* must be actively and explicitly transformed into understandable *information* (i.e., interpreted) from which actionable *messages* are created to help decisionmakers know what *actions* they might take on the basis of the data. Very specifically, **the RAND team recommends that all data presentations, including graphics, tables, or narrative reports, include an interpretive statement (translating data to information) and an actionable message (even if a set of alternatives to consider) for data users**, whether for advocacy or decisionmaking purposes.

***3. Plan for and measure data maturity.***

Mature data systems require **planning and explicit effort**. As such, our third overarching recommendation is that PMA2020 and Track20 should systematically plan for countries' data maturity in their own plans and activities. Moreover, the one-time snapshot of data maturity in the 15 program countries is just that—a proof of concept and, in some ways, a baseline assessment. The data maturity framework used in this evaluation provided an effective means for countries to assess their progress toward building an ecosystem that promotes organizational readiness, high-quality data systems, data institutionalization, data governance, and, ultimately, data use. In addition, the framework revealed key areas to prioritize as countries implement PMA2020 and Track20 in the future.

The RAND team recommends that countries use this framework, specifically the data maturity ratings, to **perform periodic self-assessments**. This will allow countries to track progress over time and across various stakeholder groups. It will also help to identify priority areas for focused program attention, especially to address the foundational elements that are early in the data maturity cycle (e.g., quality data collection, management, and analysis), which subsequently support the other building blocks of data maturity (e.g., data institutionalization, data use).

It is important to emphasize that actions and perceived progress toward data maturity will not be uniform across program countries. For some countries, achieving the basic infrastructure needed to ensure mobile data collection would be the highest priority, while other

countries may consider getting key decisionmakers to trust in PMA2020 data and Track20 estimates as a key next step in improving their data maturity.

***4. Plan for and measure institutionalization/sustainability.***

Similar to data maturity, institutionalization/sustainability of PMA2020 and Track20 data systems also requires **planning, explicit effort, and tracking progress over time**. The RAND team's fourth overarching recommendation is to be proactive rather than reactive in planning for sustainability and measuring sustainability-enabling factors related to PMA2020 and Track20 and for the programs to implement activities accordingly. Sustainability of data systems in program countries is independent of whether they reach their FP2020 goals. We heard more than once, particularly from the Gates Institute and PMA2020 External Consultative Group members, that it was "too early" to discuss sustainability during the intensive ramp-up periods in each of the PMA2020 countries. However, this approach has led to the realization that it is more challenging to retroactively structure co-financing arrangements and other specific sustainability enablers than to develop them from the outset, as was done with considerable intentionality and effort in the co-financing of PMA2020 in Côte d'Ivoire.

The RAND team's sustainability framework encompasses four principal dimensions of sustainability—financial, technical, operational, and data culture—and the specific sustainability-enabling factors associated with each. It also serves as a tool for **periodically taking stock of progress** in each country and addressing bottlenecks along the pathway toward sustainable data systems. One specific suggestion arising from the stakeholder interviews was for the programs to establish formal agreements with the governments where they operate around ownership and sustainability (i.e., not only memoranda of understanding around each program's operational elements). These agreements should articulate mutual expectations, the roles and responsibilities of key parties, and the ultimate aim for countries to take greater ownership of the systems and processes as one of the key factors that will enable their sustainability.

***5. Institutionalize data capacity development.***

As a final overarching recommendation and a unifying effort to strengthen and institutionalize data capacity across all data system functions, from collection to management, analysis, presentation, and use, the RAND team proposes establishing a **Data for Action Training Activity for Family Planning program (DATA-FP program)**. Such a program would be modeled after CDC's FETP, as described earlier in this chapter. The program could also incorporate elements of the new (as of July 2017) Gates Foundation Operational Research program implemented in partnership with the London School of Hygiene and Tropical Medicine, which provides grants to countries to build their analytic capabilities through guided operational research—i.e., investing in on-the-job data capacity-building. It could also potentially synergize efforts under way through the Gates Foundation's integrated service delivery program and other data infrastructure–oriented programs.

The DATA-FP program would build capacity in terms of both skills and number of personnel—an **ever-larger cadre of well-trained M&E personnel** who can collectively collect, manage, analyze, interpret, disseminate, and facilitate use of family planning data. This is just one of several vertically oriented Gates Foundation programs to which the proposed DATA program could be applied—others could include DATA-NUT (nutrition), DATA-WASH (water, sanitation, and hygiene), and DATA-MNCH (maternal, newborn, and child health). Use across more program areas would unify data-oriented programming in countries

and would be a unique opportunity for the Gates Foundation to systematically promote mature and sustainable data systems, a data culture, and enough staff/capacity to make it all possible. Such programs could also build on the training provided through policy workshops by the Institute for Health Metrics and Evaluation (an independent global health research center at the University of Washington).

In the same vein, we highlight the USAID-supported MEASURE Evaluation program led by the Carolina Population Center as a useful model (and potential vehicle) to consider when exploring consortium arrangements of capacity-building around data use for decisionmaking.

### Summary of Overarching Recommendations

These recommendations for further empowering countries, strengthening data use, planning for and measuring data maturity and sustainability, and institutionalizing data capacity-building are captured in a logic model reflecting a broad overview for Gates Foundation family planning programs (Figure 14.2). The Gates Foundation will likely continue to support FP2020 data-related efforts. As reflected in the figure, such support might include more subnational PMA2020 surveys and/or non–family planning survey modules. The RAND team hopes that support might also include the proposed new DATA program (DATA-FP is specifically recommended here), for which the Gates Foundation is uniquely well positioned by virtue of its vision to empower developing country partners, strong data system orientation, and broad range of development programs.

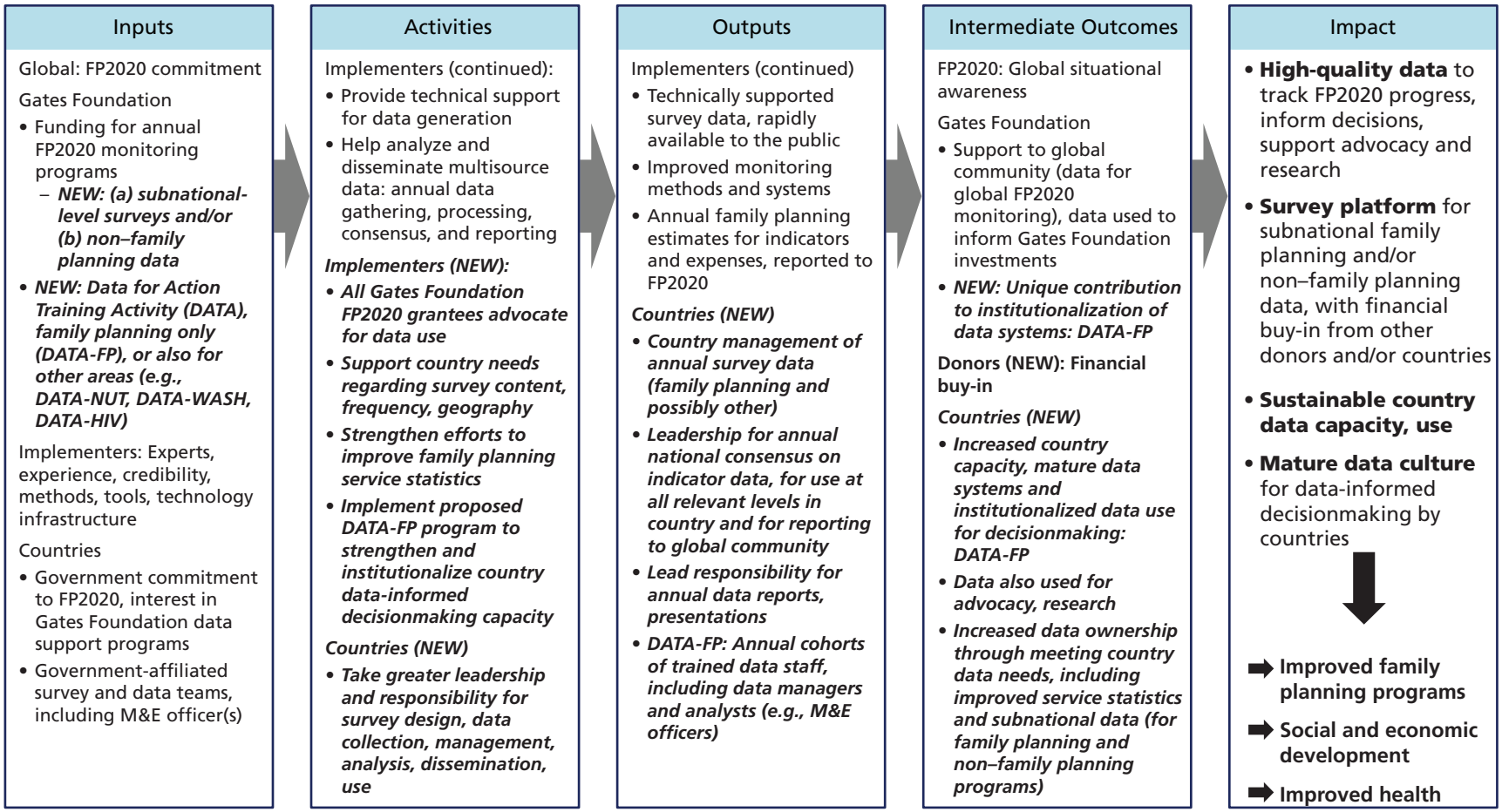
Government buy-in and expert technical assistance will continue to be required inputs for any future PMA2020 and Track20 scenario. Activities would include continuation of data generation and analysis, strengthening dissemination efforts, and broader and more intensive focus on data use to meet country needs, improvement in family planning service statistics, and potentially also the proposed DATA-FP program. Countries would assume greater leadership and management responsibilities for the data processes, including data presentations and issuance of reports. Intermediate outcomes would lead to impacts that include high-quality data that are used to inform program planning, resource allocation, and management; a survey platform that meets a broader range of government and other stakeholders' needs; institutionalized and sustainable country data capacity and use; and a sustainable data culture.

### Recommendations for PMA2020

#### *1. Reorient and operationalize the program to align with its objectives.*

This evaluation revealed that different stakeholders, from the Gates Foundation to the grantee organization (the Gates Institute) to the in-country PMA2020 staff, have varying expectations of the program's goals and objectives, which lead to wide range of opinions on future directions and potential opportunities. These divergent viewpoints lead to the RAND team's first recommendation for PMA2020, which is to critically reexamine, revise, and reach consensus around the vision, goals, objectives, and activities for PMA2020 and to follow through with activities to operationalize them. Specifically, the Gates Foundation should articulate the directions it would like PMA2020 to take, considering both its original and modified objectives, and reach agreement with the implementers. Such goals and objectives should include, but are not limited to, generating high-quality family planning data, building capacity among a cadre of resident enumerators, contributing to data maturity in program countries (including

**Figure 14.2**
**Logic Model for Potential Future for Gates Foundation Family Planning Data Programs**




RAND *RR2112-14.2*

facilitation of data use), planning for a sustainable future, and planning for a transition to increased government/country ownership.

This clarification of PMA2020's goals and objectives will be necessary to further define what successful data generation *and use* will look like, how to facilitate them, and how to measure them over time. It will clarify to what extent the PMA2020 grantee organization, the Gates Institute, is expected to both generate and facilitate the use of PMA2020 data. The RAND team's recommendation is that **the PMA2020 implementing organization should indeed play a key role in disseminating its data, interpreting them with and for decisionmakers (by providing interpretive statements and actionable messages on all of its graphics), and sharing the data with key advocates,** such as Track20 M&E officers and others who can share them with governmental decisionmakers at all levels. The PMA2020 implementer should place more emphasis on dissemination meetings (which have been occurring in some countries to a greater extent than in others), and these meetings could also function as training workshops on how to interpret and use the PMA2020 data.

A strategic reexamination of goals and potential reorientation will also enable the Gates Foundation and its implementing partners to make challenging decisions about future directions, all of which will entail trade-offs. For instance, if the Gates Foundation decides to shift its emphasis, as strongly advocated for by in-country stakeholders, toward subnational estimates over national estimates, then the survey can target certain regions or states of interest. But it will be required to stop marketing PMA2020 as producing national estimates to track progress toward a global goal (FP2020). Furthermore, PMA2020's more desirable niche could be as an important complement to service statistics rather than trying to be a complement to, or simply a more frequent, DHS.

Alignment among different stakeholders will also help to address the tension between "vertical" family planning data generation and "horizontal" data system strengthening—ideally moving toward the idea of a more "diagonal" approach that accomplishes both, as coined by Julio Frenk (Frenk, 2010). A vertical approach, which is not necessarily bad, fills important gaps in the availability and quality of family planning data and could be the focus in countries without other robust family planning data sources or could supplement data collected in other surveys, such as the DHS and MICS. These vertical data efforts could address several of the stated needs of decisionmakers, such as more qualitative information, more data on quality of services, additional populations that include adolescents and men, and more.

A vertical approach to family planning data also allows the survey to take better advantage of surveying both households and service delivery points. A related recommendation is to consider collecting service delivery point data more frequently than household data (to better inform program management, commodity tracking, and capacity-building for data use at local levels) but, during rounds in which both are sampled, to link service delivery point data more systematically to household data. On the other hand, the horizontal approach seeks to better integrate family planning data generation and use into a larger effort of building data maturity within countries across all health (and related) sectors.

**The proposed DATA-FP program described above incorporates PMA2020 and Track20 and marries the virtues of horizontal data system strengthening with vertical family planning programming—i.e., a diagonal approach.** The capacity-building feature increases the cadre of well-trained M&E personnel, while PMA2020 data generation remains vertically oriented around family planning narrowly or any other programs associated with non-family planning PMA2020 modules. By working within such a diagonal approach,

PMA2020 could position itself to help build M&E capacity for family planning while also continuing to support the generation and use of high-quality data for family planning and potentially also other programs.

***2. Engage key partners in active data dissemination.***

Secondly, the RAND team recommends that PMA2020 actively engage key partners to strengthen its data dissemination efforts and further raise its visibility within program countries. This is increasingly happening in some countries (e.g., Nigeria, Uganda) with each additional round of data collection; however, efforts have been uneven across the 11 PMA2020 countries. There are certain advantages to having the PMA2020 principal investigators primarily located within academia (e.g., their deep knowledge of survey design and their technical expertise). However, the RAND team heard loud and clear that the university-based location of the principal investigator contributes to the perception among U.S.-based stakeholders and in-country decisionmakers that the PMA2020 survey is "boutique," "academic," and "by and for researchers." Therefore, in order to overcome this perception (as well as to address the evolving realization expressed to the evaluation team that it is not enough to produce high-quality data, make them available, and hope that they are being taken up and used), the onus is on the PMA2020 in-country partners and their teams to **deliberately build strong and lasting connections to governmental decisionmakers,** including MOHs, bureaus of statistics, and other actual and potential users of its data (e.g., advocacy organizations, donors, NGOs, other researchers, and other programs, such as Advance Family Planning [an advocacy initiative] and the Challenge Initiative [an urban family planning program], both of which are supported by the Gates Foundation and led by separate teams within the Gates Institute from PMA2020). In order to effectively disseminate its data and contribute to their ultimate use for decisionmakers, PMA2020 must determine who its target audience is (as alluded to in the first recommendation) and build relationships (as well as adapt its "product" to the audience's different data needs, to the extent possible). Another potential way to address this barrier to data use is to relocate the PMA2020 survey within bureaus of statistics, which (in countries where they are highly functional and sufficiently resourced) could be in charge of its day-to-day operation, with any additional critical technical support needed to be provided by the university-based current PMA2020 principal investigators and their highly skilled teams.

Our evaluation revealed that in-country PMA2020 staff were not always invited to the national data consensus meetings organized by the Track20 M&E officer. Both programs need to be more proactive about connecting around their common mission and finding efficient ways to communicate with a shared voice with key decisionmakers.

***3. Enhance PMA2020's survey design.***

There are many possible options to enhance PMA2020's design, all representing different trade-offs and priorities. It is not possible to address the data needs of all stakeholders simultaneously. The RAND research team has integrated the family planning data needs, as articulated by a variety of stakeholders, the U.S.-based and in-country respondents' perspectives on PMA2020's design, and our analyses of the statistical properties of the PMA2020 survey, to develop the following recommendations.

- **Collect family planning data from households less frequently than every six months in the first two years** in a new PMA2020 country. Stakeholders felt that collecting data

on many of the family planning indicators every six months at first was too frequent, not only because of the resources and energy expended, but also because of how long it can take to implement interventions to measurably impact those indicators, as well as how long it takes for these indicators to change even when interventions are tried. Our statistical analyses supported this view, showing that family planning indicators, such as, but not limited to, mCPR, did not change significantly during six-month intervals, but they did change from year to year.

- The RAND team supports pilot efforts to **implement both a targeted panel survey** (as is being implemented in the maternal, newborn, and child health arena) **and a cross-sectional survey.** The panel design, while resource-intensive, would provide valuable information, including allowing PMA2020 to delve more deeply into changes in contraceptive use over time (i.e., discontinuation and method switching), which was a priority for many stakeholders. The increased resources needed for a panel design could be offset by decreasing the overall frequency of household data collection to annually, as recommended above. The other consideration is that panel data should be analyzed separately from cross-sectional data, which decreases the sample size and, therefore, the statistical power of each. Important considerations for the PMA2020 program are the trade-offs between resampling from the same enumeration area for a certain number of rounds (currently, four), then refreshing the sample to move to an adjacent enumeration area. Throughout this evaluation, we have discussed the drawbacks to sampling within the same enumeration area—namely, the high proportion of repeated households and the potential for the Hawthorne effect. However, these disadvantages must be weighed against the alternative, which is to sample a new enumeration area each time, requiring either newly trained resident enumerators or the same enumerators traveling further distances and new mapping and listing work to be done ahead of each round. The RAND team recommends further exploring this issue empirically as more rounds of data are collected in each country to determine the optimal balance between the pros and cons of refreshing the sample in a set of new, adjacent enumeration areas after a certain number of PMA2020 rounds.
- **Use the resources freed up from decreasing the frequency of household data collection to intensify explorations of other innovative but resource-intensive experiments,** in addition to the panel survey design described previously. These experiments could be performed in the interim (i.e., between annual surveys) or on an annual basis, but the key would be to leverage cost savings from decreasing the frequency of the PMA2020 survey. Options include the following:
  - Field a pilot test of conducting more rapid data collection using mobile phones.
  - Perform focused data collection around a particular intervention (i.e., for program evaluation purposes) or in a particular subnational jurisdiction.
  - Make the survey more flexible: Certain questions or modules could be asked of a smaller or larger portion of the overall sample. While complex, it would be possible to use technology to "push" the tailored survey versions to particular populations. For instance, to reduce burden on the resident enumerator and respondents, data from previous rounds of collection on the homogeneity versus heterogeneity of particular enumeration areas could guide decisions about how many households to select in an enumeration area based on design effects from the previous round, as well as how many respondents need to be asked certain questions.

  - Test new modules on a rotating basis (given the strong support for adding questions related to family planning, as well as modules beyond the core family planning topics).
  - Incorporate in-depth, open-ended questions to explore the "why" behind the data (such as around method switching and discontinuation).
- **Enhance efforts to produce robust subnational estimates** to meet data needs of decisionmakers. Because decisionmakers overwhelmingly expressed a need for family planning data on the local level (e.g., the regional, provincial, state, or even district level), this could be an important niche for PMA2020 going forward. As subnational estimates are challenging to obtain, given the sample sizes needed, the RAND team also recommends **implementing an optimal design**, using variance from the full population of the enumeration area, to estimate the number of people to survey in each cluster (if this information is not available, the optimal design could be implemented in Round 2 using the variance from Round 1 in the enumerator area). As described in Chapters Seven and Eight, this type of design accounts for the variability of a sample and permits a reduction in sample size in enumeration areas for which responses have small variability. It has the potential to improve the precision of key indicators, increase the efficiency of the PMA2020 survey, and thus reduce the necessary sample size for robust subnational estimates (or allow resources to be used to sample in more subnational jurisdictions if PMA2020 wanted to enhance the national representativeness of indicator estimates).
- **Use the service delivery point surveys to their full potential.** There are several options to consider under this recommendation.
  - Some stakeholders suggested conducting the service delivery point surveys more frequently than household surveys, given the need for more frequent facility data on items of interest that are actionable *and* change from week to week and month to month (e.g., to track, and thus address, commodity stock-outs in a more timely fashion). The RAND team recommends considering a frequency somewhere between the DHS's continuous service provision assessment model under way in Senegal and the annual service delivery point survey in countries where PMA2020 has been established beyond the first two years. This approach would allow PMA2020 data collected from service delivery points to be used to validate (i.e., triangulate) service statistics, as well as provide richer, more detailed supply-side data than can be provided by service statistics (including provider bias, more information on outreach efforts, skill level around insertion of IUDs, and approach to counseling, among other topics).
  - Another recommendation is to improve the linkage between household and service delivery point data when they are collected contemporaneously, ideally by advancing ongoing efforts within PMA2020 to use geocoded locations of both households and facilities if privacy concerns within countries can be sufficiently addressed.
  - Finally, the RAND team supports the pilot efforts under way for PMA2020 to collaborate with Track20 on **improving the quality of service statistics**. Service delivery point surveys could be more effectively used to complement (in terms of more granular data around service provision) and validate the information in the HMISs, which stakeholders noted were often of questionable quality and rigor.
- **Strengthen the technical support for data set users** and **ensure a streamlined and user-friendly data download process.** This recommendation stems from U.S.-based stakeholder views on the ease, or lack thereof, of use of PMA2020 data for research and analyses. Researchers in particular will likely benefit from efforts through a current Gates

Foundation contract to the University of Minnesota to standardize how different variables are collected from country to country (including how the variables are named) and how they are collected from round to round. This standardization will enhance comparability of data across countries and to the DHS's indicators and variables. For those who do not need or want to work with the full, raw data sets, **PMA2020 should continue to support its valuable DataLab tool.**

- Finally, from a statistical standpoint, the RAND team identified several additional recommendations:
  - Ideally, for each round of data collection, the **sample size should be estimated and updated based on the most recent mCPR** from the prior round of PMA2020 data rather than from the most recent DHS.
  - **Continue to use mCPR as the foundational indicator on which to base sample size estimates.** The statistical analyses demonstrated that the margin of error remained less than the goal of no more than 3 percent, but this margin of error should continue to be tracked (as well as that for other indicators) to ensure that it remains low.
  - Because some stakeholders expressed concern that the mCPR indicator is a "blunt tool" for understanding family planning trends, the RAND team recommends focusing on **collecting more granular information about respondents to further understand mCPR** in different subpopulations of women, such as the mCPR in women who are trying to space births or become pregnant, the mCPR in women who do not desire more children, the mCPR in married versus unmarried women, and the mCPR in women who are not sexually active or who are anovulatory in the postpartum period.
  - As we found that design effect was high, which reduced effective sample size, we recommend consideration of the option to **increase the number of enumeration areas but sample fewer respondents from each to maximize efficiency.** This option has practical challenges, so our next recommendation may be more feasible, given available resources.
  - We recommend implementing an **optimal cluster design** that takes into account the homogeneity or heterogeneity of the main indicator of interest within clusters (because variance differs by indicator). This design varies the sample size within each cluster and can improve the precision of an estimate, given the same sample size. Alternatively, it can reduce the sample size needed to obtain the same precision.

***4. Broaden the PMA2020 platform and seek cost-efficiency to attract co-financing.***

Considerable discussion has been under way among the Gates Foundation, the Gates Institute, the PMA2020 External Consultative Group, and PMA2020 countries regarding the scope of PMA2020 survey content, specifically related to **non–family planning modules**. The case in favor of supporting a broader range in survey content is that such an approach would meet a broader range of stakeholder needs (e.g., country governments, the Gates Foundation, other donors) and increase the possibility of co-financing, which is an important lens through which the Gates Foundation views sustainability. The case against such expansion involves the extra costs to program and translate new questions for collection through PMA2020's mobile survey platform and a possible dilution of the family planning focus of PMA2020 as originally conceived and implemented.

Review of documentation from the External Consultative Group meetings and the RAND team's interviews with staff from the Gates Foundation, Gates Institute, and program

countries suggest that **the time is ripe to carefully broaden the PMA2020 survey platform**. Such a pivot must be **coupled with efforts to seek co-financing, reduce survey costs, and make the core family planning modules much leaner—that is, to only ask questions that are truly actionable and necessary, not just "nice to know."** More than one country stakeholder noted the potential promise of approaching the Global Financing Facility in each country to inquire whether the PMA2020 platform could help meet its annual data needs.

The efficiency of an investment is based on a ratio between cost (numerator) and outputs (denominator). People tend to think of reducing cost as the main way to improve efficiency. Indeed, reducing the absolute costs of PMA2020 surveys would improve efficiency (e.g., reducing labor costs by conducting more surveys by phone rather than in person [or at least using this method for the subset of households re-interviewed for quality assurance purposes] or shifting labor costs more toward in-country personnel with less dependence on U.S. personnel). Conceptually, another way to improve efficiency would be to expand the outputs reflected in the denominator. As a stylized example, one can achieve greater PMA2020 efficiency for a given cost if the outputs extended beyond generating high-quality family planning data to also include **effective data use** in countries to help **improve family planning services** and, in turn, **improve health outcomes**. Simplistically, this is a comparison between

1. cost / high-quality data
2. cost / (high-quality data + effective data use + improved family planning services + improved health outcomes).

The RAND team's overarching recommendations related to country ownership and data use in particular would create this conceptually larger denominator and thereby greater efficiency of the costs invested. Another future direction under discussion by the Gates Foundation, the Gates Institute, and others is the use of PMA2020 as a methodological test bed or innovation lab. It is unclear whether this would be financially sustainable if all PMA2020 programming were to serve such a purpose, but it might be reasonable to entertain in one or a small handful of countries, perhaps selectively in those with relatively mature data systems.

### Recommendations for Track20

#### *1. Intensify the focus on contributing actionable data at different levels in increasingly decentralized health systems.*

In considering future directions for this program, the RAND team's evaluation identified a desire on the part of Gates Foundation staff, grantee organizations, and country stakeholders to continue to shift the program's emphasis toward country-level activities. As part of this reorientation toward country-specific work, we recommend intensifying the focus on two of the specific examples provided by in-country stakeholders: the new **FP Goals model** and efforts to improve the quality of **costed implementation plans**, both of which will yield the most dividends for countries as they seek to improve the use of family planning data for decisionmaking. With regard to the costed implementation plans, while a source of pride within countries that are developing them, there is much room to improve in terms of their quality, their feasibility, their degree of country ownership rather than representing donor priorities, and their relationship to family planning data. Furthermore, in order to be responsive to the family planning needs of in-country stakeholders, the RAND team recommends, based on strong consensus around the need for more subnational family planning indicator estimates,

placing more M&E officers at subnational-level positions, in addition to having at least one M&E officer in the national MOH or a comparable agency.

***2. Improve and expand the use of FPET.***

FPET is a fundamental innovation of the Track20 program and is one of the keys to its success, in addition to embedding M&E officers in MOHs. We have four recommendations to further strengthen this critical tool and methodology in the service of enhancing family planning data for decisionmaking: (1) Because the quality of the Bayesian model depends on the data sources that are entered into the tool, the RAND team supports Track20's continued work with countries to **improve their routine service statistics**; (2) in order to be responsive to in-country stakeholders who noted that the theoretical basis for FPET, as well as some of its more advanced features, were a "black box" even to statisticians, we recommend **continuing to further the understanding of FPET methodology and theoretical basis for both the users of the tool and the consumers of the data** (the decisionmakers); (3) in order to be responsive to in-country stakeholders and to reduce reliance on a single person or a few individuals who are trained in the FPET methodology, we recommend **expanding the number of people trained in FPET's modeling techniques**; and (4) we recommend expanding FPET's capacity to **produce robust subnational estimates** to meet the needs of decisionmakers.

***3. Optimize the M&E officer model.***

One of Track20's strengths is its deliberate placement of M&E officers within existing governmental structures (primarily MOHs), which facilitates their access to key decisionmakers in the family planning arena. The RAND team's third recommendation is to build on prior successes and reflect on prior challenges with M&E officer models in various countries in order to optimize several dimensions of the unique model. First, with regard to financing/salary support for these positions, determine which of the several M&E officer financing models (i.e., no, partial, or full salary support by Track20) has yielded the best results and replicate it across other countries, to the extent possible. From the results of our evaluation, it seems that having Avenir finance at least a portion of the M&E officer's salary has been beneficial and may be the preferred model going forward. This will reinforce the strong partnership among government, M&E officers, and Track20. As mentioned previously, the RAND team recommends **placing additional M&E officers at subnational levels,** as is occurring in some countries (Nigeria and India, for instance), and to continue testing which governmental office (MOH, statistics office), and at what level, yields the best result. The specific solution will not be the same in every country, but the success of the M&E officers depends on some of these structural factors. Furthermore, our interviews captured varying perspectives on whether existing M&E officers should be solely focused on family planning or should cover other areas of health and development to be less "stovepiped." M&E officers themselves felt that they could not devote enough time and energy to family planning when asked to perform multiple duties. Therefore, each country must find an optimal balance between increasing the number of M&E officers and diversifying the portfolio of existing officers.

Regardless of how much of their time is devoted to family planning activities in particular, our interviews in Track20 countries suggested that M&E officers need adequate support for their expected duties, including reasonable expectations about their workload, turnaround time for responding to data requests, capacity-building and professional development, supervision and guidance commensurate with their experience and skill level, and assistance with

connections to policymakers (as well as mechanisms to overcome the local politics and bureaucracies that impede the effective dissemination of their work). With these supports in place, M&E officers can strengthen their role as critical ambassadors of and interpreters of data for decisionmakers, as well as serve as bridges to other health and development priorities within MOHs and beyond. Track20 should continue to support M&E officers to communicate effectively with stakeholders, including ensuring access to them and the tools they need to interpret analytic results and provide actionable messages in relation to those results. Building on existing partnerships with all relevant partners (including ensuring that PMA2020 staff are routinely involved in national data consensus meetings and even more frequent meetings in the interim) and with data users will help institutionalize the Track20 program within the larger data architecture of program countries.

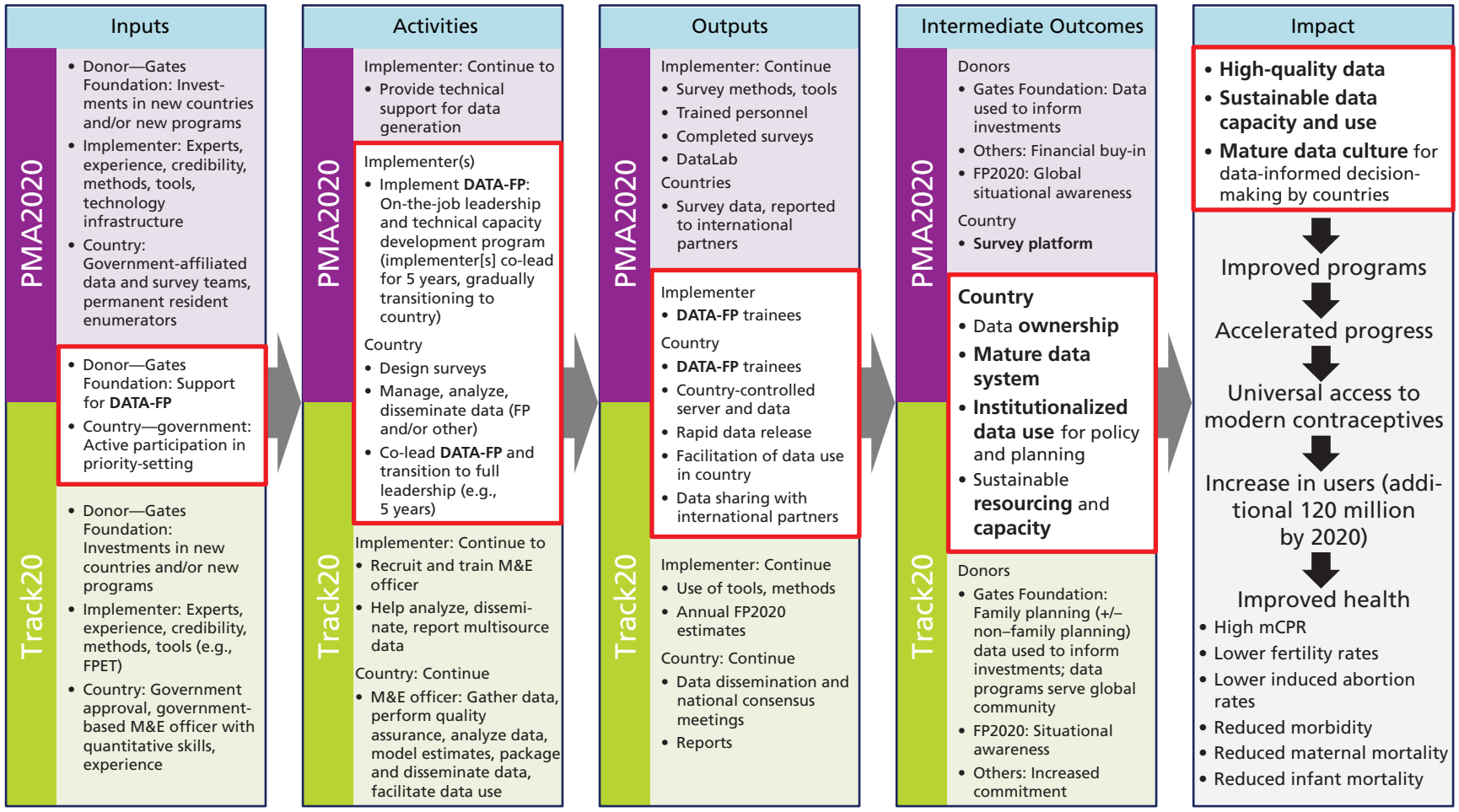
## Summary

These recommendations for engaging key partners in data dissemination, improving service statistics, optimizing the design of PMA2020 and the understanding and use of FPET, broadening the PMA2020 platform, and optimizing the M&E officer model compactly reflect recommended actions to build on achievements to date and reorient the programs to achieve even more in the future. A final logic model (Figure 14.3) extends the broad concepts reflected in Figure 14.2 and places them in the context of the two programs, also showing how the proposed new DATA-FP program (outlined in red) aligns well with those concepts.

PMA2020 and Track20 reflect the vision of Gates Foundation leadership, the commitment of participating countries, and the technical expertise of the grantee organizations. Their achievements to date are notable. The RAND research team identified a number of important opportunities for the future orientation and activities of these programs, based on analyses of the rich feedback from more than 200 stakeholders in the United States and the 15 program countries included in this evaluation. The recommendations address these opportunities and are feasible in the near term.

The RAND team is grateful for the opportunity to carry out this evaluation and appreciates the many individuals who informed the findings reported here. We wish the programs continued success in their efforts to provide high-quality family planning data to subnational, national, and global stakeholders while also empowering and enabling countries to mature and sustain their data systems.

**Figure 14.3**
**Logic Model for Potential Future Directions for PMA2020 and Track20 with Proposed DATA-FP**




RAND *RR2112-14.3*

APPENDIX A

# Details of PMA2020 Sampling Procedures

PMA2020 surveys are based on a probabilistic multistage stratified cluster design, in which the first stage is selection of enumeration areas (or clusters) and then, within those, selection of a number of households, where the surveyed population lives. This design is very similar to the design of the DHS. This design increases the ability to compare findings across surveys over time and allows validation of PMA2020 results through comparisons to DHS data, which is often seen as the gold standard. This ability to validate PMA2020's results has been identified as an advantage for promoting PMA2020 data use (though, at the same time, the fact that the measures it produces are very similar raises questions about the purpose of the survey and how it makes its own contribution). Stratification, in PMA2020's case by enumerator area, allows for reduction of "normal sampling variation and produces a sample that is more likely to look like the total population" (Fowler, 2013), as well as use of resident enumerators with specific locations.

Enumeration areas are determined by a statistical agency in each country, usually the National Statistical Office or its equivalent. The Gates Institute team and in-country PMA2020 staff together determine the sampling frame and then ask the statistical agency to draw the sample and determine the selection probabilities based on the design using "probability proportional to size (PPS) within the designated strata" (Zimmerman, 2017), a method of selecting enumeration areas at random while ensuring that larger enumeration areas have a higher probability of being selected (Lavrakas, 2008).

In addition to the list of enumeration areas, information provided includes geographic information, urban/rural designation, enumeration area selection probability, and all other relevant information used to calculate the enumeration area selection probability, which means the population/number of households in the enumeration area, the number of enumeration areas sampled per stratum, and the county or regional selection probability, where relevant (Zimmerman, 2017). This is the first stage of the two-stage selection process.

The same household sampling frame is used for the first four rounds and then is refreshed. Reportedly, the refreshed samples are drawn from neighboring enumeration areas (in order to retain the trained local resident enumerators) rather than entirely *de novo* sampling from the entire population of the country or subnational jurisdiction. At the same time that the initial enumeration areas are identified, PMA2020 now asks the government to provide "all geographically contiguous enumeration areas that share the same urban/rural designation as the original index enumeration area" to be used for alternate enumeration areas, so that enumeration area resampling can take place. While PMA2020 now requests lists of enumeration areas as well as lists of contiguous enumeration areas at the beginning of Round 1 in each country, they did not receive these contiguous lists in all countries when the project was kick-

ing off. For that reason, they have had to ask for a second set of enumeration areas to be generated later on. As a result of a miscommunication with the government in Ghana, the 100 new enumeration areas that were selected were not contiguous to the original ones. Instead, Ghana selected totally new enumeration areas. There are no repeat enumeration areas that were selected because of the large number, but this does not fit the contiguous model.

In DRC, this model of "refreshing" the original sample by drawing from contiguous enumeration areas was also not followed. There, because of the very large size of the enumeration areas, PMA2020 decided not to refresh the enumeration areas for Round 5. In most countries, by Round 4, 15 to 20 percent of respondents have already been interviewed; however, because of the large size of DRC, the rate of already-surveyed women is only 1 to 2 percent, leading to the decision to stay in the same enumeration areas for Round 5.

The plan to switch to new, contiguous enumeration areas after four rounds of data collection has been implemented in Ethiopia, Uganda, and Kenya for Round 5 (and in Ghana to new, noncontiguous enumeration areas). Five rounds of data collection are complete in all of these countries, but only for DRC are the Round 5 data cleaned and publicly available.

Once the selected enumeration areas are determined (i.e., the first stage of the two-stage cluster design), a list of households in each enumeration area is generated. PMA2020 then randomly selects a number of households in the enumeration area. The number chosen is typically 35 but ranges from 33 to 44. This decision is based on the target of matching the rate of 30 women per enumeration area in the DHS and then taking into account an estimated 10-percent nonresponse rate (customized by known prior response rates in each country) and the known population distribution in the enumeration area (some locations have more than one woman per household; others, such as Kenya and Uganda, have fewer, particularly in urban areas). Three attempts at the household level and three attempts at the female level are made. Because replacement of respondents can introduce bias, replacements are not sought, so PMA2020 applies sample-size inflation instead. All women in a household, whether permanent members or temporary members, are surveyed, if possible, though only the permanent members are included in the analyses of the female questionnaire (Zimmerman, 2017).

In addition to selecting households, service delivery points are identified for surveys. All households, health service delivery points, and key landmarks in each enumeration area are listed and mapped by the resident enumerator to create a sampling frame for the second stage of sampling. Up to three private service delivery points within each enumeration area boundary are randomly selected from the listings, and three public health service delivery points—at the community (CHPS), primary health care (health center), and primary health care referral (district hospital for that enumeration area) levels—are selected. For some enumeration areas, fewer than three service delivery points may be present.

A number of assumptions are made during this selection process. First of all, sample size is based on the mCPR rate, using the rate from the most recent DHS survey, which, depending on the country, may have been more or less recent. Similarly, the assumption is that the design effect has not changed since the last DHS. The design effect reflects the homogeneity within the selected population. The sampling approach also assumes a high response rate, about 95 percent for households and 95 percent for female interviews, for an overall rate of about 90 percent. So far, response rates in PMA2020 have been high, though not universally so (response rates for Ghana in 2013, for example, were 87.1 percent for households and 90.3 percent for women) (Performance Monitoring and Accountability 2020 Project and Kwame Nkrumah University of Science and Technology, 2013), whereas rates increase significantly by the fourth round:

97.9 percent of households agreed to an interview, with 98.0 percent of eligible women agreeing (PMA2020, 2017f). In Round 4 of the survey in Uganda, 94.5 percent of households responded, and 95 percent of eligible women then consented to the interview (PMA2020, 2017g). Overall response rates range from 88 to 97 percent.

Lastly, the design assumes that the number of women per household remains what it was under the last DHS (Zimmerman, 2017). PMA2020 chooses to err on the side of a larger number of enumeration areas, "budget permitting," to be more conservative (Zimmerman, 2017).

All of these assumptions are reasonable, given information that is available, but they depend, of course, on the amount of time since the last DHS (and the accuracy of the data from the local government).

## Sample Size

For each round, sample size is determined based on mCPR; according to PMA2020, the goal is to include a sample size that allows a national estimate of mCPR with a target margin of error: "The survey aims to include a sample size that would allow analysts to obtain a national estimate for all indicators, including calculating the modern contraceptive prevalence rate (mCPR) with a margin of error of ±3 percentage points" (PMA2020, 2017h).

Other sources have suggested that the margin of error aimed for is plus or minus 1.5 percentage points, but the Gates Institute team noted that the original plan was for 3 percentage points (PMA2020, 2014). Once the Gates Institute has begun working with countries to identify their specific needs and their budget, the margin of error estimates can become more precise. For example, in Ethiopia, the Gates Institute started working on regional estimates at a margin of error of 3 percentage points, so when the PMA2020 survey expanded to produce national estimates, that number became much more precise.

In order to minimize the design effect, sample size is determined by changing the number of enumeration areas and not the number of households per enumeration area.

The original mCPR on which the sample size is based, as noted previously, comes from DHS data. While PMA2020 has considered changing sample size in the second round to reflect mCPR as calculated in its own first round, it is not yet doing so because this change would likely require an increased sample size (when mCPR has grown), which has budget considerations.

### Sample Size Concerns

Enumeration areas vary by size but also by homogeneity on the factors that are being measured by the survey; failing to account for these factors when selecting sample sizes within each enumeration area can also introduce bias and decrease efficiency.

PMA2020 has chosen the number of enumeration areas in a country and the number of participating households (and women within households) within an enumeration area based on expectations of the acceptable margin of error of mCPR. As above, the calculation to determine sample size is based on achieving a margin of error for mCPR of no more than 3 percentage points in most cases, which is assumed to be small enough to allow policymakers to make decisions within such margin. When representativeness at any subnational level (i.e., regional level) is desired, the national-level margin of error assumed can no longer be expected. Subnationally, one can only expect to make inference with larger margins of error. The larger the national

sample size, the larger many subnational samples will be and the more confident inferences within regions can be. If information about the subnational inferences that are being made are known, the minimum margin of error at such level can be estimated in advance.

On the other hand, even though PMA2020 is designed to make inferences primarily when it comes to mCPR, decisionmakers will also wish to make inferences about other indicators, such as the use of traditional contraceptive methods. As such, confidence in inferences of secondary indicators requires the knowledge of the margin of error of such an indicator in the sample. In our analytic results section, we provide calculations of precision for a number of other indicators of interest besides the mCPR. When designing a survey such as PMA2020 around a single indicator, the accuracy of other indicator estimates can be compromised. Additionally, the mCPR around which the sample-size estimates are calculated typically comes from the most recent DHS in the respective country, sometimes several years out of date. Ideally, for each round of data collection, the estimation of sample size could be based on updated mCPR estimates coming from the most recent PMA2020 data. In cases in which mCPR has increased toward 50 percent (which is the optimal case when it comes to power calculation), one would have to increase sample size to produce estimates with the same precision, which has budget implications. Nonetheless, it is important to keep in mind that changes in mCPR are likely to be relatively small from round to round, so the need to increase sample size can also be minimal.

As Fowler (2013) argued, "It is unusual for a researcher to be able to specify a desired level of precision in more than the most general way. It is only the exception, rather than the common situation, when a specific acceptable margin for error can be specified in advance." To attain a wide variety of use, sample size should focus on the "minimum sample sizes that can be tolerated for the smallest subgroups of importance" (Fowler, 2013) while keeping cost in mind. The challenge for PMA2020 is determining the smallest subgroups for which it wants to have valid estimates: Specific regions? Specific demographic groups? Or something else?

### Potential Sources of Error

Any survey potentially has multiple sources of error. These include problems with sample selection (which includes random sampling error or bias in the sampling process), question design, interviewer effects, and nonresponse bias. Additional error can be introduced in the processing and interpreting of data collected, including data missingness. In this report, we focused on sample design.

In general, a specific population is the target of the inference sought, and individuals who are in that population are eligible to be in the sample and constitute the "sample frame" (Fowler, 2013). The quality of any design depends on how representative the actual sample frame is when compared with the population of interest. Enumeration area sampling frames that are based on outdated census data can lead to bias in the sample, especially if there is a possibility of mobility in the population of interest between the time the census was collected and the fielding of the PMA2020 survey. For example, in Nigeria, the census used for the selection of enumeration areas was fielded in the 1980s; if the population distribution at the time when PMA2020 is fielded is different from that in the 1980s in Nigeria, the enumeration area selection based on the 1980s data can lead to sampling error. The validity of PMA2020 as a random sample of women (and men) in a country hinges on the representativeness of the selected enumeration areas. For this reason, we looked closely at sample representativeness.

Furthermore, when the sample of PMA2020 enumeration areas is refreshed, which usually occurs after the fourth round of a PMA2020 survey, in general only enumeration areas adjacent to those initially selected have been considered. After the sample of enumeration areas is selected for the first round of PMA2020, households in any enumeration area other than those selected have no chance of being in the samples for the PMA2020 Round 2, 3, and 4 surveys. Similarly, for Round 5, only households in enumeration areas adjacent to Round 1 enumeration areas have a chance of being selected.

## Quality Assurance

Once PMA2020 data are collected, they undergo a comprehensive process of quality assurance.

First, data managers in the country conduct a series of routine analyses daily or every two days. They generate a report to look for specific errors. For example, the report flags duplicate submissions for the same household. The data managers in the country also check to make sure that a household form is present for every household with a female form, and vice versa. If matching forms are missing, the appropriate resident enumerator can be contacted to inquire about and help resolve the problem. All surveys include GPS coordinates, to ensure that the survey is submitted in proximity to the relevant household. If the location of submission is far from the known house location, there is concern that the survey was submitted inappropriately and that the data are perhaps not accurate. The data managers also analyze nonresponse and "do not know" responses for certain questions. There is concern that there may be interviewers who are just selecting "do not know" rather than asking some of the more sensitive questions, such as "When did you last have sex?" High rates of "do not know" or nonresponse would indicate that the interviewer is either not putting the respondent at ease or is simply not asking the more sensitive question. The local data manager also has a tool that reports how much time is spent on each question, so as to identify whether a resident enumerator perhaps quickly swiped through the survey and answered all the questions without actually conducting an interview. These and other quality control checks are done in-country so that the data manager can quickly identify problems and communicate directly with the supervisor or with the resident enumerators in the field to address any issues. If at any point they identify data that are inaccurate or falsified, they would go back and redo the entire numeration area. That has rarely occurred.

Variables are similar across all survey countries, though the survey is translated into relevant languages and the survey uses country- and culture-appropriate terminology. Choices about methods include only choices relevant for a given country. Other country-specific differences, which are minimal, are listed in Box A.1.

Once the data arrive at the Gates Institute, further quality assurance is conducted. A data manager there runs the same reports that are run in the country but with less frequency. This ensures that the problems have been resolved. The scripts that identify these errors are written and updated at the Gates Institute.

At the end of data collection in a specific country, 10 percent of the sample is identified, and supervisors go back and administer much shorter surveys to verify that the people there are the ones who were surveyed, that the survey was actually done, and that the right women were included. This is a verification exercise. If concerns are raised via this process, a specific household can be reinterviewed, or if problems are broader, data from the whole enumeration

**Box A.1**
**Variables That Differ by Country on PMA2020 Survey, per PMA2020 Data Documentation**

1. Livestock questions: The specific livestock options (cow, rabbit, etc.) vary across countries.
2. School: Education categories for female schooling vary across countries.
3. Family planning provider: Provider of current or most recent method of family planning varies across countries.
4. Roof/wall/floor: Household materials vary across country.
5. Assets: The household assets used to construct wealth scores vary across countries, as do the binary variables that are created from the multi-select asset question.
6. Wealth quintile or tertile: In some countries, wealth quintiles are provided; in others, wealth tertiles. The continuous variable score is included to allow for reconstruction of various wealth categories.

SOURCE: PMA2020, 2017i.

area can be recollected. Every round of PMA2020 undergoes this verification exercise. Rarely has it demonstrated any issues.

Data sets are also updated when there are changes or adjustments to calculated variables or labeling, to add variables to match variables in other data sets, or to fix skip patterns. For example, the survey asks about LAM use, but PMA2020 also calculated an LAM variable to include only those with a birth in the past six months, which is the technical definition for when it would be considered a reliable, modern method. The original variable was restored in a corrected data set so that researchers could re-create the PMA2020 data analysis. These edits increase uniformity and improve the accuracy of the data, but it can be disconcerting for data users to receive an updated data set from 2013 when they have already performed their analyses.

APPENDIX B

# Indicators Collected by PMA2020, FP2020, and DHS

The FP2020 initiative has developed 17 core indicators that can be compared across countries and over time (AvenirHealth, 2017). PMA2020 surveys were designed to support the larger FP2020 effort and collect data on a number of family planning indicators, including measures of contraceptive utilization, demand, choice, quality, and access. Table B.1 compares the FP2020 core indicators that are captured by PMA2020. Table B.2 defines the PMA2020 indicators and how they are calculated, and Table B.3 compares which PMA2020 indicators are also collected in the DHS and MICS and the FP2020 core indicator to which each one corresponds.

The main outcome of interest within FP2020 is the "number of additional women (or their partners) of reproductive age currently using a modern contraceptive method compared to 2012" (AvenirHealth, 2017). To determine this number, the mCPR, which refers to the proportion of all women ages 15–49 who are using (or whose partners are using) a modern method of contraception, is the critical indicator. For this reason, PMA2020's sampling strategy is based on the rate of change in mCPR: Consistent with FP2020's goal of enabling 120 million more women and girls to use contraceptives by 2020 (Family Planning 2020, 2017), the sample is powered, for each round, to detect an annual change in mCPR with less than a 1.5-percentage-point margin of error at the national level, representing 95-percent confidence (PMA2020, 2014).

**Table B.1**
**FP2020 Core Indicators Captured by PMA2020**

| FP2020 Core Indicator | PMA2020 Indicator |
|---|---|
| 1. Number of additional users of modern methods of contraception | |
| 2. Contraceptive prevalence rate, modern methods (mCPR) | Modern contraceptive prevalence rate |
| 3. Percentage of women with an unmet need for modern methods of contraception | Unmet need for spacing and for limiting |
| 4. Percentage of women whose demand is satisfied with a modern method of contraception | |
| 5. Number of unintended pregnancies | |
| 6. Number of unintended pregnancies averted due to modern contraceptive use | |
| 7. Number of unsafe abortions averted due to modern contraceptive use | |
| 8. Number of maternal deaths averted due to modern contraceptive use | |
| 9. Percentage of women using each modern method of contraception | Total contraceptive prevalence rate |
| 10. Percentage of facilities stocked out, by method offered, on the day of assessment | |
| 11a. Percentage of primary service delivery points that have at least 3 modern methods of contraception available on day of assessment | |
| 11b. Percentage of secondary/tertiary service delivery points with at least 5 modern methods of contraception available on day of assessment | |
| 12. Annual expenditure on family planning from government domestic budget* | |
| 13. Couple-years of protection (CYP)* | |
| 14. Method Information Index | Told of other methods and counseled on side effects |
| 15. Percentage of women who were provided with information on family planning during their last contact with a health service provider | |
| 16. Percentage of women who decided to use family planning alone or jointly with their husbands/partners | Method chosen alone or jointly |
| 17. Adolescent birth rate | |

* These indicators do not use PMA2020 data.

**Table B.2**
**Definition of PMA2020 Family Planning Indicators**

| Indicator | How Calculated |
|---|---|
| Contraceptive prevalence rate (CPR) | Proportion of women ages 15–49 who are using (or whose partners are using) any contraceptive method at the time of the survey |
| Modern contraceptive prevalence rate (mCPR) | Proportion of women ages 15–49 who are using (or whose partners are using) a modern method of contraception, which includes hormonal and barrier methods, sterilization, emergency contraception, lactational amenorrhea method (LAM), and the standard days/cycle beads method |
| Long-acting CPR | Proportion of women ages 15–49 using a long-acting contraceptive method, which includes an IUD, implant, and sterilization (male and female) |
| Total number of modern contraceptive users | A count of the number of females ages 15–49 who are current users of modern methods of contraception |
| Unmet need | Percentage of fecund, sexually active women ages 15–49 who are not using contraception and do not wish to become pregnant at all (unmet need for limiting) or within the next two years (unmet need for spacing) |
| Demand satisfied by modern contraception | Percentage of women ages 15–49 who do not want to get pregnant who are using modern contraception |
| Total fertility rate | Number of children who would be born to a woman if she were to pass through her reproductive years bearing children according to the current schedule of age-specific fertility rates (ASFR) |
| Adolescent fertility rate | Age-specific fertility rates for women ages 15–19 |
| Intention to use contraception | Percentage of women not currently using a method of contraception who intend to use a method in the future* |
| Unintended births | Percentage of births in the past 5 years to females ages 15–49 that are reported to be mistimed (wanted later) or unwanted |
| Method mix | Composition of current methods used by women ages 15–49 |
| Method chosen by self or jointly | Percentage of women ages 15–49 currently using a modern contraceptive method, reporting that they decided on method themselves or jointly with a partner or provider |
| Obtained method of choice | Percentage of women ages 15–49 currently using a modern contraceptive method, reporting that they obtained their contraceptive method of choice |
| Method information index | Percentage of recent/current users reporting that they were informed about other methods and side effects and, if informed of side effects, were told what to do* |
| Paid for services | Percentage of women ages 15–49 currently using a modern contraceptive method reporting that they paid for family planning services in the past 12 months |
| Satisfaction with provider | Percentage of women ages 15–49 currently using a modern contraceptive method who would return to their provider and would refer a relative or friend to that provider |
| Received method from public service delivery point | Percentage of women ages 15–49 currently using a modern contraceptive method who obtained their contraceptive method from a public service delivery point |
| Sterilized users told method was permanent | Percentage of sterilized users counseled on method |
| Reasons for non-use | Reasons for non-use of contraceptive methods among married women who express a desire to postpone their next birth by two or more years |

**Table B.2—continued**

| Indicator | How Calculated |
|---|---|
| Age at first marriage | Median age at first marriage for women ages 25–49 |
| Age at first sex | Median age at first sexual intercourse for women ages 15–49 |
| Age at first contraceptive use | Median age at first contraceptive use for female ever users ages 15–49 |
| Number of living children at first contraceptive use | Average number of living children at first contraceptive use among women ages 15–49 who have ever used contraception |
| Women having first birth by age 18 | Percentage of all women ages 18–24 who had their first birth before age 18 |
| Received family planning information from provider | Percentage of women ages 15–49 reporting that they received family planning information from a provider at a facility or in their home in the past 12 months |
| Recent exposure to mass media family planning messages | Percentage of women ages 15–49 reporting exposure to family planning messages via radio, television, or print in the past few months |

SOURCE: PMA2020, 2017c.

* This is an original FP2020 indicator. It can be calculated using PMA2020 data but is not currently presented in PMA2020 publications. Calculations for indicators are made according to these definitions, which were updated in December 2016. Calculations may have changed slightly from those used in earlier PMA2020 publications.

**Table B.3**
**Key Indicators Measured by PMA2020 and Other Selected Surveys**

| PMA2020 Survey | Category | Indicator (PMA2020, 2017b) | Definition (PMA2020, 2017c; PMA2020, 2017d) | DHS | MICS | FP2020 |
|---|---|---|---|---|---|---|
| Individual and household | Utilization | mCPR | Proportion of women ages 15–49 who are using (or whose partners are using) a modern method of contraception, which includes hormonal and barrier methods, sterilization, emergency contraception, lactational amenorrhea method (LAM), and the standard days/cycle beads method | X | X | Indicator 2, used to calculate other indicators, including 1, 4, 6, 7, and 8 |
| | | Traditional contraceptive prevalence rate | | X | X | |
| | | Total contraceptive prevalence rate | Proportion of women ages 15–49 who are using (or whose partners are using) any contraceptive method at the time of the survey | X | X | Indicator 9 |
| | Demand | Unmet need for spacing and for limiting | Percentage of fecund, sexually active women ages 15–49 who are not using contraception and do not wish to become pregnant at all (unmet need for limiting) or within the next two years (unmet need for spacing) | X (unmet need aggregate) | X | Indicator 3 |
| | | Total demand | | X | X | |
| | | Percentage of non-users who intend to adopt a contraceptive method in the future | Percentage of women not currently using a method of contraception who intend to use a method in the future | | | |
| | Choice | Obtained method choice | Percentage of women ages 15–49 currently using a modern contraceptive method, reporting that they obtained their contraceptive method of choice | | | |
| | | Method chosen alone or jointly | Percentage of women ages 15–49 currently using a modern contraceptive method, reporting that they decided on method themselves or jointly with a partner or provider | X | | Indicator 16 |
| | | Told of other methods and counseled on side effects | Percentage of recent/current users reporting that they were informed about other methods and side effects and, if informed of side effects, were told what to do | X | | Indicator 14 |

**Table B.3—continued**

| PMA2020 Survey | Category | Indicator (PMA2020, 2017b) | Definition (PMA2020, 2017c; PMA2020, 2017d) | DHS | MICS | FP2020 |
|---|---|---|---|---|---|---|
| Service delivery point | Quality and access | Offer family planning counseling and services to adolescents | Percentage of health facilities that provide contraceptive counseling, as well as contraceptive provision or prescription, to adolescents | | | |
| | | Client feedback system | | | | |
| | | Percentage of service delivery points offering family planning methods (by type) | Percentage of all health facilities that counsel, provide, or prescribe each of the following contraceptive methods: pills, injectables, IUDs, implants, male condoms, and emergency contraception | | | Indicator 11, stock-outs used for indicator 10 |
| | | Mobile outreach teams worked from facility in the facility | Percentage of health facilities reporting that a mobile team visited to deliver supplementary/additional family planning services in past 6 (or 12) months | | | |

APPENDIX C

# Comparison of Selected Family Planning Data Sources

**Table C.1**
**Comparison of Selected Family Planning Data Sources**

| Characteristic | PMA2020 | DHS[a] | MICS[b] |
|---|---|---|---|
| Year launched | 2013 | 1984 | 1995-96 |
| Sponsoring organization | Gates Foundation | USAID | UNICEF |
| Implementing organization | Johns Hopkins University and country partners | ICF International[c] | Local governments |
| Where implemented | 11 of 69 countries committed to FP2020 | 93 USAID-supported countries[d] | 107 low- and middle-income countries |
| Survey goal | Nationally representative data on key family planning indicators | Nationally representative and internationally comparable data on fertility, family planning, health, and nutrition | Internationally comparable data on women and children and statistical system capacity-building[e] |
| Collection frequency | Twice a year for two years, then annually | Usually every five years | Every five years initially, every three years since 2009 |
| Sampling design | Multistage stratified cluster sampling | Multistage stratified cluster sampling | Multistage stratified cluster sampling |
| Sample size | 3,000–7,000 households | 5,000–30,000 households | 3,000–6,000 households |
| Reporting domains | National and in some cases regional | National and regional | National and regional |
| Primary sampling unit | Enumeration area based on recent census | Enumeration area based on recent census | Enumeration area based on recent census |
| Cluster size | 30–35 households per enumeration area | 20–30 households per enumeration area[f] | One or more enumeration areas, with an average of 100 households, depending on the country |
| Stratification for sample | Urban/rural | Age, residence, states/regions, education, wealth quintile | Residence, age, gender, subnational |
| Sample weights | Household and individual weights; survey dependent | Household and individual weights; survey dependent | Household and individual weights; survey dependent |
| Data collection approach | Smartphone-based | Paper-based initially; collected directly into laptop or on personal digital assistant at selected sites[g] | Paper-based |

**Table C.1—continued**

| Characteristic | PMA2020 | DHS[a] | MICS[b] |
|---|---|---|---|
| Fieldwork period | 6-week collection period | 3- to 6-month collection period | 2- to 6-month collection period |
| Family planning indicators | Utilization, demand, choice, quality, and access | Utilization, demand, choice, and access | Utilization, demand |
| Other topics covered | Water, sanitation, and hygiene; menstrual hygiene management; sustainable development goals; schistosomiasis | Women's health (female genital cutting, empowerment, domestic violence, nutrition, child health (infant and child mortality, anemia, malaria, nutrition), environmental health, HIV/AIDS, tobacco; independent variables (education, household characteristics, wealth); special surveys on HIV, malaria, gender, service provision, and specialized project-specific surveys | Water, sanitation, hygiene, child health (child development, nutrition, protection), HIV and sexual behavior, tobacco and alcohol use, mortality, poverty, victimization, household energy use, literacy and education, mass media and information and communications technologies, social transfers, functioning, subjective well-being, adolescent health, malaria, salt iodization, female genital cutting, and life satisfaction |
| Response rate | 88–97% | 90–99% | 87–99% |

[a] Aliaga and Ren, 2006.

[b] Plowman and Fotso, 2013.

[c] ICF International was founded as the Inner City Fund. The name of the DHS implementing organization has changed over time, but key personnel have remained the same.

[d] DHS Program STATcompiler, undated.

[e] UNICEF, September 2015.

[f] Corsi et al., 2012.

[g] International Union for Scientific Study of Population, 2009.

APPENDIX D

# Additional Background on Logic Models

The idea of a logic model, originally called a "logical framework" or "log frame," was first conceived by Rosenberg et al. (1970) during their evaluation of USAID's project appraisal reporting system (Rosenberg, Posner, and Hanley, 1970). The original framework consisted of inputs, outputs, project purpose, and sector or programming goal. Between 1970 and 1971, USAID applied this framework to evaluate projects for technical assistance in 30 program countries. Subsequently, USAID decided to expand its use to other types of programs. In the mid-1970s, the former Canadian International Development Agency became the first non-U.S. agency to employ this approach for evaluating its programs around the world (Practical Concepts Incorporated, undated). Since then, the use of logic models has expanded within and beyond the international development community. Currently, many organizations use logic models in a more strategic and systematic way (O'Mahony et al., forthcoming).

Logic models are grounded in *theories of change*. A theory of change (TOC) is a depiction of how and why program activities accomplish the desired impacts in a particular context (Center for Theory of Change, undated); a TOC establishes the theory for how and why an initiative works (Weiss, 1995). A TOC begins with identification of the desired impacts and is then developed by "backward mapping" to identify the pre-conditions necessary to achieve these effects (Weiss, 1995). A TOC can be represented in different formats, including figures that have different shapes linked through arrows, lines, or feedback loops that imply cyclical or direct relationships (Bullen, 2013). When a TOC is depicted as a sequential framework that links program inputs and activities to their outputs and outcomes, followed by the desired impact (Gertler et al., 2011), it becomes a logic model or log frame (Bullen, 2013).

A logic model establishes a linear relationship between activities and impacts and the intermediate products and results between them (Bullen, 2013). It shows the trajectory of how *inputs* (resources) (Gertler et al., 2011) support program *activities* that, in turn, produce *outputs* (tangible, short-term products, services, or results) (O'Mahony et al., forthcoming), which result in *intermediate outcomes* (what is achieved using outputs) and, eventually, *long-term impacts* (desired long-term effects) (Haggard and Burnett, 2006).

When designing a logic model, it is important to be aware of external factors that might affect program outcomes (Greenfield, Williams, and Eiseman, 2006). For instance, in a program that provides training to participants, how participants apply the training they receive in practice might be affected by multiple factors outside of the trainers' control. Therefore, logic model developers should take this into consideration and try to anticipate issues that might impede achievement of desired results.

By providing a structure and visual representation lining up operations, short-term products, intermediate-term outcomes, and long-term goals, a logic model is a useful tool for pro-

gram strategic planning, monitoring, and evaluation (Haggard and Burnett, 2006). Moreover, program planners and evaluators can use logic models to assess how activities contribute to or hinder the achievement of a program's objectives and identify key entry points for improving the probability of attaining the desired outcomes (Haggard and Burnett, 2006). While logic models lend themselves to measurement across all elements, from inputs to impacts, we prefer to follow the direction of others in focusing on achievements related to outputs and intermediate outcomes (Young et al., 2014). This approach focuses on what is produced and achieved in the short and medium terms and can be tracked as part of the regular M&E process.

APPENDIX E

# Additional Background on Data Maturity Models

**Table E.1**
**Review of Existing Data Maturity Models**

| Source | Maturity Levels | Domains Measured and Areas Within Each |
|---|---|---|
| University of Chicago Center for Data Science and Policy Maturity Model | Lagging<br>Basic<br>Advanced<br>Leading | Data and technical readiness scorecard: Collection (relevance and sufficiency, quality—among others), storage (accessibility, storage, integration), other (privacy, documentation)<br>Organizational readiness scorecard: Buy-in of (a) staff, (b) data collector, (c) leadership, (d) intervenor, (e) funder; people resource; data use policy |
| Open Data Maturity Model | Initial<br>Repeatable<br>Defined<br>Managed<br>Optimizing | Data management processes: Data release process, standards development and adoption—among others<br>Knowledge and skills: Open data expertise, knowledge management<br>Customer support and engagement: Engagement process, open data documentation, community norms<br>Investment and financial performance: Financial oversight, valuation process—among others<br>Strategic oversight: Open data strategy, asset catalogue |
| Master Data Management Maturity Model (MD3M) | Initial<br>Repeatable<br>Defined process<br>Managed, measurable<br>Optimized | Data model: Definition of master data, master data model, data landscape<br>Data quality: Assessment of data quality, impact on business, awareness of quality gaps, improvement<br>Usage and ownership: Data usage, data ownership, data access<br>Maintenance: Storage, data life cycle |
| Capability Maturity Model Integrated (CMMI), Carnegie Mellon University | Performed<br>Managed<br>Defined<br>Measured<br>Optimized | Data management strategy: Communications, data management function, business case, funding<br>Data governance: Governance management<br>Data quality: Data quality strategy, data quality assessment, data cleaning<br>Data operations: Data requirements definition, data life-cycle management, provider management<br>Platform and architecture: Architecture approach and standards, data management platform, data integration<br>Supporting processes: Measurement and analysis, process management—among others |
| Informs Analytic Maturity Model | Beginning (levels 1–3)<br>Developing (levels 4–7)<br>Advanced (levels 8–10) | Organization: People, leadership impact, measures, processes<br>Analytics capability: Analytic framework, roles and skills, analytic services, analytic processes<br>Data and infrastructure:<br>Health, access, traceability, analytics architecture |

**Table E.1—continued**

| Source | Maturity Levels | Domains Measured and Areas Within Each |
|---|---|---|
| TDWI Analytics Maturity Model and Assessment Tool | Nascent<br>Pre-adoption<br>Early adoption<br>Corporate adoption<br>Mature/visionary | Organization: Leadership/culture, strategy, skills<br>Infrastructure: Development, technologies, architecture<br>Data management: Kinds of data, integration, data quality<br>Analytics: Scope, culture, delivery methods<br>Governance: Policies, structure, compliance, stewardship, security<br>Demographics |
| IBM | Ad hoc<br>Foundational<br>Competitive<br>Differentiating<br>Breakaway | Business strategy<br>Information<br>Analytics<br>Cultural and operational execution<br>Governance<br>Architecture |

APPENDIX F

# Additional Background on Sustainability Enablers

Sustainability is a nearly universal priority in the development community, yet it is a concept that is often difficult to pin down precisely. The United Nations (UN) defines sustainability as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (UN, 2016). For the purposes of this evaluation, RAND researchers drew from a published definition of sustainability of health information systems:

> Sustainability refers to the tendency of the [health information] system to endure over time and space and is directly concerned with the system becoming institutionalized in the workings of the health department. Institutionalization can thus be described as the process by which [health information systems] can be sustained over time (Kimaro and Nhampossa, 2007).

Sustainability in the strengthening of country-level health systems and data management depends on the degree to which local systems and communities internalize and incorporate related principles into their everyday operations.

## Key Elements of Sustainability

For a system to be truly sustainable, the international development literature consistently emphasizes that it must be *locally driven, owned, and operated*. Pretty (1995, as cited in Mog, 2004) asserts that "the most useful way to conceptualize sustainable development is as a process of social change that tackles underlying structural problems and is rooted in learning, continual innovation and 'perpetual novelty'" (Mog, 2004). Because of the growing appreciation of the importance of sustainability, the UN has shifted its focus from the Millennium Development Goals (MDGs) to the Sustainable Development Goals (SDGs). The SDGs are consistent with the multiple UN conferences focused on this topic since the Rio Summit in 1992 (e.g., also in 2002, 2012, 2016); they were adopted in 2015 and comprise 17 goals to help eradicate poverty and inequality globally and 169 targets to help measure progress toward those goals. They continue the mission of the MDGs to alleviate poverty but adopt a broader social and environmental perspective that focuses on education, health, social protection, and economic growth. The focus on environmentally conscious development solutions recognizes the need to address these problems in an integrated and holistic fashion. In the specific realm of Health, Goal 3 of the SDGs aims to "ensure healthy lives and promote well-being at all ages" with a specific target to "ensure universal access to sexual and reproductive health care services (including family planning or FP)." Recent evidence has highlighted how investing in family

planning can be considered a "best buy" for development (Starbird, Norton, and Marcus, 2016). However, what remains elusive is how governments and planners can track their continued progress to achieve lasting gains as part of a long-term socioeconomic development strategy linked to SDG achievement. By promoting family planning services and programs in an integrated, community-owned, and inclusive manner, the UN hopes that member countries can more readily and quickly achieve their stated goals.

## Sustainability and Family Planning

In the joint areas of sustainability and successful family planning programming, the literature emphasizes many similar key points. *Strong internal (government) and external (donor) support* are critical, as is *collaboration* in program design and implementation across key actors, including all levels of government within a country (from national to local), donors, and civil society (Jamison et al., 2006). Also essential is consistent, predictable *funding support* alongside programmatic approaches that *build the capability* of the health system in terms of human and other resource capacity as well as operational function. From a local perspective, programs and the tools they provide must be *simple yet effective*, while ensuring regular participation of households and community members in all phases of the project. Related technological tools used to advance family planning programming, data tracking, and usage need to be *user-friendly* if they are to be adopted and used over a sustained period.

## Sustainability and Data Use

A key element of successful family planning programming is the collection and use of appropriate data to track progress against goals and objectives. Sustainability of data usage follows several core principles. Consistent with the broader literature, *buy-in and consultation with the local community* remains essential so that data systems can be used effectively and maintained (Moore, 2007). Building on this initial buy-in, regular *training* for and *funding* of data systems is also a determinant of lasting impact and success (Iwelunmor et al., 2016). If local users do not possess the technical knowledge to independently update and maintain an information system, the chances for sustainability are greatly reduced.

Beyond the mere usage of data and systems, establishing a *norm* of *data-informed decisionmaking* refers to the interactive processes that consider data during program planning, monitoring, review, and improvement; advocacy; and policy development and review (Nutley, 2012). Such a process ensures sustainability of data usage as it seeks to create a *data culture* that supports data-informed decisionmaking and involves sound norms of primary and secondary data collection and usage among all levels of data generators, manipulators, and *end users* (Bond, Bertrand, and Mera, 1994). Creating a mature data culture in the health sector requires senior-level MOH involvement in creating a *supportive environment*, strong support for and training in useful data manipulation tools at senior levels of the MOH, *institutionalized training programs* for key professional staff, and the development of tools in a *collaborative and participatory* fashion (Bond, Bertrand, and Mera, 1994). In addition, it is pertinent to address potential barriers that affect the sustainable use of data, such as disincentives stemming from lack of *decisionmaking authority*; failure of health systems to *meet the needs of policymakers* in

terms of content, format, and timeliness of data; lack of *trust in the accuracy of the data*; and fear of *political, economic, and social consequences* (Pappaioanou et al., 2003).

In the realm of sustainability of data usage in global health, relevant case studies from Mozambique, Tanzania, and South Africa highlighted several key relationships and conclusions relevant for sustainability of effective HMIS platforms. In these instances, the *relationships* between the respective MOHs and the software development agency, as well as between the MOH and donors, were deemed critical. Based on such findings, all users need to be able to effectively use the system in question, but the design should be a *flexible* one that evolves over time to meet the specific needs of the country in question—ideally, the HMIS should be *simple, of high quality, responsive, scalable, adaptable, stable, and robust* (Kimaro and Nhampossa, 2007). Furthermore, local networks should conform to national standards or templates in terms of data usage and input and should expand on existing frameworks where necessary (Jacucci, Shaw, and Braa, 2006).

APPENDIX G

# Contextual Information on the 15 Countries Evaluated

Understanding the specific country contexts, including the survey data landscape for family planning in PMA2020 and Track20 program countries, provides important context for these programs. Table G.1 shows the demographic, economic, and social profile of the 15 countries we evaluated. The Human Development Index is a composite index ranging from 0 to 1 that measures the average achievements in a country related to life expectancy, access to knowledge, and standard of living, with lower numbers indicating less development.

**Table G.1**
**Demographic, Economic and Social Profile of the 15 Countries**

| Country | Mid-2017 Population | World Bank Economic Classification | Human Development Index (rank, among 188 countries) |
|---|---|---|---|
| Burkina Faso | 19.6 million | Low income | 0.402 (#185) |
| Côte d'Ivoire | 24.4 million | Lower middle income | 0.474 (#171) |
| DRC | 81.5 million | Low income | 0.435 (#176) |
| Ethiopia | 105.0 million | Low income | 0.448 (#174) |
| Ghana | 28.8 million | Lower middle income | 0.579 (#139) |
| India | 1.21 billion (2011) | Lower middle income | 0.624 (#131) |
| Indonesia | 257.6 million (2015) | Lower middle income | 0.689 (#113) |
| Kenya | 49.7 million | Lower middle income | 0.548 (#148) |
| Lao PDR | 6.8 million (2015) | Lower middle income | 0.586 (#137) |
| Niger | 20.6 million | Low income | 0.353 (#187) |
| Nigeria | 190.9 million | Lower middle income | 0.527 (#152) |
| Pakistan | 203.4 million (2016) | Lower middle income | 0.550 (#147) |
| Tanzania | 57.5 million | Low income | 0.531 (#151) |
| Uganda | 42.8 million | Low income | 0.493 (#163) |
| Zimbabwe | 16.6 million | Low income | 0.516 (#154) |

SOURCES: Population Reference Bureau (Kaneda and Dupuis, 2017); World Bank, 2017; United Nations Development Programme, 2016.

**Table G.2**
**PMA2020, DHS, and MICS Surveys and DHIS2 Use in the 15 Countries**

| Country | PMA2020 Years | DHS Years | MICS Years | DHIS2 User: Year Started |
|---|---|---|---|---|
| Burkina Faso | 2014, 2015, 2016 | 1993, 1998–1999, 2003, 2010, 2014* | 1996, 2006 | 2013 |
| Côte d'Ivoire | 2017 | 1994, 1998–1999, 2005*, 2011–2012, 2016* | 1996, 2000, 2006, 2016 | 2016 (pilot phase) |
| DRC | 2013, 2014, 2015 | 2007, 2013–2014 | 1995, 2001, 2010, 2017 | 2014 |
| Ethiopia | 2014, 2015, 2016 | 2000, 2005, 2011, 2016 | 1995 | — |
| Ghana | 2013, 2014, 2015 | 1989, 1993, 1998, 1999*, 2003, 2004*, 2008, 2011, 2014, 2016*, 2017* | 1995, 2006, 2011, 2017 | — |
| India | 2016 | 1992–1993, 1998–1999, 2005–2006, 2015–2016 | 1995–1996, 2000 | (HMIS since at least 1980s) |
| Indonesia | 2015 | 1987, 1991, 1994, 1997, 2002–2003*, 2002–2003, 2007*, 2007, 2012*, 2012, 2017 | 1996, 2000, 2011 | — |
| Kenya | 2014, 2015 | 2008–2009, 2010*, 2014, 2015* | 1996, 2000, 2008, 2009, 2011, 2013–2014 | — |
| Lao PDR | — | 4 (1996–2012) | | — |
| Niger | 2015, 2016 | 1992, 1998, 2006, 2012, 2017 | 1996 | — |
| Nigeria | 2014, 2015, 2016 | 1990, 1999, 2003, 2008, 2010*, 2013, 2015* | 1995, 1999, 2007, 2011, 2016–2017 | 2013 |
| Pakistan | — | 8 (1995–2017) | | — |
| Tanzania | — | 1996 | | 2013 |
| Uganda | 2014, 2015, 2016 | 1988–1989, 1995–1996*, 1995, 2000–2001, 2004–2005*, 2006, 2007*, 2009*, 2011, 2011*, 2014–2015*, 2016 | None | 2012 |
| Zimbabwe | — | 2009, 2014 | | 2013 |

* Indicates special survey—e.g., Malaria Indicator Survey, AIDS Indicator Survey (see DHS Program, undated).

Table G.2 shows the years that PMA2020, DHS, and MICS surveys have been conducted in the 15 countries we evaluated.

The sections that follow provide a brief snapshot of each country included in this evaluation.

## Burkina Faso

The RAND team member conducted six interviews (in French) with eight individuals in Ouagadougou and attended the national consensus meeting. Based on those interviews, he

concluded that Burkina Faso has clearly progressed in its approach to incorporating data into decisionmaking for family planning since PMA2020 and Track20 began. Government officials and NGO stakeholders demonstrated a profound appreciation for the data being generated, and they reported progress in regard to the way in which data are used to inform programs and set targets. The main PMA2020 country partner is based at the University of Ouagadougou. The Track20 M&E officer is based in the MOH, with full salary support from the ministry. There seems to be solid collaboration between PMA2020 and Track20 and also with the NGO community.

**Notable Strengths.** With the growing availability and reliability of family planning data, decisionmakers have become accustomed to incorporating data into their decisionmaking to achieve national family planning objectives and family planning project targets. It was the opinion of several interviewees that PMA2020 and Track20 contributed significantly to this shift toward greater data use. These firm targets are indicative of the data-driven mindset that has become commonplace in the MOH (and elsewhere in the government), supported by the National Institute of Statistics. Burkina Faso has established a Reproductive Health Technical Group, which brings together a wide range of stakeholders to discuss data (especially DHS, PMA2020, and DHIS2) and family planning targets and to construct the National Plan for Family Planning 2017–2020. The level of data sophistication, according to several in-country and external sources, is considerably higher than that of other countries in the region.

**Notable Opportunities for Improvement.** The country's fertility rate is 5.4 children per woman, indicating that the country's family planning programs have not yet achieved their desired goals. A key opportunity for improvement was the provision of demand-driven data in terms of both content and geographic granularity—i.e., greater responsiveness of PMA2020 to country needs. The prospects for sustainability of PMA2020 and Track20 are uncertain.

## Côte d'Ivoire

The RAND team member conducted nine interviews (in French) with ten individuals in Abidjan and attended the national consensus meeting. PMA2020 programming was only recently under way, with the first PMA2020 survey to be conducted during 2017. Both the Track20 staff and PMA2020 principal investigator are based in the National Statistics Office. While Avenir supports Track20, the French Development Agency is co-funding PMA2020 with the Gates Institute.

**Notable Strengths.** Interviewees indicated their familiarity with various family planning data sources and the core FP2020 indicators. In addition, they viewed both Track20 and PMA2020 favorably. The country has a plan and budget for family planning programming through 2020. Decisionmakers understand the importance of family planning data and expect to use them to inform policy and resourcing decisions.

**Notable Opportunities for Improvement.** While the Track20 and PMA2020 staff are housed in the same government institution, and the PMA2020 principal investigator is also the Track20 director, the two programs are not yet optimally coordinated. For example, PMA2020 staff were not invited to the June 2017 national consensus meeting organized through Track20. Some interviewees raised concerns about data security with data from mobile devices uploaded to a server in Baltimore (not Abidjan). Recognizing local variability

in contraceptive use, others expressed a desire for more geographically granular data; some also expressed a desire to include non–family planning modules (e.g., nutrition) in PMA2020 surveys. Finally, the government was due to co-finance PMA2020 but had not yet approved (or released) funding as of June 2017, which may potentially delay the first and subsequent planned surveys.

## Democratic Republic of the Congo

The RAND team member conducted eight interviews (in French) with eight individuals in and around Kinshasa. Both PMA2020 and Track20 have contracted with Tulane University to provide technical assistance for the two programs. The main PMA2020 country partner is based in the University of Kinshasa, but his salary is supported entirely by the government. The Track20 M&E officer is based in the MOH, with full salary support by Track20.

**Notable Strengths.** The general perception among the key stakeholders is that data are expected, valued, and used by policymakers in the decision process, including national and provincial governments and the NGO community. The Track20 M&E officer noted that he has good access to the different heads of departments within the government and NGOs through the consensus meetings and various trainings. All provinces regularly attend the annual national consensus meetings, and their representatives receive valuable training. There is also an effective and widely respected multisectoral coordinating committee run by the M&E officer's supervisor.

**Notable Opportunities for Improvement.** DRC is still in the beginning stages of data maturity. The level of data maturity is generally highest in Kinshasa and the surrounding provinces, dropping greatly in more distant provinces. There is a cadre, albeit small, of highly qualified, capable data professionals (the head of the national technical coordinating committee, the Track20 M&E officer, and the PMA2020 co-principal investigators, for instance). Efforts are under way to train more individuals at both the national and provincial levels.

## Ethiopia

The RAND team member conducted 11 interviews with 12 stakeholders in Addis Ababa. The PMA2020 country partner is based in Addis Ababa University; the university pays this individual's full salary, and PMA2020 provides a supplemental consultancy fee. The Track20 M&E officer works within the Directorate of Planning and Programming, an especially strong unit within the MOH. Track20 fully supports the officer's salary.

**Notable Strengths.** The core PMA2020 team is knowledgeable and committed, has high technical capacity, and places great emphasis on data quality and data use for decisionmaking. In addition, the resident enumerators and their supervisors believe in the importance of the program to help make family planning services accessible to all women, and they have great respect for the PMA2020 country leadership. Overall, there is more awareness of PMA2020 compared to Track20 in Ethiopia, but there is strong government support for both programs. The team also has strong rapport with the MOH and is committed to making sure that the data collected and reports produced are easily accessible to and used by decisionmakers at all levels of government, as well as university and NGO researchers. Stakeholders at the Federal

Ministry of Health commented that annual PMA2020 surveys and Track20 estimates have been essential to help track progress toward Ethiopia's FP2020 commitments and improve the rate of contraceptive use.

**Notable Opportunities for Improvement.** Nearly all interviewees called for subnational PMA2020 data and subnational family planning estimates, and several called for data in program areas beyond family planning (e.g., nutrition, vaccination, migration). Current aggregated estimates do not meet the needs of either regional or national health authorities, who continue to use DHS data for their planning purposes. Some interviewees expressed a desire for less centralization of the PMA2020 process and greater management of the survey forms in the country, which they felt would reduce the overall time for translation into three languages and launching of the surveys. The country's Track20 M&E officer has multiple responsibilities, which constrain the time available for quality checks and analysis of the country's family planning data. Some interviewees commented that, at times, decisionmakers do not understand the technical elements of the indicator estimates. Enhancing their understanding would likely enhance buy-in by key decisionmakers. Finally, there is little to no apparent interaction between PMA2020 and Track20 staff in Ethiopia.

## Ghana

Ghana is the only country in which PMA2020 operates that does not also participate in Track20. The RAND team member conducted 11 interviews with 12 individuals in Accra and Kumasi. Ghana has a strong tradition of using data for decisionmaking, previously relying mainly on the Ghana DHS and now using the additional PMA2020 data. The national surveys are well institutionalized within the (governmental) Ghana Statistical Service. The PMA2020 lead country partner is based in the School of Medical Sciences at the Kwame Nkrumah University of Science and Technology in Kumasi. The surveys are conducted in collaboration with the University of Development Studies and with the support of the Ghana Statistical Service and the Ghana Health Service.

**Notable Strengths.** The PMA2020 country team is competent, knowledgeable, and proud of their work. All end users interviewed find PMA2020 data useful, citing in particular the regularity, quality, and presentation of the data. The program also appears to have a strong relationship with the government's implementing agency for family planning programs. The university-based PMA2020 team has helped build local capacity—for example, by training Ghana Statistical Service staff in mobile data collection.

**Notable Opportunities for Improvement.** Interviewees expressed the need for increased awareness, dissemination, and engagement with key stakeholders. Also, some interviewees commented that the government's sense of ownership of PMA2020 is not strong since the program is based out of a university. The RAND team member opined that institutionalization of PMA2020 will be proportional to awareness, data use, and ownership by the government.

## India

The RAND team member conducted 26 interviews with 29 individuals in Jaipur (Rajasthan state) and New Delhi. PMA2020 and Track20 programs are operating in different states, so

most interviewees were familiar with just one or the other program; an exception was an official from the national Ministry of Health and Family Welfare, who was familiar with both. India has a very large family planning program. The PMA2020 country partners are based in the Indian Institute for Health Management Research University. Track20 has one team lead and four M&E officers, who operate at both the national and state levels. Track20 supports the full salaries of all in-country staff.

**Notable Strengths.** Track20 estimates are generally considered high quality, well respected across stakeholders, and garnering government buy-in. In-country staff members received high praise for their capacity, responsiveness to government needs, and involvement in activities beyond their Track20 roles. The National Data Consensus Meetings also received high praise, with a notable improvement this year, including an expert's panel for demographers and statisticians to discuss quality of data sources and estimate methods. In addition, the meeting to disseminate FPET estimates to stakeholders was run by the Ministry of Health and Family Welfare, demonstrating support for and ownership of the data.

The PMA2020 program is relatively new, but the principal investigator is very experienced (e.g., conducting National Family Health Surveys) and highly respected among peers. The rest of the PMA2020 team is young, energetic, and fully invested in the program's mission. State-level stakeholders rate the quality of PMA2020 data as high, whereas national-level stakeholders gave PMA2020 mixed reviews. General consensus is that further judgments should be withheld until more survey rounds have been conducted.

**Notable Opportunities for Improvement.** More extensive and proactive capacity-building of government staff will be integral to the sustainability of Track20. Track20 can also raise awareness of the program among government officials and other stakeholders—for example, through a more robust dissemination plan targeted to data user needs, including more geographically granular indicator estimates.

## Indonesia

The RAND team member conducted 13 interviews with 20 individuals in Jakarta and one subsequent telephone interview with one individual in Sulawesi. Track20 and PMA2020 are officially housed within BKKBN. PMA2020 is additionally supported by Universitas Hasanuddin and Universitas Sumatera Utara. In addition, Avenir has placed a highly experienced, full-time consultant in Indonesia to help work on a service statistics initiative. Track20 has covered the full salary of the M&E officer through 2017, after which BKKBN will assume responsibility for the full salary. Because the country has a decentralized health system, data collected at the subnational level are perceived to be most relevant and useful. BKKBN has its own annual family planning survey. Because of perceived duplication between this survey and PMA2020, the Gates Institute halted the 2017 PMA2020 survey and is working with BKKBN and the university on a plan to subsume PMA2020 under the BKKBN survey, which had already adopted PMA2020 data collection methodology and key content. Numerous agencies work in family planning and reproductive health, complicating identification of who the decisionmakers are and further complicating issues of data ownership and use.

**Notable Strengths.** Indonesia has trusted and experienced family planning advisers, who provide strong mentorship for junior researchers. There is a rich appreciation for data in general and recognition of the importance of publishing these data. Lastly, there is an understanding

of how to sustain Track20 and PMA2020 within a deeply bureaucratic landscape; for example, the deliberate lack of branding associated with the two platforms is anticipated to increase local buy-in and ownership.

**Notable Opportunities for Improvement.** Indonesia's enormous amount of data is not equally matched by analysis, dissemination, collaboration, and coordination. The lack of clarity over who ultimately controls decisionmaking drives confusion over trust of data sources and how to actually apply data to decisionmaking. Much of the decisionmaking should be occurring within the Ministry of Planning, but lines of communication and data sharing with the ministry are unclear. As in several other countries, family planning service statistics are unreliable and of poor quality. There are high demands placed on the Track20 M&E officer, who also serves as a researcher and whose first-line duties are for research rather than for Track20. PMA2020 is useful for its methodology and rigor of data collection, but confusion remains over ownership, especially if PMA2020 becomes subsumed under the BKKBN survey.

## Kenya

The RAND team member conducted 12 interviews with 12 individuals in Nairobi and Kilifi County (near Mombasa). Both PMA2020 and Track20 are well established and enjoy relatively high visibility in the country. The PMA2020 country partner is based in the University of Nairobi, with salary paid by the International Center for Reproductive Health—Kenya. The Track20 M&E officer is based in the MOH. The Track20 M&E officer is formally considered an employee of Kenya's National Council for Population and Development (part of the Ministry of Planning) seconded to the MOH. Her salary is paid entirely by Track20 and is distributed through her home agency, which retains some personnel oversight but does not manage her day-to-day activities.

Kenya has decentralized its family planning resources since 2014, now to more than 40 jurisdictions (counties).

**Notable Strengths.** The government is supportive of family planning, but interviewees were uncertain about the outcome of the upcoming election in August 2017 and implications for ongoing support for family planning. PMA2020 staff have a good working relationship with the MOH, and the surveys have a good reputation for accuracy (bolstered by confirmation of results from the most recent DHS). Respondents described the Track20 M&E officer as proactive and a good communicator. Both PMA2020 data and Track20 estimates are used in decisionmaking; in particular, the government's implementing partners draw from subnational (provincial-level) PMA2020 data to target their technical assistance programming and for advocacy purposes.

**Notable Opportunities for Improvement.** PMA2020 has a comparative advantage because it includes data from both the public and private sectors. However, the data are not tagged as such. In addition, PMA2020 does not cover the entire country, missing some of the poorest and lowest-performing counties. Track20 provides technical assistance to counties, but its own capacity is limited and does not meet the apparent demand; it is more prominent at the national level.

## Lao People's Democratic Republic

The RAND team member conducted 11 interviews with 13 individuals in Vientiane. The Track20 program provides technical support to the MOH and UNFPA, which pay the salaries of Track20 staff. Family planning is a relatively low government health priority, and familiarity with the Track20 program among policymakers was low. However, at the time of the RAND country visit, only one Track20 national consensus meeting had been conducted. Since 2011, Lao PDR's transition from a World Bank least developed country into a lower middle income country triggered donor requirements for greater government contributions for projects and resulted in a reduction in overall official development assistance. Despite the time that has passed since then, multiple MOH respondents mentioned that funding and personnel constraints remain a threat to program sustainability.

**Notable Strengths.** Most respondents noted that data use, while still relatively weak, has become stronger in recent years. Most respondents rate the sense of Track20 ownership in the government as medium to high. Of the various data processes, collection and management seem to be the best-institutionalized, with most respondents rating the institutionalization of data dissemination lower, and of data analysis lower still.

**Notable Opportunities for Improvement.** Some respondents at international agencies point out that Track20 is sequestered in a low-impact unit within the MOH—the Maternal and Child Health Center. An especially important missing link in communications is apparently between this center and the Statistics Division, which controls the funds for family planning programming (including donor contributions), but whose director was unfamiliar with Track20. The three M&E officers have duties beyond Track20 and, by many accounts, are overloaded with multiple disparate tasks, of which Track20 is only one. The culture of data use is weak even at the national level and gets progressively weaker in outlying jurisdictions. Some interviewees commented that more training is needed, especially at subnational levels. Numbers are uncritically reported up the chain and not used for local analysis and program management. Planning is a top-down process. Many decisions are made based on heuristics, not evidence. Also, service statistics are widely considered to be unreliable (a theme echoed from other countries included in this evaluation).

## Niger

The RAND team member conducted nine interviews (in French) with 13 individuals in Niamey. He was also able to witness a well-attended, high-quality training session for PMA2020 resident enumerators. The Track20 M&E officer is based in the MOH, but the salary of the M&E officer is supported by Pathfinder International. Track20 officials also stressed the importance of the close collaboration they have enjoyed with Johns Hopkins University and how that has built their capacity and skill sets as well.

**Notable Strengths.** The highly regarded rigor of PMA2020 data collection and Track20 statistical analysis derive at least in part from the professionalism and passion of the country teams for the production and analysis of quality, meaningful data. The Track20 M&E officer receives solid support from the Gates Institute. The close ties of both programs to the government are another notable strength—both are housed in government offices and have easy and

direct access to ministry-level decisionmakers; they report no problems in getting their data seen and making their voices heard.

**Notable Opportunities for Improvement.** Challenges noted by both program staff and NGOs were problems with communication, dissemination, education, and utilization of data and reports. Both PMA2020 and Track20 staff felt that their data are used in government briefings and publications, but opinions differed between program staff and the NGO community concerning the degree to which family planning decisions are actually data-driven. According to NGO leadership, PMA2020 data and Track20 reports are not publicized well enough and are not reaching all necessary audiences (NGOs, local communities). Furthermore, many stakeholders called for greater geographic granularity of PMA2020 surveys, especially for rural areas where access to health and reproductive services is lower and fertility rates are higher. PMA2020 wants to generate more provincial and local data but is hamstrung by limited funding and personnel.

## Nigeria

The RAND team member conducted 24 interviews with 27 individuals in Ibadan, Abuja, and Lagos. PMA2020 appears to be a well-coordinated program with a team of qualified staff. PMA2020 has two country principal investigators and is operated primarily out of the private research organization led by one of the principal investigators. The co-principal investigator is based in a leading university. Track20 has two M&E officers based in the Federal Ministry of Health and two in each state MOH in the two states in which Track20 presently operates. Their salaries are fully covered by the government. Track20 has also hired and pays the salary of a full-time Nigerian consultant.

**Notable Strengths.** The PMA2020 team is competent and committed to ensure high-quality data and has developed an impressive system for data auditing system and quality assurance. The country team carries out these functions mostly independently, with some support from the Baltimore team. The team also engages with state governments through the dissemination meetings. Across the board, end users of data mentioned their appreciation for the quick turnaround and timeliness of PMA2020 data. Respondents cited specific examples in which actions by state governments were initiated following presentation of PMA2020 data during dissemination—e.g., in Lagos State, where PMA2020 data contributed to the development of the state's costed implementation plan.

The Track20 M&E officers are motivated and collaborate well with one another. Track20 is deeply engaged with the Federal Ministry of Health and the MOHs in the states in which it operates. In addition, there is a strong focus on ensuring that the M&E officers are able to analyze family planning data and generate projections by themselves, which helps to further improve in-country capacity in the federal and state MOHs.

**Notable Opportunities for Improvement.** The absence of government involvement in the operation of the PMA2020 surveys reduces the sense of ownership, although the team has built strong relationships with the state governments where they operate. Key gaps remain in regard to improving engagement and dissemination at the federal government level. There is also an opportunity to expand the survey to other states in the country while exploring potential funding and sustainability mechanisms with interested states even before the survey begins.

The sense of ownership of Track20 estimates is relatively strong, likely due to government commitment to the capacity-building of the M&E officers (who are government employees). However, there appeared to be an ongoing need for training updates to further improve the quantitative skills of M&E officers—e.g., in moderate/advanced Excel techniques and the use of the FPET tool. Track20 might consider phased expansion of the program to the other 34 states in the country to build capacity of M&E officers across the country and to increase the visibility of the program.

## Pakistan

The RAND team member conducted ten interviews with ten individuals in Pakistan by telephone (in Urdu and English). Pakistan is one of just two Track20 countries where the M&E officer is not based in a government office (Tanzania is the other country). The M&E officer is based in the Population Council, which supports 25 percent of his salary; Track20 supports the remaining 75 percent of the salary. All stakeholders commented on the unique nature of Track20 in the context of Pakistan's increasing decentralization of health authority to the provincial level across all four provinces and various regions in the country. The sole Track20 M&E officer is responsible for all provinces and regions and provides valuable support to provincial governments in terms of keeping them on track toward FP2020 goals. Track20 appears to have played an influential role across both provinces and existing data collection platforms (e.g., at the Pakistan Bureau of Statistics) by encouraging the validity of available data and improving data quality. Track20 data are now being used in the five-year Annual Plan for Pakistan, which is reportedly one of the most important government planning documents—it occupies a key role in monitoring and planning for key targets throughout various sectors.

**Notable Strengths.** Pakistan's Country Engagement Working Group is an effective coordinating mechanism; the Working Group is a forum for provincial- and national-level government stakeholders, NGOs, and donors to discuss data-related issues and the use of Track20 data in the development of costed implementation plans. Track20 has also provided valuable opportunities for capacity-building within the National Institute for Population Studies, both alone and in conjunction with a USAID-supported implementing partner. There are several concrete examples of how family planning data have been used, both to improve family planning services and to inform multiyear provincial-level costed implementation plans and the national-level plan.

Stakeholders indicated that Track20 is slowly helping to create ownership of family planning data at the highest level, as well as providing data that enable provinces to compare their progress against one another.

**Notable Opportunities for Improvement.** Several stakeholders commented on the need for more training and engagement at the provincial level so that provincial authorities can use the FPET tool and, more specifically, their local (district- and subdistrict-level) service statistics, for which the governments feel great ownership and are motivated to improve. Some stakeholders suggested embedding Track20 staff members within the M&E units of the Population Welfare Department. Such an approach would provide active, hands-on capacity-building through daily interactions, which, in turn, would facilitate an improved culture of data use within the provinces.

## Tanzania

The RAND team member conducted 13 interviews with 14 individuals in Dar es Salaam and was also able to attend the national consensus meeting. The Track20 M&E officer is an employee of FHI360 (an NGO) but is currently seconded to the HMIS unit of the MOH's Reproductive and Child Health Department; the salary is split between Track20 (75 percent) and FHI360 (25 percent). The M&E officer has become a go-to point of contact for DHIS2 data requests by both government staff and development partners alike, as access to the database is very restricted.

**Notable Strengths.** Many development partners pointed to the utility of the annual Track20 consensus workshops as the only multistakeholder function focused specifically on family planning data. There was broad consensus that the Track20 M&E officer has been able to build strong relationships with his colleagues within the MOH and has proven useful in data analysis and supporting data quality assessments. Further, the growing culture and focus on data within the family planning unit is attributed to his presence within the MOH. The government is also very pleased with the incorporation of HMIS data into the FPET model in 2017.

**Notable Opportunities for Improvement.** The extent of Track20's capacity-building within the HMIS and family planning departments remains a concern. Interviewees commented on the heavy reliance on a single person—the M&E officer—for regular access and analysis of data, at the expense of dedicated focus and time for more proactive analysis of family planning data. Several stakeholders expressed concerns about the quality of DHIS2 data and noted that it could be beneficial for the M&E officer to champion improvements in data quality. However, this would require travel to the district levels, and funding for travel is not currently supported under his contract. Many development partners left the consensus workshop feeling that they were exposed to a large amount of very important data and findings but had insufficient time for in-depth discussion or complete understanding of the implications. Track20 could facilitate improvements through advance circulation of discussion materials, scheduling more discussion time during the meetings, and further strengthening the M&E officer's presentation and communications skills.

## Uganda

The RAND team member conducted 17 interviews with 26 individuals in and around Kampala. She was able to observe resident enumerators conducting interviews in the outskirts of Kampala and also attended the national consensus meeting organized by Track20. The PMA2020 country partner is based at Makerere University. The Track20 M&E officer is based in the MOH, and the M&E officer's salary is fully supported by Track20.

**Notable Strengths.** There is considerable awareness of the *importance* of data to attract investments in family planning, lend credibility to interventions and campaigns, and identify such issues as stock-outs of contraceptive materials, with one interviewee commenting that "You cannot have advocacy without data." This signals a cultural shift in terms of the importance of establishing an evidence base for interventions, as well as the various levels involved in improving family planning statistics, from information supplied by providers to understanding the contraceptive method mix.

PMA2020 is considered to be high quality, or "authentic," because of its support from the Ugandan Bureau of Statistics, which houses the national census and the DHS. Because it is conducted through the Makerere School of Public Health, it is likewise thought to be an unbiased data source. The 2016 DHS estimates were in consonance with PMA2020 estimates, which further strengthened the credibility of the PMA2020 data. Stakeholders also appreciated the brevity and accessibility of the PMA2020 reports and the more detailed family planning indicators it provides (in contrast with the DHS).

Track20 is in regular contact with advocacy groups involved in family planning, reproductive health, and child and maternal health. Advocates are well informed and strive for improvements within the programs, particularly by pushing for more data on family planning behavioral factors, improvements in latent demand for modern contraceptives, and improved service delivery statistics. Track20 and PMA2020 collaborate with one another through regular meetings and phone calls. The national consensus meeting, while driven by Track20 and the Family Planning Focal Point personnel, was supplemented by input from the PMA2020 principal investigator.

**Notable Opportunities for Improvement.** Despite awareness of the importance of data, there is only limited understanding of what the data mean and how information should be used to guide decisionmaking. The national consensus meeting revealed critical misunderstandings in statistics and major inconsistencies in the reporting of service delivery data, which themselves are unreliable. Without reliable service statistics and stronger data capacity at subnational levels, operationalization of Track20 estimates and use of PMA2020 data may be limited.

## Zimbabwe

The RAND team member conducted eight interviews with ten individuals in Harare and one additional interview in New York City. She also attended the national consensus meeting organized by Track20. The Track20 M&E officer had previously worked in academia and the NGO sector and was recruited into a newly created government position. The Track20 M&E officer is based in the Ministry of Health and Child Care, with her salary fully supported by Track20.

**Notable Strengths.** The M&E officer is well respected and has elevated the role of family planning in the ministry. Her supervisor, the director of family health, has promoted Track20's work to the director of preventive health (similar to the assistant secretary level in the United States) and has used its estimates when working with parliamentarians. Track20 estimates are considered accurate, reliable, and comparable to the DHS. While users would like data that are disaggregated to the provincial level, there were no calls for more frequent data collection.

**Notable Opportunities for Improvement.** Track20 appears to be well embedded within the country's data structure. However, implementers still plan new projects based on data from the DHS instead of Track20. Track20 and data use need to be strengthened at the provincial and district levels. Provincial representatives at the national data consensus meeting were largely silent. Some respondents suggested that local representatives are not comfortable using data and need training on how to interpret and use data. Because provincial leaders cannot calculate estimates, they have to wait for the annual meeting to receive progress updates. The lack of funding for family planning commodity procurement, distribution, or training suggests

that Zimbabwe will not be able to co-finance or fully assume responsibility for the Track20 program in the near term.

APPENDIX H

# Additional Detail on Statistical Methods and Country-Specific Analyses

This appendix describes in more detail optimal probability design, including how it is calculated. We also report results of the in-depth analyses for the five countries that are summarized more generally in Chapter Seven.

## Optimal Probability Design

As described in brief in Chapter Three, the variance of the estimate for an indicator in a cluster provides information on how many individuals need to be surveyed in each cluster. For PMA2020 specifically, each enumeration area is a cluster; individuals are nested within households, and households are nested within clusters. For each indicator, assuming that the estimated variance of **each indicator** is a good approximation of the population variance of each indicator in the enumeration areas, the RAND team used the "optimal" design scheme proposed by Cochran and William (1977) to choose the best probability design that allows for the best precision of each indicator. As precision is the inverse of variance (the smaller the variance, the larger the precision of the estimate), this sampling strategy chooses the number of survey respondents in each cluster that minimizes the variance of the population estimate of the indicator within the country. The formula for such variance was derived as follows:

Assuming that there are $S$ different enumeration areas and that for a particular design allocation $k$ (that takes value 1 for the current PMA2020 design and 2 for an optimal design in this study),

- $N_i$ is the total population number of eligible people in the $i^{th}$ enumeration area. Because the true value of $N_i$ is unknown in this study, for ease of computation, we will assume that all enumeration areas are large and have approximately the same number of people that can be sampled.
- $n_{ik}$ is the size of the sample that will be taken in this $i^{th}$ enumeration area for the design allocation $k$. This is the sample size that would optimize the design.
- $\sigma_i$ is the standard deviation of the indicator of interest (e.g., modern contraceptive method use) in the $i^{th}$ enumeration area. In this study, we will assume that the within-enumeration-area standard deviation observed from the PMA2020 sample is a good approximation of the standard deviation of the population.
- $V_k$ is the variance estimate of the population indicator of interest for the allocation $k$, the variance that is minimized in an "optimal design."

- $RSE_{ik}$ is the relative standard error of the estimate of the family planning indicator in the $i^{th}$ enumeration area for the allocation $k$.
- $\mu_{ik}$ is the mean of the indicator (e.g., prevalence of modern contraceptive method use), the variable of interest, in the $i^{th}$ enumeration area for the allocation $k$.

Cochran and William showed that

$$V_k = \sum_{i=1}^{S} \frac{N_i^2 \sigma_i^2}{n_{ik}} = \sum_{i=1}^{S} N_i^2 \mu_{ik} RSE_{ik}^2 .$$

For two different design allocations $j$ and $k$, the better design is the one with the smaller variance or a study of the ratio $V_k/V_j$. This optimal probability design for a specific indicator, also called the Neyman allocation design, has a closed-form solution. The sample size within enumeration areas that will result in the specific indicator's optimal sample is

$$n_k = n \frac{N_k \sigma_k}{\sum_{s=1}^{S} N_s \sigma_s},$$

where $n$ is the total sample size desired, here the sample size obtained in the PMA2020. As described in the main text of the report, for traditional contraceptive methods, for example, which have little variation (i.e., small $\sigma_i$) in prevalence across enumeration areas, this optimization will require only taking a small number of observations in such areas. However, for other enumeration areas with an indicator showing large variation, a larger number of observations will be required to achieve the same level of accuracy. The optimal sample size for one indicator (e.g., mCPR) will not necessarily be optimal for another indicator (e.g., traditional contraceptive prevalence rate). In this evaluation, comparisons were made to assess the indicators for which variance can be minimized when using optimal design. The actual survey sample size can be determined based on the indicator of main interest (such as mCPR) or the indicator that requires the largest sample size among different indicators of interest.

The main goal of this analysis is to determine the gain that would be obtained if using an optimal design relative to the PMA2020 design, a useful ratio is that of the variances of two different allocations:

$$R = \frac{V_2}{V_1} = \frac{\sum_i \frac{N_i^2 \sigma_i^2}{n_{i2}}}{\sum_i \frac{N_i^2 \sigma_i^2}{n_{i1}}} .$$

With the assumption that for different enumeration areas $N_i \approx N_j$, this yields

$$R = \frac{V_2}{V_1} = \frac{\sum_i \frac{\sigma_i^2}{n_{i2}}}{\sum_i \frac{\sigma_i^2}{n_{i1}}}.$$

For each indicator, the ratio of the optimal design variance by the PMA2020 variance, termed *R*, hereafter referred to as percentage gain from optimal design, is reported. $R = 0.5$, for example, indicates a survey design that can help reduce the variance observed in the PMA2020 data by 50 percent.

## Country-Specific Statistical Analyses

The sections that follow present the analyses undertaken for each of the five countries we examined: comparison of PMA2020 to the DHS in terms of representativeness of the sample; round-to-round comparisons across PMA2020 surveys; detectable margin of error; and precision, design effect, and frequency of data collection.

### Ghana

Ghana had four rounds of PMA2020 survey data available for analysis, gathered between 2013 and 2015 (Table H.1).

#### *Comparison of Ghana PMA2020 to the DHS*

To assess the representativeness of the PMA2020 survey in Ghana, we compared PMA2020 data with those in the DHS fielded in Ghana between January and March 2014. For this comparison, we focus mainly on the Round 2 Ghana PMA2020, which was conducted between February and May 2014, since it is the most contemporaneous to the 2014 Ghana DHS.

When we compared the sample characteristics between PMA2020 and the DHS, there was a statistically significant difference in the average age of the surveyed population (29.0 for PMA versus 29.9 for the DHS, $p < 0.001$; Table H.2), but such difference is practically small (close to one year). A similar difference was observed when we compared the next-closest PMA2020 (Round 1, fielded between September and October 2013); the average age in that

**Table H.1**
**Ghana PMA2020 Rounds and Sample Size**

| Round | Data Collection Period | Sample Size (number of women) |
|---|---|---|
| 1 | September to October 2013 | 3,766 |
| 2 | February to May 2014 | 4,009 |
| 3 | October to December 2014 | 4,663 |
| 4 | May to June 2015 | 5,271 |

**Table H.2**
**Ghana Population Characteristics—DHS and PMA2020**

| Characteristic Definition | Level | DHS (Jan–Mar 2014) | PMA2020 Round 1 (Sep–Oct 2013) | PMA2020 Round 2 (Feb–May 2014) | PMA2020 Round 4 (May–June 2015) | *P* Value DHS to Round 2 | *P* Value Across PMA2020 Rounds |
|---|---|---|---|---|---|---|---|
| **Person-level variables (women)** | | | | | | | |
| Age | | 29.87 (29.62, 30.12) | 29.03 (28.78, 29.28) | 29.04 (28.78, 29.29) | 28.56 (28.31, 28.81) | 0.000 | 0.000 |
| Marital status | 1. Currently married | 45.16% | 53.74% | 54.08% | 47.41% | 0.000 | 0.000 |
| | 2. Currently living with partner | 12.91% | 7.97% | 9.30% | 11.22% | | |
| | 3. Divorced or separated | 6.70% | 4.17% | 4.95% | 5.55% | | |
| | 4. Widow or widower | 2.86% | 2.10% | 1.90% | 2.07% | | |
| | 5. Never married | 32.36% | 32.02% | 29.76% | 33.74% | | |
| Highest level of school completed | 0. Never attended | 24.28% | 26.25% | 26.46% | 22.93% | 0.000 | 0.000 |
| | 1. Primary | 18.59% | 18.57% | 19.50% | 18.06% | | |
| | 2. Middle | 37.55% | 34.39% | 33.27% | 34.94% | | |
| | 3. Senior | 14.11% | 14.51% | 14.57% | 16.81% | | |
| | 4. Higher | 5.47% | 6.27% | 6.21% | 7.26% | | |
| **Household-level variables** | | | | | | | |
| | Has electricity | 70.99% | 68.38% | 71.58% | 76.60% | 0.506 | 0.000 |
| | Has wall clock | 29.59% | 32.81% | 33.41% | 32.60% | 0.000 | 0.000 |
| | Has black and white television | 1.99% | 4.53% | 2.97% | 3.73% | 0.001 | 0.000 |
| | Has color television | 54.31% | 51.73% | 51.69% | 56.39% | 0.007 | 0.000 |
| | Has mobile phone | 82.20% | 80.09% | 79.04% | 82.06% | 0.000 | 0.000 |
| | Has landline | 1.17% | 1.89% | 1.63% | 1.65% | 0.032 | 0.000 |
| | Has refrigerator | 28.77% | 28.87% | 29.40% | 31.49% | 0.478 | 0.011 |
| | Has generator | 2.62% | 2.37% | 1.37% | 1.70% | 0.000 | 0.000 |
| | Has computer | 11.87% | 12.25% | 10.17% | 10.44% | 0.006 | 0.000 |
| | Has bicycle | 30.12% | 27.57% | 26.46% | 24.36% | 0.000 | 0.000 |
| | Has motorcycle | 12.41% | 9.43% | 8.98% | 9.55% | 0.000 | 0.000 |
| | Has car | 7.56% | 7.74% | 6.81% | 7.95% | 0.139 | 0.225 |

**Table H.2—continued**

| Characteristic Definition | Level | DHS (Jan–Mar 2014) | PMA2020 Round 1 (Sep–Oct 2013) | PMA2020 Round 2 (Feb–May 2014) | PMA2020 Round 4 (May–June 2015) | *P* Value DHS to Round 2 | *P* Value Across PMA2020 Rounds |
|---|---|---|---|---|---|---|---|
| Wealth quintile | 1. Lowest quintile | 21.20% | 20.18% | 24.44% | 25.93% | 0.000 | 0.000 |
| | 2. Lower quintile | 20.46% | 19.69% | 20.84% | 19.91% | | |
| | 3. Middle quintile | 21.77% | 19.84% | 19.70% | 17.39% | | |
| | 4. Higher quintile | 19.32% | 20.18% | 18.03% | 17.94% | | |
| | 5. Highest quintile | 17.25% | 20.12% | 17.00% | 18.83% | | |

NOTES: To minimize the number of columns in each of these tables comparing the DHS with PMA2020, we present results for the first two rounds of data and the last or most recent round of data only. *P* values were calculated using Tukey-Kramer adjustment to account for multiple comparisons.

round was also 29.0, statistically different from the 2014 DHS ($p < 0.001$). Note that due to the large sample sizes in the different surveys, very small differences can be statistically significant. As such, the age difference observed does not necessarily challenge the representativeness of the PMA2020 survey.

More-substantial differences are seen for marital status. Both the Round 1 and Round 2 PMA2020 surveys for Ghana found that about 54 percent of the women were currently married, while the DHS results reported 45 percent, even though the DHS mean age was slightly higher. (Typically, younger women are less likely to be married than their older counterparts.) The differences between the PMA2020 and DHS estimates are statistically significant and may reflect some of the differences in the actual implementation of the sampling strategies between the PMA2020 and the DHS. Regarding education, there is a statistically significant difference between PMA2020 and DHS estimates: Slightly more of the women surveyed by PMA2020 had never attended school (about 26.5 percent, compared with 24.3 percent in the Ghana DHS, $p < 0.001$), while DHS women had completed middle school by more than 4 percentage points.

At the household level, the PMA2020 results differed slightly from the DHS results, with fewer PMA households reporting that they had a bicycle, for example, but some of these differences, though statistically significant, were quite small (a few percentage points or less). Differences between PMA2020 and the DHS were also observed in the household-level wealth categorized in quintiles, though the differences varied across PMA2020 rounds, with Round 1 of PMA2020 having slightly fewer in the lowest quintile and more in the wealthiest compared with the DHS, but Round 2 having more in the poorest and the same in the wealthiest; while statistically different overall, all differences between wealth quintiles in the DHS compared with PMA2020 Round 1 and Round 2 were 3.2 percentage points or less.

### *Comparison Across PMA2020 Rounds*

When assessing variation across the rounds of the PMA2020 surveys in Ghana, no noteworthy differences were found in age or education. There were some statistically significant differences in marital status across rounds (for example, currently married ranged from 47 percent to 54 percent across the rounds, peaking in Round 2). Even with the variation

among PMA2020 rounds, the percentage of women who are currently married is never as low as that in the 2014 DHS. Household-level variables were similar in most cases across rounds, with most differences less than 3 percentage points. One exception is with the rate of having electricity or mobile phones in the households. For example, 80 percent of households reported having a mobile phone in Round 1 of PMA2020 in Ghana; this dropped to 69 percent of households in Round 3, while 82 percent had them in Round 4. Other variables were more consistent across rounds, with no statistical significance found in differences in the percentages of households having cars or motorcycles, and the statistically significant differences that did exist were generally quite small.

### *Detectable Margin of Error and Precision*

As noted elsewhere, sample sizes for the PMA2020 surveys were determined based on the mCPR, with a goal of a less-than-3-percentage-point margin of error. Power calculations for subsequent rounds did not take updated mCPR into account. However, our analysis shows that the actual estimated margin of error for that indicator remains quite small, 1.1 or 1.2 percent, in all four rounds in Ghana, and that even though the sample size focuses on that single indicator, the margins of error for the other indicators considered here were all also under 3 percentage points, except for current use of the oral contraceptive pill in Round 2, for which the margin of error was 3.5 percent. This is indicative of the smaller number of women answering the question about oral contraceptive pills, especially in Round 2, where only 578 respondents answered that question, as opposed to 3,766 in Round 1, 949 in Round 3, and 1,321 in Round 4. This may be due to the way this indicator was calculated at the Gates Institute: Missing values may be set to "no" for binary indicators.

In general, particular questions where only a smaller part of the sample is eligible to participate will lead to larger margin of error and, as such, should be used with caution, especially when greater accuracy is of interest. When it comes to the observed variation in the mCPR indicator (standardized in coefficient of variation [CoV]), such variation seemed to be high and slightly decreasing over time, at 4.01 in Round 1, 3.74 in Round 2, 3.09 in Round 3, and 2.48 in Round 4, suggesting a decrease in variance. For reference, a CoV less than 1 is reflective of a standard deviation smaller than the mean of an indicator, while a value more than 1 reflects a standard deviation larger than the mean; for the different rounds, the standard deviation is two to four times greater than the mean. Note that for binary outcomes, under a binomial distribution, the variance and the rate of mCPR are related. Across the different outcomes, there is much variation in the CoV, which is probably explained by the fact that many of these indicators have varying scales and have different rates. The use of traditional contraceptive methods, for example, is rare in Ghana (1.1 percent in Round 1), while use of any modern method (mCPR) is more common (rate of 14.2 percent in Round 1).

### *Design Effect*

Across all indicators, very small ICCs were observed (Table H.3). Most of them are estimated at 0.1 or less, with some variation from indicator to indicator. The design effect, which is a combination of ICC and the number of observations in an enumeration area, varied across the different indicators. Though there was similarity in ICCs across the indicators, the variation in design effect from 3.37 to 6.78 across all indicators and rounds in Ghana can potentially be explained by the average number of respondents in enumeration areas (the clusters), which was around 38 in Round 1, 40 in Round 2, 46 in Round 3, and 52 in Round 4. For PMA2020,

**Table H.3**
**Ghana: Design Effect and Optimal Design Gain—Round-to-Round Comparisons**

| | Round 1 | | | Round 2 | | | Round 3 | | | Round 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain |
| Current use of any contraceptive method | 0.07 | 3.37 | 80.0% | 0.08 | 4.11 | 73.9% | 0.13 | 6.78 | 84.0% | 0.10 | 6.14 | 88.0% |
| Current use of any modern contraceptive method | 0.07 | 3.36 | 78.9% | 0.07 | 3.69 | 74.1% | 0.10 | 5.46 | 84.0% | 0.08 | 5.14 | 88.5% |
| Current use of any traditional contraceptive method | 0.11 | 4.79 | 83.2% | 0.11 | 5.20 | 79.1% | 0.11 | 5.73 | 77.8% | 0.05 | 3.65 | 83.1% |
| Total unmet need (spacing and limiting) | 0.05 | 2.83 | 80.8% | 0.03 | 2.34 | 83.6% | 0.04 | 2.96 | 88.1% | 0.04 | 3.01 | 91.6% |
| Current use of the oral contraceptive pill | 0.02 | 1.74 | 77.7% | 0.14 | 1.68 | 70.4% | 0.11 | 1.97 | 82.1% | 0.10 | 2.21 | 85.7% |

NOTE: *Optimal design gain* refers to the gain in variance for using an optimal design.

nearly equal numbers of respondents were selected in each enumeration area for each specific round, even though some enumeration areas might have more variability than others. If an optimal design were used in which the sample sizes within each enumeration areas were selected to optimize the variability (i.e., find the optimal variance) in an indicator, for mCPR, we found that between a 12 percent to 26 percent reduction in variance could be obtained, improving the accuracy of inferences being made. Across the other indicators, similar improvements in variance could be obtained with an optimal design. This expected improvement should be understood with the caveat that the variances used for such estimation are the ones obtained from the sample, not the true variance at the population level.

### *Frequency of Data Collection*

The ideal frequency of collecting these different indicators is a question of great practical significance for PMA2020. Results from Ghana indicate that mCPR is significantly different at each six-month interval after the first round, with differences ranging between 3.4 and 5.4 percentage points (Tables H.3 and 7.2). The one-year differences and the 18-month differences are also significant, with differences as high as 8.8 and 9.1 percentage points, respectively. Likewise, other indicators, such as current use of any contraceptive method and current use of the pill, vary significantly over six-month periods. These results suggest that PMA2020 seemed to produce an added value every six months that can help shed light on some of the rapid changes occurring in the use of contraceptive methods in Ghana.

### Ethiopia

PMA2020 surveys were fielded in Ethiopia between 2014 and 2016 (Table H.4). The number of respondents in PMA2020's four rounds, between 2014 and 2016, climbed slightly from round to round, from 6,631 in the first round (of a total sample of 28,538) to 7,547 (of a total of 31,563) in Round 4.

#### *Comparison of Ethiopia PMA2020 with the DHS*

The most recent DHS in Ethiopia was conducted in 2010–2011 (USAID, undated). To assess PMA2020 sample representativeness among women in Ethiopia, we compare characteristics of the Ethiopia DHS sample from the survey fielded in 2011 with the nearest two PMA2020s, which were Round 1 (January–March 2014) and Round 2 (October–December 2014). We see no statistically significant differences in age (27.8 and 27.5 for the first two PMA2020 rounds; 27.7 for the DHS).

Like in Ghana, the breakdown of marital status was significantly different between the DHS and PMA2020, though in this case the Ethiopia PMA2020 percentage married was lower and the difference was smaller (less than 3 percentage points, as compared with a 9-percentage-point difference in Ghana; Table H.5). Despite the same average age, however, PMA2020 data showed a much higher rate (almost 5 percentage points) of women who had never married than did the DHS. However, unlike in Ghana, the amount of time between the two surveys' data collection period is large: DHS data collection ended more than three years before PMA2020 data collection began. Therefore, any statistically significant differences between these two surveys' responses may be partially or completely attributable to the populations' behavioral changes over time, rather than attributable to differing survey methods.

With that caveat in mind, we also observed statistically significant differences in education: Almost 15 percent fewer women in PMA2020 had never attended school, while significantly more PMA2020 respondents had completed secondary education—more than 10 percentage points, as well as technical and higher education ($p = 0.000$ for this variable).

These results suggest that the population surveyed by the two surveys (DHS and PMA2020) is not quite the same. Indeed, household-level variables showed large differences, including larger percentages with electricity (a difference of more than 20 percentage points), television, mobile phone, and more among the PMA2020 respondents. However, these differences may reflect, at least to some degree, real changes that occurred during the three-year

**Table H.4**
**Ethiopia PMA2020 Rounds and Sample Size**

| Round | Data Collection Period | Sample Size (number of women) |
|---|---|---|
| 1 | January to March 2014 | 6,631 |
| 2 | October to December 2014 | 6,729 |
| 3 | April to May 2015 | 6,944 |
| 4 | March to April 2016 | 7,547 |

**Table H.5**
**Ethiopia Population Characteristics—DHS and PMA2020**

| Characteristic | Response | DHS (2011) | PMA2020 Round 1 (Jan–March 2014) | PMA2020 Round 2 (Oct–Dec 2014) | PMA2020 Round 4 (Mar–Apr 2016) | *P* Value DHS to Round 1 | *P* Value Across PMA2020 Rounds |
|---|---|---|---|---|---|---|---|
| **Person-level variables (women)** | | | | | | | |
| Age | | 27.68 (27.51, 27.86) | 27.78 (27.40, 28.16) | 27.46 (27.28, 27.63) | 27.85 (27.68, 28.03) | 0.992 | 0.033 |
| Marital status | 1. Currently married | 57.39% | 54.94% | 55.18% | 56.89% | 0.992 | 0.000 |
| | 2. Currently living with partner | 4.40% | 2.14% | 2.28% | 1.70% | | |
| | 3. Divorced or separated | 7.97% | 8.67% | 8.19% | 7.73% | | |
| | 4. Widow or widower | 3.52% | 3.05% | 2.74% | 2.57% | | |
| | 5. Never married | 26.72% | 31.20% | 31.61% | 31.11% | | |
| Highest level of school completed | 0. Never | 50.12% | 35.17% | 33.28% | 32.62% | 0.000 | 0.000 |
| | 1. Primary | 35.47% | 34.22% | 35.35% | 35.97% | | |
| | 2. Secondary | 8.45% | 18.77% | 22.81% | 22.50% | | |
| | 3. Technical | 2.80% | 5.95% | 4.71% | 5.30% | | |
| | 4. Higher | 3.15% | 5.89% | 3.86% | 3.62% | | |
| **Household-level variables** | | | | | | | |
| | Has electricity | 32.15% | 52.75% | 55.01% | 53.78% | 0.000 | 0.000 |
| | Has wall clock | 42.52% | 11.89% | 10.46% | 15.35% | 0.000 | 0.000 |
| | Has television | 18.28% | 27.88% | 29.22% | 28.82% | 0.000 | 0.000 |
| | Has mobile phone | 31.67% | 60.77% | 65.23% | 69.82% | 0.000 | 0.000 |
| | Has landline | 7.85% | 6.88% | 7.09% | 5.64% | 0.011 | 0.000 |
| | Has refrigerator | 7.21% | 9.98% | 10.47% | 10.93% | 0.000 | 0.000 |
| | Has bicycle | 2.74% | 2.35% | 2.42% | 2.10% | 0.086 | 0.000 |
| | Has motorcycle | 0.35% | 1.08% | 0.87% | 1.17% | 0.000 | 0.000 |
| | Has car | 31.67% | 60.77% | 65.23% | 3.37% | 0.078 | 0.000 |

**Table H.5—continued**

| Characteristic | Response | DHS (2011) | PMA2020 Round 1 (Jan–March 2014) | PMA2020 Round 2 (Oct–Dec 2014) | PMA2020 Round 4 (Mar–Apr 2016) | *P* Value DHS to Round 1 | *P* Value Across PMA2020 Rounds |
|---|---|---|---|---|---|---|---|
| Wealth quintile | 1. Lowest quintile | 23.78% | 11.92% | 12.17% | 12.94% | 0.000 | 0.000 |
| | 2. Lower quintile | 15.17% | 15.03% | 14.41% | 13.58% | | |
| | 3. Middle quintile | 14.15% | 12.03% | 13.15% | 13.89% | | |
| | 4. Higher quintile | 14.48% | 15.92% | 17.29% | 17.96% | | |
| | 5. Highest quintile | 32.41% | 45.10% | 42.98% | 41.63% | | |

NOTES: *P* values indicating significant changes across all rounds with $p < 0.05$ are highlighted in pink. To minimize the number of columns in each of these tables comparing the DHS with PMA2020, we present results for the first two rounds of data and the last or most recent round of data only. *P* values were calculated using Tukey-Kramer adjustment to account for multiple comparisons.

interval between the DHS and subsequent PMA2020 survey. The population surveyed by PMA2020 was significantly better off, with over 40 percent in the highest wealth quintile (as opposed to 32 percent in the last DHS) and only about 12 percent in the lowest (compared with 24 percent in the last DHS), which may explain the differences in household appliances. The rates of households owning bicycles and cars, though both very small, were not statistically different between the two surveys.

These results suggest potential population difference between the Ethiopia DHS and the PMA2020 survey; differences in rates of family planning may derive from differences in the underlying populations or may reflect real changes in modern contraceptive use.

### *Comparison Across PMA2020 Rounds*

When comparing different rounds of the PMA2020, the populations were similar in age: Every round had an average age between 27.4 and 27.9 years old (Table H.5). Marital status was similarly stable, and no significant differences were found round to round except between Rounds 3 and 4, where the percentage married went up (from 55.2 percent to 56.9 percent).

Education improved between rounds, with an increase in women who responded that they had some secondary education and a drop in those responding no education or higher education; the percentage change was 4 points or less, but these changes were statistically significant. These changes were statistically significant between Rounds 1 and 2 and between Rounds 2 and 3; Round 4 rates were similar to those in Round 3. This could be due to secular trends and could reflect improvement in education in the country and not differences in sampled population.

Mobile phone ownership increased significantly across the rounds, from 61 percent in Round 1 to 70 percent by Round 4, and this can possibly be explained by the penetration of mobile phone in the country over time. There were some changes in the numbers reporting having electricity, with a drop in Round 3, leading to significant changes between Rounds 2

and 3 and Rounds 3 and 4. Other differences were small and statistically significant only between some rounds.

These differences suggest some slight variation in population sampled across time even within PMA2020, but most characteristics remained stable. This suggests that observed differences in the indicators from round to round cannot be attributed to population differences.

### *Detectable Margin of Error and Precision*

The margin of error for mCPR is quite low in Ethiopia—1.0 percent for each of the four rounds. While the margin of error did not change much between rounds, the coefficients of variation did fluctuate, with higher numbers in Round 2 (2.17 for mCPR) decreasing to 1.92 in the last round, but still not as low as the 1.04 in Round 1.

As in Ghana and other countries, the variation can be partly explained by the varying scales for some of these outcomes; for example, using only traditional contraceptive methods is rarer than use of modern methods.

### *Design Effect*

Across all indicators, very small ICCs were observed (Table H.6). All ICCs were less than 0.1, and the ones for traditional contraception were even lower. Again, though ICC did not vary much, there was variation in design effect between indicators. This variation in design effect is because of a combination of the number of observations in enumeration areas and the small variations in the ICC, which reflects homogeneity within the enumeration areas.

The impact of optimal design was much smaller in Ethiopia than in Ghana: With an optimal design used for mCPR, only a 7- to 8-percent reduction in variance could be obtained,

**Table H.6**
**Ethiopia: Design Effect and Optimal Design Gain—Round-to-Round Comparisons**

| | Round 1 | | | Round 2 | | | Round 3 | | | Round 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain |
| Current use of any contraceptive method | 0.088 | 3.81 | 92.0% | 0.075 | 3.40 | 92.3% | 0.087 | 3.86 | 93.7% | 0.088 | 3.89 | 92.9% |
| Current use of any modern contraceptive method | 0.088 | 3.83 | 91.9% | 0.077 | 3.46 | 92.1% | 0.087 | 3.89 | 93.3% | 0.091 | 4.00 | 92.3% |
| Current use of any traditional contraceptive method | 0.025 | 1.80 | 89.0% | 0.008 | 1.25 | 92.8% | 0.019 | 1.61 | 92.8% | 0.027 | 1.90 | 90.9% |
| Total unmet need (spacing and limiting) | 0.078 | 3.50 | 84.7% | 0.060 | 2.97 | 86.6% | 0.074 | 3.51 | 86.0% | 0.093 | 4.06 | 85.2% |
| Current use of the pill | 0.066 | 1.52 | 83.4% | 0.040 | 1.32 | 85.6% | 0.069 | 1.62 | 84.5% | 0.078 | 1.70 | 84.8% |

NOTE: *Optimal design gain* refers to the gain in variance for using an optimal design.

which would improve, only slightly, the accuracy of inferences being made. Greater improvements could be seen for the indicator of total unmet need (14–16 percent) and for indicators for specific methods (14–17 percent improvement with optimal design for an indicator regarding use of the pill, and much higher for less–frequently used methods), but for the indicators for rates of use, the gain with optimal design was not more than 10 percent.

Again, any expected improvement should be understood with the caveat that the variances used for such estimation are the ones obtained from the sample, not the true variance at the population level, since such population-level variance is not available.

#### *Frequency of Data Collection*

In Ethiopia, the data collection every six months did not add meaningful information in many cases (Table H.5). For example, for mCPR and total contraceptive prevalence, the differences between Round 2 and 3 are statistically significant, but not those between Rounds 1 and 2 or between Rounds 3 and 4; the one-year and 18-month differences were statistically significant. While this is different than what we found in Ghana, this pattern indicated that moving to a yearly survey would be sufficient to capture meaningful changes in family planning indicators, provided that these changes continue to occur at the rates observed during the past few years.

### Democratic Republic of the Congo

The PMA2020 survey conducted in DRC began as a local survey in Kinshasa and expanded to Kongo Central in Round 4. To enable comparison with DHS data and comparison over time, only Kinshasa data are presented here, even though more data are available in later rounds. For this same reason, wealth quintile data were not included, since the quintiles in the DHS were determined nationally. Also, we note that DRC is the first country for which we have five rounds of data, but enumeration areas were not changed even for the fifth round because of the large size of the enumeration area. Therefore, we were unable to look at the question of the impact of changing enumeration areas on the estimates. Dates of each round are shown in Table H.7.

#### *DHS and DRC—Kinshasa PMA2020*

DHS data in DRC were collected between August 2013 and February 2014. The average age of the women included in the first round of PMA2020 (27.4) was lower than the DHS estimate (28.1), but this difference was not significant (Table H.8). The percentage of women

**Table H.7**
**DRC—Kinshasa PMA2020 Rounds and Sample Size**

| Round | Data Collection Period | Sample Size (number of women) |
|---|---|---|
| 1 | October 2013 to January 2014 | 2,136 |
| 2 | August to September 2014 | 2,881 |
| 3 | May to June 2015 | 2,693 |
| 4 | October 2015 to January 2016 | 2,780 |
| 5 | July to September 2016 | 2,609 |

**Table H.8**
**DRC—Kinshasa Population Characteristics—DHS and PMA2020**

| Characteristic | Response | DHS (Aug 2013–Feb 2014) | PMA2020 Round 1 (Oct 2013–Jan 2014) | PMA2020 Round 2 (Aug–Sep 2014) | PMA2020 Round 5 (Jul–Sep 2016) | *P* Value DHS to Round 1 | *P* Value Across PMA2020 Rounds |
|---|---|---|---|---|---|---|---|
| **Person-level variables (women)** | | | | | | | |
| Age | | 28.20 (27.25, 29.15) | 27.38 (27.05, 27.71) | 28.03 (27.70, 28.37) | 28.10 (27.76, 28.43) | 0.600 | 0.034 |
| Marital status | 1. Currently married | 26.55% | 29.12% | 28.66% | 28.71% | 0.000 | 0.000 |
| | 2. Currently living with partner | 16.80% | 18.26% | 18.03% | 16.56% | | |
| | 3. Divorced or separated | 8.48% | 1.73% | 3.72% | 3.45% | | |
| | 4. Widow or widower | 2.16% | 1.73% | 1.88% | 1.34% | | |
| | 5. Never married | 46.01% | 49.16% | 47.72% | 49.94% | | |
| Highest level of school completed | No education | 0.72% | 5.82% | 1.39% | 2.15% | 0.000 | 0.000 |
| | Primary | 9.81% | 47.46% | 9.46% | 18.00% | | |
| | Secondary | 73.45% | 39.44% | 73.16% | 64.04% | | |
| | Higher/ tertiary | 16.02% | 7.28% | 15.99% | 15.81% | | |
| **Household-level variables** | | | | | | | |
| | Has electricity | 88.30% | 79.59% | 79.84% | 74.20% | 0.000 | 0.000 |
| | Has wall clock | 75.45% | 63.70% | 66.00% | 63.04% | 0.000 | 0.000 |
| | Has television; in Rounds 1 and 2, combined color and black-and-white television | 81.55% | 74.69% | 81.26% | 80.22% | 0.000 | 0.000 |
| | Has mobile phone | 94.48% | 90.92% | 84.63% | 84.39% | 0.000 | 0.000 |
| | Has landline | 2.31% | 2.25% | 3.42% | 1.95% | 0.926 | 0.022 |
| | Has refrigerator | 32.48% | 9.30% | 9.21% | 9.92% | 0.000 | 0.000 |
| | Has bicycle | 3.95% | 2.59% | 2.37% | 1.19% | 0.036 | 0.000 |
| | Has motorcycle | 2.22% | 3.04% | 2.84% | 1.52% | 0.175 | 0.011 |
| | Has car | 9.06% | 5.92% | 7.16% | 5.96% | 0.001 | 0.001 |

**Table H.8—continued**

| Characteristic | Response | DHS (Aug 2013–Feb 2014) | PMA2020 Round 1 (Oct 2013–Jan 2014) | PMA2020 Round 2 (Aug–Sep 2014) | PMA2020 Round 5 (Jul–Sep 2016) | *P* Value DHS to Round 1 | *P* Value Across PMA2020 Rounds |
|---|---|---|---|---|---|---|---|
| Wealth quintile (round-to-round comparison only) | 1. Lowest quintile | | 19.90% | 19.05% | 19.39% | | 0.000 |
| | 2. Lower quintile | | 20.57% | 18.89% | 20.75% | | |
| | 3. Middle quintile | | 21.08% | 20.68% | 20.48% | | |
| | 4. Higher quintile | | 21.93% | 21.53% | 20.70% | | |
| | 5. Highest quintile | | 16.52% | 19.84% | 18.69% | | |

NOTES: *P* values indicating significant changes across all rounds with $p < 0.05$ are highlighted in pink. To minimize the number of columns in each of these tables comparing the DHS with PMA2020, we present results for the first two rounds of data and the last or most recent round of data only. *P* values were calculated using Tukey-Kramer adjustment to account for multiple comparisons.

who were married differed significantly, with 29.1 percent of PMA2020 respondents currently married compared with 26.6 percent of DHS respondents. At the same time, 49.2 percent of PMA2020 respondents were never married, compared with 46.0 percent of DHS respondents. PMA2020 found much higher rates of secondary education, with 7.3 percent (Round 1, $p < 0.001$), compared with DHS's 16.0 percent, though the differences disappeared when comparing DHS with Round 2 PMA2020 data (16.0 percent, $p = 0.207$).

Other household variables also showed significant differences: PMA2020 respondents were less likely to say they had electricity, less likely to say they had a mobile phone, and much less likely to report having a refrigerator (all $p < 0.001$), even though the dates of the surveys overlapped.

Wealth quintile was not included in this comparison because it was calculated country-wide in the DHS data.

This suggests that the PMA2020 sample in DRC is not completely similar to the population that was surveyed in the DHS; thus, differences in rates of family planning uptake may derive from differences in the underlying populations.

### *Comparison Across PMA2020 Rounds*

There were large changes in household-level variables between Rounds 2, 3, and 4 attesting to access to electricity (79.9 percent to 67.2 percent to 73.3 percent). The percentage reporting having a television was 75 percent in Round 1, which increased to 81.2 in Round 2 and stayed stable over the following three rounds. The percentage reporting access to mobile phone dropped significantly between Rounds 1 and 2 and then remained relatively stable. Other variables had slight changes, with a few statistically significant differences, but most round-to-round changes were less than 1 percentage point different.

The only significant difference between Rounds 4 and 5 was in the level of education, with more women reporting higher education completed, comparable to Round 2, and signifi-

cantly higher for Rounds 3 and 4. Otherwise, all of the demographic variables were statistically equivalent when comparing Round 4 and Round 5, including distribution across wealth quintiles, which changed very little between rounds (round-to-round differences were not statistically significant, though the overall *p* value for a comparison across all rounds was).

### *Detectable Margin of Error and Precision*

The margin of error for mCPR is low in DRC, with estimates ranging from 1.3 percent to 1.6 percent across the rounds. In addition, for all indicators assessed, they remained below the 3 percent target set for margin of error for mCPR. The CoV, which standardizes the variance, is higher in Round 1, with a slight decrease in variation in all indicators by Round 5.

### *Design Effect*

We next looked at the ICCs and design effects in the PMA2020 survey for DRC—Kinshasa (Table H.9). In Round 1, the ICCs across all indicators were very small and less than 0.01, but for all other rounds they were slightly larger, even though still smaller than 0.1. In Round 1, the design effect was close to 1, meaning that the effective sample size is close to the actual

**Table H.9**
**DRC—Kinshasa: Design Effect and Optimal Design Gain**

| | Round 1 | | | Round 2 | | | Round 3 | | | Round 4 | | | Round 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Indicator | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain |
| Current use of any contraceptive method | 0.006 | 1.20 | 61.3% | 0.106 | 6.09 | 95.3% | 0.101 | 5.52 | 94.3% | 0.063 | 3.88 | 95.8% | 0.069 | 4.03 | 95.3% |
| Current use of any modern contraceptive method | 0.007 | 1.24 | 56.2% | 0.057 | 3.72 | 94.9% | 0.084 | 4.78 | 93.0% | 0.055 | 3.53 | 92.4% | 0.047 | 3.07 | 91.7% |
| Current use of any traditional contraceptive method | 0.014 | 1.49 | 63.9% | 0.091 | 5.36 | 86.6% | 0.053 | 3.38 | 91.0% | 0.046 | 3.10 | 93.7% | 0.041 | 2.82 | 93.1% |
| Total unmet need (spacing and limiting) | 0.007 | 1.25 | 55.6% | 0.045 | 3.21 | 93.1% | 0.054 | 3.42 | 91.7% | 0.072 | 4.31 | 89.1% | 0.078 | 4.42 | 89.2% |
| Current use of the pill | 0.009 | 1.09 | 74.9% | 0.019 | 1.91 | 89.7% | 0.022 | 1.98 | 92.0% | 0.020 | 1.95 | 87.4% | 0.016 | 1.71 | 90.8% |

sample size. However, the design effect is much higher in Round 2 and Round 3, meaning that there is a lot of similarity within clusters and a much lower effective sample size. With such a large design effect, it might be more efficient to increase the number of enumeration areas while decreasing the number of observations collected within enumeration areas. Still, even with large design effects, the margin of error for the different indicators is still small. The potential gain from optimal design was highest in Round 1, which is indicative of different precisions within enumeration areas in that round, while in later rounds little potential gain was seen.

#### *Frequency of Data Collection*

Because we have data for five rounds of surveys, the data from DRC offered us the opportunity to look at the largest number of changes between rounds. While all of the major indicators had significant overall change (i.e., between Round 1 and Round 5), few had statistically significant changes every six months. mCPR only had a significant change between Rounds 3 and 4; no other six-month change was significant. Even the change over the first year (Round 1 versus Round 3) was not statistically significant. However, the 12-month changes over the second year and into the third—that is, Rounds 2 to 4 and Rounds 3 to 5—were statistically significant for all of the major indicators we examined. Similarly, the differences over 18 months were mostly significant, although unmet need did not change significantly between Rounds 1 and 4; when comparing Round 1 to Round 5, all indicators did change significantly.

### Nigeria (Kaduna and Lagos only)

PMA2020 collected three rounds of data in Nigeria between 2014 and 2016 (Table H.10). The first two were only conducted in Kaduna and Lagos. The third round used a national sample but focused on urban areas. We used DHS data only for Kaduna and Lagos to make comparisons to Rounds 1 and 2 of PMA2020 data and limited PMA2020 data to these two states as well to enable comparisons.

The most recent census in Nigeria was in the early 1980s. The Gates Institute recognized the limitation of using quite outdated census data for PMA2020's sampling frames in Nigeria and noted that it is also the one used by the DHS, so to the extent that a goal is comparability to the DHS, the sampling design relies on the same census data. We acknowledge this limitation and examine the available data to explore representativeness, margins of error, design effects, and frequency.

**Table H.10**
**Nigeria PMA2020 Rounds and Sample Size**

| Round | Data Collection Period | Sample Size (number of women) |
|---|---|---|
| 1 (Kaduna and Lagos only) | September to October 2014 | 3,380 |
| 2 (Kaduna and Lagos only) | August to September 2015 | 4,421 |
| 3 (National) | May to June 2016 | 11,209 (4,323 in Kaduna and Lagos) |

### *DHS and Nigeria PMA2020*

The average respondent age in the first round of PMA2020 data is 27.3, statistically significantly lower than the DHS average of 28.9 (Table H.11). There are also significant differences in marital status, with more PMA2020 survey respondents reporting that they are currently married (76.2 percent versus 64.8 percent) and fewer reporting that they were never married compared with the DHS. The education results were also significantly different, with 5 percentage points fewer in PMA2020 reporting that they had never attended school and almost 5 percentage points more reporting that they had completed higher education. The households also varied significantly for all compared indicators. PMA2020 respondents were 13 percentage points less likely to have a mobile phone than DHS respondents (75.0 percent versus 87.9 percent). As with DRC above, wealth was not included in this comparison because it was calculated country-wide in the DHS data.

### *Comparison Across PMA2020 Rounds*

Three rounds of PMA2020 data are available for Nigeria. Our analysis here is limited to data from Kaduna and Lagos only.

The rounds differed slightly on average age, with the difference of 1.2 years between Rounds 1 and 2 being statistically significant.

The percentage of households reporting having electricity increased in Round 2 (68.1 percent) and dropped back down in Round 3 (63.0 percent). The percentage of mobile phone ownership remained stable between Rounds 1 (75.0 percent) and 2 (74.0 percent) but increased in Round 3 (76.1 percent), and the percentage of refrigerator ownership and car ownership fluctuated by a number of percentage points (8 and 3, respectively).

### *Detectable Margin of Error and Precision*

The margins of error for mCPR were small, far below the 3-percent maximum target for PMA2020: 1.0 percent in Round 1 and 1.1 percent in Rounds 2 and 3 for mCPR. For all other indicators, the margin of error was small, below the 3-percent target except for the use of the pill in Round 1 (4.3 percent). An examination of the data showed that in Round 1, there was a large amount of missing data for the question about "use of the pill" (only a 10.7-percent response rate in Round 1, compared with more than 99 percent in later rounds), which explains the lack of precision for that indicator in the round. This indicates that inferences can be made about all the indicators in PMA2020 with good precision even though mCPR has been the main focus of the survey design.

### *Design Effect*

Very small ICCs were observed across all indicators, with most of them less than 0.1 (with the exception of pill use in Round 1, which we already noted had a lot of missing data) (Table H.12).

The design effect for mCPR was large (more than 3.5) and increased over the three rounds, as it did for most of the other indicators measured. Even with such large design effects, very small margins of error were estimated, which is indicative of good precision. However, the design showed potential for improvement with optimal design—that is, changing the number of individuals in each cluster could improve the efficiency of the survey in Nigeria by about 20 percent for most indicators.

**Table H.11**
**Nigeria Population Characteristics—DHS and PMA2020 (Kaduna and Lagos only)**

| Characteristic | Response | DHS (Feb–June 2013) | PMA2020 Round 1 (Sep–Oct 2014) | PMA2020 Round 2 (Aug–Sep 2015) | PMA2020 Round 3 (May–June 2016) | *P* Value DHS to Round 1 | *P* Value Across PMA2020 Rounds |
|---|---|---|---|---|---|---|---|
| **Person-level variables (women)** | | | | | | | |
| Age | | 28.91 (28.62, 29.20) | 27.31 (27.16, 27.46) | 28.55 (28.39, 28.70) | 28.56 (28.26, 28.85) | 0.000 | 0.000 |
| Marital status | 1. Currently married | 64.84% | 76.16% | 73.57% | 71.57 | 0.000 | 0.000 |
| | 2. Currently living with partner | 3.45% | 1.21% | 1.29% | 1.23 | | |
| | 3. Divorced or separated | 2.42% | 1.39% | 1.74% | 1.58 | | |
| | 4. Widow or widower | 2.17% | 1.42% | 1.74% | 1.3 | | |
| | 5. Never married | 27.12% | 19.82% | 21.65% | 24.31 | | |
| Highest level of school completed | 0. Never attended | 21.69% | 26.96% | 25.73% | 24.38 | 0.000 | 0.000 |
| | 1. Primary | 14.83% | 21.53% | 21.05% | 19.4 | | |
| | 2. Secondary | 47.78% | 38.70% | 38.93% | 40.5 | | |
| | 3. Higher | 15.71% | 12.81% | 14.30% | 15.71 | | |
| **Household-level variables** | | | | | | | |
| | Has electricity | 81.88% | 59.13% | 68.13% | 62.96% | 0.000 | 0.000 |
| | Has wall clock | 82.60% | 73.22% | 73.13% | 72.79% | 0.000 | 0.000 |
| | Has radio | 67.58% | 71.00% | 66.05% | 66.29% | 0.000 | 0.000 |
| | Has black and white television | 71.67% | 3.45% | 3.54% | 3.07% | 0.000 | 0.000 |
| | Has color television | 71.67% | 53.91% | 59.34% | 57.27% | 0.000 | 0.000 |
| | Has mobile phone | 87.90% | 74.96% | 73.99% | 76.11% | 0.000 | 0.000 |
| | Has landline | 1.77% | 1.01% | 1.48% | 1.79% | 0.012 | 0.037 |
| | Has refrigerator | 36.57% | 24.72% | 32.52% | 30.76% | 0.000 | 0.000 |
| | Has bicycle | 8.77% | 16.52% | 15.47% | 16.25% | 0.000 | 0.001 |
| | Has motorcycle | 19.26% | 21.78% | 21.50% | 20.94% | 0.017 | 0.077 |
| | Has car | 12.75% | 10.89% | 14.28% | 12.87% | 0.026 | 0.000 |
| Wealth quintile (round-to-round comparison only) | 1. Lowest quintile | | 19.57% | 20.04% | 18.67% | | 0.000 |
| | 2. Lower quintile | | 19.95% | 19.70% | 17.69% | | |
| | 3. Middle quintile | | 20.33% | 19.38% | 17.96% | | |
| | 4. Higher quintile | | 20.58% | 20.29% | 22.85% | | |
| | 5. Highest quintile | | 19.57% | 20.59% | 22.83% | | |

NOTES: *P* values indicating significant changes across all rounds with $p < 0.05$ are highlighted in pink. *P* values were calculated using Tukey-Kramer adjustment to account for multiple comparisons.

**Table H.12**
**Nigeria—Kaduna and Lagos: Design Effect and Optimal Design Gain**

| | Round 1 | | | Round 2 | | | Round 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Indicator | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain |
| Current use of any contraceptive method | 0.090 | 3.78 | 72.2% | 0.127 | 5.58 | 79.4% | 0.131 | 5.71 | 79.0% |
| Current use of any modern contraceptive method | 0.086 | 3.65 | 72.9% | 0.100 | 4.61 | 80.1% | 0.111 | 5.00 | 78.5% |
| Current use of any traditional contraceptive method | 0.012 | 1.36 | 78.3% | 0.063 | 3.26 | 74.5% | 0.052 | 2.87 | 79.7% |
| Total unmet need (spacing and limiting) | 0.057 | 2.78 | 82.4% | 0.055 | 2.99 | 88.2% | 0.084 | 4.02 | 79.3% |
| Current use of pill (A lot of missing data in Round 1) | 0.214 | 1.64 | 52.7% | 0.041 | 2.47 | 82.0% | 0.031 | 2.10 | 83.7% |

### *Frequency of Data Collection*

With only three rounds of data, we have limited comparisons about the ideal frequency of data collection (only six months and one year). However, we did see significant changes in the indicators of interest between Rounds 1 and 2 ($p < 0.001$) but not from Round 2 to Round 3 on any indicators. The differences from Round 1 to Round 3, or one-year differences, were all statistically significant ($p < 0.001$). This provides an inconclusive result, as the observed difference between Rounds 3 and 1 seemed to be driven by the observed differences between Round 2 versus Round 1. If Round 1 turned out to be an outlier (e.g., it was the first implementation in Nigeria, and so implementation issues could have caused a biased sample in Round 1, as seen in the round-to-round representativeness), then the Round 2 to Round 3 comparisons will seem to be the only viable result. All the different issues observed in the Nigeria data argue further for the need for consistent data collection if one of the goals of the PMA2020 is to assess changes happening in family planning over time.

## Kenya

Four rounds of PMA2020 data were collected in Kenya between 2014 and 2015 (Table H.13). The DHS in Kenya was last fielded in 2014.

### *DHS and Kenya PMA2020*

The average ages in the PMA2020 Round 1 (28.8) and Round 2 (28.7) surveys are not significantly different from that in the DHS (28.9, $p > 0.05$; Table H.14). Data collection for both Rounds 1 and 2 of the PMA2020 surveys occurred in 2014, the same year as the DHS. The timing of Round 1 data collection completely overlaps with the timing of the DHS's data collection, so any difference in the survey results should be mostly attributable to differences

**Table H.13**
**Kenya PMA2020 Rounds and Sample Size**

| Round | Data Collection Period | Sample Size (number of women) |
|---|---|---|
| 1 | May to July 2014 | 3,826 |
| 2 | November to December 2014 | 4,386 |
| 3 | June to July 2015 | 4,440 |
| 4 | November to December 2015 | 4,968 |

in the survey designs, not attributable to changes in the populations' behavior over time. The marriage data differ slightly between the two sources; the DHS's currently married percentage was 57.1 percent, higher than PMA2020's 56.0 percent in Round 1 and 53.3 percent in Round 2. These differences in marital status are practically small, though they showed statistical significance ($p < 0.001$). Education levels were also slightly different in the PMA2020 data, though the biggest difference was 3 percentage points, with fewer PMA2020 respondents finishing only primary school and almost 2 percentage points more PMA2020 respondents finishing college.

Household variables also showed some differences: The percentage of mobile phone ownership was 76.1 percent in PMA2020, statistically significantly lower than the 81.6 percent found in the DHS ($p < 0.001$). Electricity access, in contrast, was 4 percentage points higher in the PMA2020 data. Household socioeconomic status also seemed to vary, with some differences between the DHS and the two PMA2020 rounds found in the wealth of the households.

#### *Comparison Across PMA2020 Rounds*

Consecutive-round differences were not statistically significant for average age until the step from Round 3 to Round 4; here the average age went down by slightly less than a year, from 28.8 to 28.0 ($p < 0.001$). The percentage reporting being currently married varied somewhat, with a large drop between Rounds 3 and 4 (from 56.2 percent to 49.9 percent, $p < 0.001$). Distribution of education levels stayed relatively stable from Round 2 on. The percentage of respondents reporting having access to electricity did not change significantly in the first few rounds but did drop by almost 4 percentage points between Rounds 3 and 4. Mobile phone rates increased between 2 and 5 percentage points between each round, although this is by nature a quickly changing variable, so it may reflect an actual increase in access.

These findings raise some questions about the consistency of the representativeness of the PMA2020 sample from round to round, particularly between Rounds 3 and 4.

#### *Detectable Margin of Error and Precision*

The margin of error for all indicators considered for Kenya was well below the 3 percent maximum targeted by PMA2020, including for specific contraceptive methods. The margin of error was about 1.5 percent for mCPR across the different rounds and similar for other indicators, except for the traditional method use, where the margins of error were less than 0.4. The use of traditional methods was rare (less than 2 percent), with a not-so-small standard deviation;

**Table H.14**
**Kenya Population Characteristics—DHS and PMA2020**

| Characteristic | Response | DHS (May–Oct 2014) | PMA2020 Round 1 (May–July 2014) | PMA2020 Round 2 (Nov–Dec 2014) | PMA2020 Round 4 (Nov–Dec 2015) | *P* Value DHS to Round 1 | *P* Value Across PMA2020 Rounds |
|---|---|---|---|---|---|---|---|
| **Person-level variables (women)** | | | | | | | |
| Age | | 28.88 (28.70, 29.06) | 28.76 (28.58, 28.94) | 28.72 (28.54, 28.90) | 28.04 (27.87, 28.22) | 0.898 | 0.000 |
| Marital status | 1. Currently married | 57.12% | 55.98% | 53.30% | 49.86% | 0.000 | 0.000 |
| | 2. Currently living with partner | 4.13% | 10.71% | 8.60% | 7.85% | | |
| | 3. Divorced or separated | 7.33% | 4.53% | 4.54% | 4.55% | | |
| | 4. Widow or widower | 3.83% | 3.12% | 2.92% | 2.92% | | |
| | 5. Never married | 27.59% | 25.66% | 30.65% | 34.82% | | |
| Highest level of school completed | 1. Primary | 56.80% | 55.47% | 53.75% | 52.48% | | |
| | 2. Postprimary/ vocational | 1.25% | 2.61% | 1.73% | 2.24% | | |
| | 3. Secondary/ A-level | 31.96% | 29.97% | 32.94% | 31.83% | | |
| | 4. College | 7.49% | 9.09% | 9.20% | 10.23% | | |
| | 5. University | 2.51% | 2.86% | 2.39% | 3.22% | | |
| **Household-level variables** | | | | | | | |
| | Has electricity | 26.92% | 30.69% | 32.45% | 36.08% | 0.000 | 0.000 |
| | Has television; in Rounds 1 and 2, combined color and black and white television | 27.37% | | 30.21% | 32.43% | | 0.000 |
| | Has mobile phone | 81.63% | 76.11% | 78.11% | 85.43% | 0.000 | 0.000 |
| | Has landline | 0.43% | 0.53% | 0.43% | 1.06% | 0.355 | 0.000 |
| | Has refrigerator | 4.64% | 4.23% | 4.60% | 5.91% | 0.216 | 0.000 |
| | Has bicycle | 20.18% | 18.59% | 19.05% | 18.34% | 0.012 | 0.002 |
| | Has motorcycle | 7.59% | 4.43% | 5.76% | 6.11% | 0.000 | 0.000 |
| | Has car | 3.97% | 3.26% | 3.28% | 4.07% | 0.019 | 0.014 |

**Table H.14—continued**

| Characteristic | Response | DHS (May–Oct 2014) | PMA2020 Round 1 (May–July 2014) | PMA2020 Round 2 (Nov–Dec 2014) | PMA2020 Round 4 (Nov–Dec 2015) | *P* Value DHS to Round 1 | *P* Value Across PMA2020 Rounds |
|---|---|---|---|---|---|---|---|
| Wealth quintile | 1. Lowest quintile | 25.02% | 34.67% | 21.92% | 21.85% | 0.000 | 0.000 |
| | 2. Lower quintile | 19.20% | 26.06% | 22.74% | 22.14% | | |
| | 3. Middle quintile | 18.80% | 18.20% | 21.52% | 20.28% | | |
| | 4. Higher quintile | 19.95% | 12.88% | 19.11% | 20.22% | | |
| | 5. Highest quintile | 17.04% | 8.19% | 14.70% | 15.51% | | |

NOTES: *P* values indicating significant changes across all rounds with $p < 0.05$ are highlighted in pink. To minimize the number of columns in each of these tables comparing the DHS with PMA2020, we present results for the first two rounds of data and the last or most recent round of data only. *P* values were calculated using Tukey-Kramer adjustment to account for multiple comparisons.

this led to a very large CoV for that indicator. Most of the other indicators have small CoVs of 2 or less.

### *Design Effect*

ICCs were very small across all of the indicators in each round in Kenya (mostly around 0.05; Table H.15). The design effect also was relatively small across rounds, even though some of them were three or more. When it comes to the potential for improvement using optimal design, very limited gain will be expected.

### *Frequency of Data Collection*

Of the three six-month differences, each indicator only showed one significant change. However, when looking at one-year differences, all changes were statistically significant, where, for example, a 5- to 6-percent increase in mCPR rate was observed. The 18-month changes were also significant, with all changes going in expected directions.

**Table H.15**
**Kenya: Design Effect and Optimal Design Gain**

| | Round 1 | | | Round 2 | | | Round 3 | | | Round 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain | ICC | Design Effect | Optimal Design Gain |
| Current use of any contraceptive method | 0.054 | 2.62 | 91.4% | 0.063 | 3.20 | 91.1% | 0.065 | 3.35 | 92.9% | 0.061 | 3.46 | 91.8% |
| Current use of any modern contraceptive method | 0.053 | 2.59 | 91.3% | 0.065 | 3.28 | 91.2% | 0.062 | 3.25 | 92.9% | 0.069 | 3.75 | 91.5% |
| Current use of any traditional contraceptive method | 0.003 | 1.09 | 90.1% | 0.044 | 2.54 | 85.5% | 0.015 | 1.53 | 91.6% | 0.022 | 1.87 | 72.4% |
| Total unmet need (spacing and limiting) | 0.025 | 1.77 | 91.7% | 0.052 | 2.81 | 90.3% | 0.033 | 2.18 | 89.8% | 0.040 | 2.60 | 86.7% |
| Current use of the pill | 0.025 | 1.77 | 77.2% | 0.052 | 1.72 | 87.9% | 0.095 | 2.53 | 85.7% | 0.137 | 6.48 | 74.9% |

# References


Alexander, Charles H., "Still Rolling: Leslie Kish's 'Rolling Samples' and the American Community Survey," *Survey Methodology*, Vol. 28, No. 1, 2002, pp. 35–42.

Aliaga, Alfredo, and Ruilin Ren, *Optimal Sample Sizes for Two-Stage Cluster Sampling in Demographic and Health Surveys*, Washington, D.C.: United States Agency for International Development, No. 30, 2006. As of November 9, 2017:
http://dhsprogram.com/pubs/pdf/WP30/WP30.pdf

AvenirHealth, "FP2020 Core Indicators," 2017. As of August 4, 2017:
http://www.track20.org/pages/data_analysis/core_indicators/StatTrack.php

Bond, Lawrance W., William E. Bertrand, and Robertino Mera, *Data for Decision Making for the Health Sector Project: A Mid-Term Evaluation*, Office of Health and Nutrition, Bureau for Global Programs, Field Support and Research, Washington, D.C.: United States Agency for International Development, 1994. As of November 9, 2017:
http://pdf.usaid.gov/pdf_docs/PDABK343.pdf

Brand, Joel, "First Ladies of Africa Partner with RAND, U.S. Dept. of State and Corporate Council on Africa to Advance Women's Leadership and Economic Empowerment in Africa," Santa Monica, Calif.: RAND Corporation, September 23, 2011. As of November 9, 2017:
https://www.rand.org/news/press/2011/09/23.html

Bullen, Piroska Bisits, "Theory of Change vs Logical Framework—What's the Difference?" *Tools4dev*, April 11, 2013. As of November 9, 2017:
http://www.tools4dev.org/resources/theory-of-change-vs-logical-framework-whats-the-difference-in-practice

CDC—*See* Centers for Disease Control and Prevention.

Center for Theory of Change, "What Is Theory of Change," undated. As of November 9, 2017:
http://www.theoryofchange.org/what-is-theory-of-change

Centers for Disease Control and Prevention, Division of Global Health Protection, "Field Epidemiology Training Program: Where We Work," May 12, 2017. As of August 1, 2017:
https://www.cdc.gov/globalhealth/healthprotection/fetp/where/index.html

Cochran, W. G., and G. William, *Sampling Techniques*, New York: John Wiley & Sons, Inc, 1977.

Corsi, D. J., M. Neuman, J. E. Finlay, and S. V. Subramanian, "Demographic and Health Surveys: A Profile," *International Journal of Epidemiology*, Vol. 41, No. 6, December 2012, pp. 1602–1613. As of November 9, 2017:
https://www.ncbi.nlm.nih.gov/pubmed/23148108

Demographic and Health Survey, "Continuous Demographic and Health Survey," Washington, D.C.: USAID, undated. As of November 9, 2017:
https://dhsprogram.com/pubs/pdf/DM34/DM34.pdf

DHIS2—*See* District Health Information Software, version 2.

DHS—*See* Demographic and Health Survey.

DHS Program, "DHS Overview," undated. As of November 9, 2017:
http://dhsprogram.com/What-We-Do/Survey-Types/DHS.cfm

DHS Program STATcompiler, home page, undated. As of November 9, 2017:
http://statcompiler.com/en/

District Health Information Software, version 2, homepage, undated(a). As of November 9, 2017:
https://www.dhis2.org/

District Health Information Software, version 2, "DHIS 2 in Action," undated(b). As of November 9, 2017:
https://www.dhis2.org/inaction

Family Planning 2020, " About Us," 2017. As of November 9, 2017:
http://www.familyplanning2020.org/microsite/about-us

Fowler, Floyd J., Jr., *Survey Research Methods*, 5th ed., Sage Publications, Inc., 2013.

Frenk, Julio, "The Global Health System: Strengthening National Health Systems as the Next Step for Global Progress," *PLoS Medicine,* Vol. 7, No. 1, 2010, p. e1000089.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch, *Impact Evaluation in Practice*, Washington, D.C.: World Bank, 2011. As of November 9, 2017:
http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf

Greenfield, Victoria, Valerie L. Williams, and Elisa Eiseman, *Using Logic Models for Strategic Planning and Evaluation: Application to the National Center for Injury Prevention and Control*, Santa Monica, Calif.: RAND Corporation, TR-370-NCIPC, 2006. As of December 14, 2017:
https://www.rand.org/pubs/technical_reports/TR370.html

Haggard, Lois M., and Shandra J. Burnett, "Measuring the Impact of a Web-Based Data Query System: The Logic Model as a Tool in the Evaluation Process," *Journal of Public Health Management and Practice*, Vol. 12, No. 2, 2006, pp. 189–195.

Hawes, Meagan, Sally Safi, Abigail Greenleaf, Amy Tsui, and the PMA2020 Study Group, *PMA2020 Methodological Report No. 1: Response Patterns on Behavioral Outcomes in Relation to Use of Resident Enumerators Over Multiple Survey Rounds*, PMA2020, 2017. As of November 9, 2017:
https://pma2020.org/sites/default/files/PMA2020 Methodological Report_v12-2017-05-08-MMH-sed-ep.pdf

INFORMS, *Informs Analytics Maturity Model User Guide*, Catonsville, Md., 2017.

International Telecommunication Union, World Telecommunication/ICT Development Report and database, 2017. As of November 9, 2017:
https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx

International Union for the Scientific Study of Population, *The Demographic and Health Surveys (DHS) Project Past, Present and Future*, 2009. As of January 18, 2018:
http://iussp2009.princeton.edu/papers/93520

IPUMS-DHS, "Health Data on Women and Children in the Developing World," undated. As of November 9, 2017:
https://www.idhsdata.org/idhs/

Iwelunmor, Juliet, Sarah Blackstone, Dorice Veira, Ucheoma Nwaozuru, Collins Airhihenbuwa, Davison Munodawafa, Ezekiel Kalipeni, Antar Jutal, Donna Shelley, and Gbenga Ogedegbe, "Toward the Sustainability of Health Interventions Implemented in Sub-Saharan Africa: A Systematic Review and Conceptual Framework," *Implementation Science*, Vol. 11, No. 1, 2016, p. 43.

Jacucci, Edoardo, Vincent Shaw, and Jørn Braa, "Standardization of Health Information Systems in South Africa: The Challenge of Local Sustainability," *Information Technology for Development*, Vol. 12, No. 3, 2006, pp. 225–239.

Jamison, Dean T., Joel G. Breman, Anthony R. Measham, George Alleyne, Mariam Claeson, David B. Evans, Prabhat Jha, Anne Mills, and Philip Musgrove, eds., *Disease Control Priorities in Developing Countries*, 2nd ed., Washington, D.C.: World Bank, 2006.

Kaneda, Toshiko, and Genevieve Dupuis, *2017 World Population Data Sheet*, Washington, D.C.: Population Reference Bureau, August 2017. As of November 9, 2017:
http://www.prb.org/pdf17/2017_World_Population.pdf

Keating, Caitlin, "Melinda Gates: Contraceptives Are 'One of the Greatest Antipoverty Innovations' in the World," *People*, June 26, 2017. As of June 27, 2017:
http://people.com/human-interest/melinda-gates-foundation-contraceptives-greatest-anti-poverty-innovations/

Kimaro, Honest, and Jose Nhampossa, "The Challenges of Sustainability of Health Information Systems in Developing Countries: Comparative Case Studies of Mozambique and Tanzania," *Journal of Health Informatics in Developing Countries*, Vol. 1, No. 1, 2007.

Lavrakas, Paul, ed., *Encyclopedia of Survey Research Methods*, Thousand Oaks, Calif.: Sage Publications, 2008. As of June 22, 2017:
http://sk.sagepub.com/reference/survey

Marquis, Jeff, Michael McNerney, S. Rebecca Zimmerman, Merrie Archer, Jeremy Boback, and David Stebbins, *Developing an Assessment, Monitoring, and Evaluation Framework for U.S. Department of Defense Security Cooperation*, Santa Monica, Calif.: RAND Corporation, RR-1611-OSD, 2016. As of November 9, 2017:
https://www.rand.org/pubs/research_reports/RR1611.html

MEASURE Evaluation, "Data Demand and Use," 2017. As of August 4, 2017:
https://www.measureevaluation.org/our-work/data-demand-and-use

Mog, Justin M., "Struggling with Sustainability—A Comparative Framework for Evaluating Sustainable Development Programs," *World Development*, Vol. 32, No. 12, 2004, pp. 2139–2160.

Moore, Audrey-Marie Schuh, "Data for Effective Decision Making," *EQ Review: Educational Quality in the Developing World*, Vol. 5, No. 2, 2007. As of April 19, 2017:
http://www.equip123.net/EQ_Review/5_2.pdf

Moore, Melinda, David J. Dausey, Bounlay Phommasack, Sok Touch, Lu Guoping, Soe Lwin Nyein, Kumnuan Ungchusak, Nguyen Dang Vung, and Moe Ko Oo, "Sustainability of Sub-Regional Disease Surveillance Networks," *Global Health Governance*, Vol. 5, No. 2, 2012.

Nutley, Tara, *Improving Data Use in Decision Making: An Intervention to Strengthen Health Systems*, MEASURE Evaluation Special Report, 2012.

O'Mahony, Angela, Ilana Blum, Gabriela Armenta, Nicholas E. Burger, Joshua Mendelsohn, Michael J. McNerney, Steven W. Popper, Jefferson P. Marquis, and Thomas S. Szayna, *Assessing, Monitoring, and Evaluating Army Security Cooperation: A Framework for Implementation*, Santa Monica, Calif.: RAND Corporation, forthcoming.

Otieno, Yvonne, and Tom Arunga, "Managing Data with DHIS2: Improving Health Commodities Reporting and Decision Making in Kenya," July 31, 2014. As of November 9, 2017:
https://www.msh.org/news-events/stories/managing-data-with-dhis2-improving-health-commodities-reporting-and-decision

Pappaioanou, Marguerite, Michael Malison, Karen Wilkins, Bradley Otto, Richard A. Goodman, R. Elliott Churchill, Mark White, and Stephen B. Thacker, "Strengthening Capacity in Developing Countries for Evidence-Based Public Health: The Data for Decision-Making Project," *Social Science & Medicine,* Vol. 57, No. 10, 2003, pp. 1925–1937.

Performance Monitoring and Accountability 2020 Project and Kwame Nkrumah University of Science and Technology, "Detailed Indicator Report: Ghana 2013," Baltimore, Md.: PMA2020, 2013. As of November 9, 2017:
https://www.pma2020.org/sites/default/files/Detailed-Indicator-Report-GHPMA2013-110314.pdf

Plowman, Beth Ann, and Jean Christophe Fotso, *UNICEF Evaluation of the Multiple Indicator Cluster Surveys (MICS)—Round 4. Cluster 1: Response to Lessons Learned in Prior Rounds and Preparations for Round 5*, UNICEF, 2013. As of November 9, 2017:
https://www.unicef.org/evaluation/files/MICS_evaluation_Cluster_1__final.pdf

PMA2020, "PMA2020 Technical Note on Statistical Properties & Costing Final," Baltimore, Md., June 2014.

PMA2020, "Highlights and Recommendations: PMA2020 Consultative Meeting," New York, N.Y., May 9–10, 2016a.

PMA2020, "PMA DataLab," 2016b. As of November 9, 2017:
http://datalab.pma2020.org/

PMA2020, "Calculations from the PMA2020 Program," Baltimore, Md., May 5, 2017a.

PMA2020, "Data & Research: Indicators by Topic Area," 2017b. As of November 9, 2017:
https://pma2020.org/indicators-topic-area

PMA2020, "Glossary of Family Planning Indicators," 2017c. As of November 9, 2017:
https://www.pma2020.org/glossary-family-planning-indicators

PMA2020, "Glossary of Service Delivery Point Indicators," 2017d. As of November 9, 2017:
https://pma2020.org/glossary-service-delivery-point-indicators

PMA2020, "Highlights and Recommendations: PMA2020 Consultative Meeting," Chicago, Ill., April 26, 2017e.

PMA2020, "PMA2015/GHANA-R4 SOI: Snapshot of Indicators," 2017f. As of August 8, 2017:
https://pma2020.org/pma2015ghana-r4-soi

PMA2020, "PMA2016/UGANDA-R4 SOI: Snapshot of Indicators," 2017g. As of August 8, 2017:
https://pma2020.org/pma2016uganda-r4-soi

PMA2020, "Sampling Overview," 2017h. As of November 9, 2017:
https://www.pma2020.org/sampling-overview

Practical Concepts Incorporated, *Guidelines for Teaching Logical Framework Concepts*, undated. As of November 9, 2017:
http://pdf.usaid.gov/pdf_docs/pnaec576.pdf

Prochaska, James O., and Carlo C. DiClemente, "Stages and Processes of Self-Change of Smoking: Toward an Integrative Model of Change," *Journal of Consulting and Clinical Psychology*, Vol. 51, No. 3, 1983, p. 390.

Rosenberg, Leon J., L. D. Posner, and E. J. Hanley, "Project Evaluation and the Project Appraisal Reporting System," U.S. Agency for International Development and Fry Consultants Incorporated, 1970.

Rutstein, Shea O., and Ann Way, *The Peru Continuous DHS Experience*, Rockville, Md.: ICF International, 2014. As of November 9, 2017:
http://dhsprogram.com/pubs/pdf/OP8/OP8.pdf

Ryan, Gery W., and H. Russell Bernard, "Techniques to Identify Themes," *Field Methods*, Vol. 15, No. 1, 2003, pp. 85–109.

Seltzer, Judith R., *The Origins and Evolution of Family Planning Programs in Developing Countries*, Santa Monica, Calif.: RAND Corporation, MR-1276-WFHF/DLPF/RF, 2002. As of November 9, 2017:
https://www.rand.org/pubs/monograph_reports/MR1276.html

SocioCultural Research Consultants, "Dedoose Version 7.0.23," 2017. As of November 9, 2017:
http://www.dedoose.com/

Spruit, Marco, and Katharina Pietzka, "MD3M: The Master Data Management Maturity Model," *Computers in Human Behavior*, Vol. 51, 2015, pp. 1068–1076.

Starbird, Ellen, Maureen Norton, and Rachel Marcus, "Investing in Family Planning: Key to Achieving the Sustainable Development Goals," *Global Health: Science and Practice*, Vol. 4, No. 2, 2016, pp. 191–210.

Track20, "FP Goals," 2017. As of January 18, 2018:
http://www.track20.org/pages/our_work/innovative_tools/FPgoals.php

UN—*See* United Nations.

UNICEF—*See* United Nations Children's Fund.

United Nations, "The Sustainable Development Agenda," 2016. As of November 9, 2017:
http://www.un.org/sustainabledevelopment/development-agenda

United Nations Children's Fund, *Monitoring the Situation of Children and Women for 20 Years: The Multiple Indicator Cluster Surveys (MICS) 1995–2015*, New York, September 2015.

United Nations Development Programme, *Human Development Report*, New York, 2016. As of November 9, 2017:
http://hdr.undp.org/en/content/human-development-index-hdi

U.S. Agency for International Development, "Problem with Dates in the Ethiopia Datasets," undated. As of August 2017:
http://userforum.dhsprogram.com/index.php?t=msg&goto=66&

U.S. Census Bureau, "Design and Methodology: American Community Survey," Washington, D.C., 2009.

USAID—*See* U.S. Agency for International Development.

Weiss, Carol H., "Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families," *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts,* Vol. 1, 1995, pp. 65–92.

World Bank, "World Bank Country and Lending Groups," 2017. As of August 24, 2017:
https://datahelpdesk.worldbank.org/knowledgebase/articles/906519

World Health Organization, "Service Availability and Readiness Assessment (SARA)," 2017. As of November 9, 2017:
http://www.who.int/healthinfo/systems/sara_indicators_questionnaire/en/

Young, Stephanie, Henry H. Willis, Melinda Moore, and Jeffrey Engstrom, *Measuring Cooperative Biological Engagement Program (CBEP) Performance: Capacities, Capabilities, and Sustainability Enablers for Biorisk Management and Biosurveillance*, Santa Monica, Calif.: RAND Corporation, RR-660-OSD, 2014. As of December 14, 2017:
https://www.rand.org/pubs/research_reports/RR660.html

Zimmerman, Linnea, "PMA2020 General Sampling Strategy Memo," Johns Hopkins University and Bill & Melinda Gates Institute for Population and Reproductive Health, 2017. As of November 9, 2017:
https://pma2020.org/sites/default/files/
PMA2020 Survey Sampling Strategy Memo_v7_2017.05.02_lz-web.pdf

In 2013, the Bill & Melinda Gates Foundation launched two programs to help monitor progress toward a new global goal to increase modern contraceptive use by 2020. The Performance Monitoring and Accountability 2020 (PMA2020) program aimed to support annual, rapid-turnaround, nationally representative surveys of households and service delivery points in nine countries. Track20 was designed to support global standardization of key family planning indicators and country-level monitoring and capacity-building in 22 countries. This evaluation of both programs is based on interviews with more than 260 stakeholders in the United States and 15 program countries, statistical analysis of the PMA2020 survey, and analysis of stakeholder ratings of data maturity and sustainability.

Stakeholders felt that PMA2020 has successfully conducted annual, rapid-turnaround surveys with high-quality data. However, it has not fully achieved its original objectives of promoting data use, meeting local data needs, or integrating PMA2020 into country data systems. The team's statistical analysis of PMA2020 surveys identified opportunities for modifications in survey frequency, design, and content. Stakeholders felt that Track20 is on target to achieve most of its objectives. Monitoring and evaluation officers are the core of Track20 in program countries: They are highly skilled personnel, typically embedded within ministries of health, giving them ready access to decisionmakers. The RAND research team recommended that both programs promote country-driven agendas for data collection, use, and ownership; intensify focus on data use; and plan for and measure data maturity and data system sustainability. The research team also recommended a new program—Data for Action Training Activity for Family Planning (DATA-FP)—to increase country capacity for data system management.

www.rand.org

$70.00

ISBN-10 0-8330-9940-X
ISBN-13 978-0-8330-9940-2

RR-2112-BMGF