Measuring impact qualitatively

October 2018

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

Measuring impact qualitatively

Susan Pietrzyk, PhD
with Reeti Hobson, MPH, Lwendo Moonzwe, MPH, PhD,
and Debra Prosnitz, MPH, PMP

October 2018

MEASURE Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, North Carolina 27516
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of the MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004.MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-17-230
ISBN: 978-1-64232-094-7

USAID FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

ACKNOWLEDGMENTS

The Measuring Impact Qualitatively study was undertaken as an activity of the U.S. Agency for International Development (USAID)-funded MEASURE Evaluation project. The study was led by Susan Pietrzyk, with contributions from Reeti Hobson, Lwendo Moonzwe, and Debra Prosnitz.

Special thanks go to Sian Curtis, Jessica Fehringer, Heidi Reynolds, Ani Hyslop, and Shannon Salentine for their technical review and excitement about the study.

We also thank Gretchen Bitar for her support in implementing the webinar, as well as Marie Mikulak and Cindy Young-Turner for editing and graphics support.

ABSTRACT

Based on a document review of 32 USAID-funded HIV/AIDS-related evaluation reports from the USAID Development Experience Clearinghouse (DEC) and participant research, the Measuring Impact Qualitatively paper provides insights and strategies for evaluators undertaking performance- and impact-focused evaluations. In outlining some of the ways in which integrating qualitative research methods into evaluations occurs in perfunctory ways, the paper advocates expanding discussions and moving beyond the status quo with end line performance evaluations. The way forward, the paper suggests, centers on improving recognition that the use of qualitative research methods within evaluations cannot simply be quick and simplistic or a last-minute addition or afterthought; yet, improving qualitative research methods within evaluations is not overly complicated. To this end, the paper provides logical, common sense suggestions that emerged from reviewing 32 evaluation reports and thinking through choices, strengths, and weaknesses. The paper concludes that the endeavor to "measure impact qualitatively" is less about developing a new framework and much more about recognizing that evaluations have long sought to conceptualize, understand, and assess impact in qualitative terms. The improvements needed involve ensuring that (1) the use of qualitative research methods is thoughtful and attentive to detail, explanatory text, and transparency; and (2) the evaluation considers gender appropriately.

CONTENTS

TABLES

ABBREVIATIONS

DEC	Development Experience Clearinghouse
GAO	U.S. Government Accountability Office
NGO	nongovernmental organization
PEPFAR	U.S. President's Emergency Plan for AIDS Relief
PMTCT	prevention of mother-to-child transmission
USAID	U.S. Agency for International Development

EXECUTIVE SUMMARY

Introduction

The aim of the Measuring Impact Qualitatively study was to provide insights and strategies for evaluators undertaking performance and impact evaluations. The endeavor sought to include critical, judicious, and pragmatic thought about past and present HIV/AIDS-related U.S. Agency for International Development (USAID) project evaluations. The objectives of the study were as follows:

1. Examine the various ways qualitative research methods are applied and how the associated terminology is used in evaluations and evaluation reports, and how qualitative work can continue to contribute to understanding and assessing impact
2. Develop recommendations that evaluators can use to help strengthen the use of qualitative research methods in performance and impact evaluations

Methods

The primary method used for the study was a document review of USAID-funded HIV/AIDS-related evaluation reports from the USAID Development Experience Clearinghouse (DEC). Four social science researchers selected the documents using an iterative process of searching and decision making. The objectives of the document review were as follows:

1. To review a purposefully selected set of USAID-funded HIV/AIDS-related evaluation reports critically and practically
2. To assess and discuss qualitative research methodological choices, strengths, and weaknesses, as well as the integration of gender into evaluations

The documents were coded using ATLAS.ti software, using both predetermined and emergent codes. The intent in coding was to identify commonly used methods, terms, and writing conventions across the 32 evaluation reports. After completing the document review coding, the study team held a webinar to present some of the preliminary findings and solicit feedback, both incorporated in this paper. A total of 585 individuals registered for the webinar, with 308 attending.

In addition to document review, the study team engaged in activities such as attending webinars, mapping out evaluation-focused organizations, learning the history of those who have advocated for impact evaluations, and networking via Twitter and LinkedIn. Engaging in these activities was a form of participant research and facilitated peer-to-peer interaction as well as reflection regarding the field of evaluation, including the role of qualitative research methods.

Findings

The documents reviewed consist of 32 USAID-funded HIV/AIDS evaluation reports (2003–2015). Thirteen evaluations are described by the report authors as impact evaluations. The 19 non-impact evaluations largely qualify as performance evaluations. Across the 32 evaluations, 28 focus on a single project in a single country, with 21 different countries, 25 evaluating organizations, and 26 project implementers represented. The key informant interview is the most commonly referenced qualitative data collection method across the 32

reports. The findings and possible lessons that emerged from the document review and participant research fell into three broad themes, as noted below.

Strategies for establishing an evaluation approach: In its 2011 policy around evaluation, USAID asserts that impact evaluations are to have a rigorously defined counterfactual, whereas performance evaluations focus on descriptive and normative questions. The strictness of these definitions appears less apparent across the 32 evaluation reports reviewed, probably because some date back to 2003, before the policy. A lack of commonality in design, method, and overall approach appears across the reports. Assessing this lack of commonality helped to tease out strategies for establishing a more consistent qualitative evaluation approach, which are as follows:

- Across the 32 evaluation reports reviewed, 13 indicate the use of both qualitative and quantitative methods, whereas 10 assert the use of mixed methods, worded as such. A **mixed-methods approach** is a specific strategy. Taking a deep dive into such a strategy generally was not the norm across these 23 reports indicating the use of either both qualitative and quantitative methods, or mixed methods. Instead, the assertion was often largely in the use of words only and less the result of an evaluation team recognizing and embracing the details and nuances of what "mixed" actually means. For example, an evaluation that undertakes a survey as well as focus group discussions does not automatically mean taking a mixed-methods approach unless those involved in designing and implementing the evaluation make it a mixed methods approach.
- All 32 of the evaluation reports reviewed include a section entitled "Methodology." Beyond this common section heading, however, the content, level of detail, and writing quality in the different methodology sections varies. A sub-section entitled "Design" is included in 11 of the reports. A **strong presentation of design and method** comes about by (1) distinguishing (not conflating) design and method; (2) developing a specific design statement; and (3) not equating design and method to lists of activities, disclaimers about what could not be accomplished, protective statements, and circular logic that merely reverts back unproductively to the scope of work.
- The 32 evaluation reports present 32 unique evaluation approaches. This variety suggests a certain level of reliance on conceptualizing each evaluation as fit-for-purpose. A **fit-for-purpose** approach tailors an evaluation to meet the latter's needs. The specifics of fit-for-purpose evaluation relate to, for example, budget, time, data availability, data logistics, data quality, programming, and topical subject matter. In consideration of the seeming prominence of fit-for-purpose thinking, it would potentially be beneficial to be frank and acknowledge fit-for-purpose elements.

Traps to understand and avoid: Many manuals and other publications exist to provide guidance for integrating qualitative data collection and data analysis methods in evaluations. In ways that manuals often do not, this review of 32 evaluation reports often sought to identify weaknesses within and across evaluations. The intent of this critical lens has been to provide insights into traps to understand and avoid, including the following:

- Many of the evaluation reports that undertook primary qualitative data collection used the phrase "qualitative data collection" instead of providing the details. This phrase **lacks specificity,** as do "key informant interview," "focus group discussion," and "in-depth interview." These types of phrases provide little detail concerning specific data collection choices. Often, the writing across the

reports treats qualitative data collection almost as an afterthought. The presentation of qualitative data collection methods requires description and explanatory text. Taking a short cut and merely including a term is not adequate.

- Across the evaluation reports, there are numerous instances in which it appears that data collection methods may have come about almost as if they were thrown in at the end. In these instances, the details are lacking on two fronts—the reader learns nothing about the logic or intent of certain aspects of the data collection strategy or what certain aspects of that strategy yielded. There seems to be a reliance on a data collection strategy that simply involves turning on an evaluation **automatic pilot** button. Further, the review makes apparent the importance of maintaining transparency regarding recording, storing, managing, and analyzing data from interviews and focus group discussions.
- Across the reports, no single or universal evaluation design, data collection method, or data analysis technique leap out as the most commonly used approach, nor does a common lexicon emerge. A strong evaluation necessitates being precise with words as part of avoiding **simply mapping quantitative onto qualitative.** Such a strategy represents an exercise in knowing that, in some instances, the words, phrases, and techniques common to quantitative work will not be logical for qualitative work.

Addressing gender in qualitative research methods in evaluation: A gender-sensitive evaluation is suitable for any type of project, regardless of whether it integrates gender as part of its overall approach. USAID now requires that all projects integrate gender, and the international development community increasingly recognizes the benefits of all evaluations adopting gender-sensitive approaches. Some of the important elements for gender-sensitive evaluation are as follows:

- Know the definitions and commit to them. Sex is our anatomy. Gender is a socioculturally defined set of economic, social, and political roles, responsibilities, rights, entitlements, and obligations.
- Understand the continuum. Generally, projects are gender blind or gender aware in the extent to which they address gender norms, relations, and inequities. The type and degree of gender awareness is likely to fall along a spectrum of being gender exploitative, accommodating, or transformative.
- Recognize intersectionality. Gender, as an identity and marker of difference, intersects with other markers of difference and vulnerabilities, such as age, disability, ethnicity, race, religion, and sex. Recognizing intersecting identities is important when developing a data collection plan.
- Disaggregate data by sex (not by gender).

Conclusions and Recommendations

The review of the 32 evaluation reports and resulting Measuring Impact Qualitatively study overall reveal that the planning and implementation associated with integrating qualitative research methods in evaluations often come about only in perfunctory ways. The main recommendations revolve around ensuring that (1) the use of qualitative research methods is thoughtful and attentive to detail, explanatory text, and transparency; and (2) the evaluation considers gender appropriately.

INTRODUCTION

In broad strokes, international development projects seek to create a positive impact. The word "impact," however, has come to denote a specific evaluation methodology, as illustrated in the definition of impact evaluation used by USAID:

> *Impact evaluations measure change in a development outcome attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change.* (USAID, 2011)

In recognition of the important role that qualitative research methods play within evaluations, MEASURE Evaluation approved a study to explore ways to broaden conversations about the possible ways to conceptualize, understand, and assess impact in the context of HIV/AIDS-related international development projects. The aim of the Measuring Impact Qualitatively study was to provide insights and strategies for evaluators undertaking evaluations. The endeavor sought to include critical, judicious, and pragmatic thought about past and present HIV/AIDS-related USAID project evaluations. The objectives of the study were as follows:

1. Examine the various ways qualitative research methods are applied and how the associated terminology is used in evaluations and evaluation reports, and how qualitative work can continue to contribute to understanding and assessing impact
2. Develop recommendations that evaluators can use to help strengthen the use of qualitative research methods in performance and impact evaluations

Typical of trends, the impact evaluation trend appears to be fading. In a statement issued on March 21, 2017, the President's Emergency Plan for AIDS Relief (PEPFAR) moved away from supporting impact evaluations, citing cost, difficulties with tracking, and challenges in translating results into program improvement (Reynolds & Curtis, 2017). The PEPFAR statement signals a shift in approach and method, not a shift in the desire to understand and assess project impacts.

METHODS

The primary research method used for the study was a document review of 32 USAID-funded HIV/AIDS-related evaluation reports from the USAID DEC. The evaluation reports were coded in ATLAS.ti software, using both predetermined and emergent codes. The intent in coding was to identify commonly used methods, terms, and writing conventions across the reports. In addition, the study team engaged in activities such as attending webinars, mapping out evaluation-focused organizations, learning the history of those who have advocated for impact evaluations, and networking via Twitter and LinkedIn. Engaging in these activities was a form of participant research and facilitated peer-to-peer interaction as well as reflection regarding the field of evaluation, including the role of qualitative research methods. After completing the document review, the study team held a webinar to present some of the preliminary findings and solicit feedback, both incorporated in this paper. A total of 585 individuals registered for the webinar, with 308 attending.[1, 2]

Document Review

As the largest online resource for USAID-funded technical and project materials, the DEC currently includes nearly 200,000 documents, with thousands of items added each month. Four social science researchers selected the documents using an iterative process of searching and decision making. The objectives of the document review were as follows:

1. To review a purposefully selected set of USAID-funded HIV/AIDS-related evaluation reports critically and practically
2. To assess and discuss qualitative research methodological choices, strengths, and weaknesses, as well as the integration of gender into evaluations

Three members of the study team searched the DEC for USAID-funded HIV/AIDS-related evaluation reports during specific time periods (2003–2008, 2009–2013, and 2014–present). The fourth member of the team searched for pertinent background and cross-cutting materials (e.g., USAID policy documents, policy reports, technical notes, manuals or best practice documents, reviews and advocacy around gender-sensitive evaluations, and other relevant technical reports). Setting the lower bound of the search parameter at 2003 aligned with the establishment of PEPFAR.

All "items" in the DEC have been categorized as one of eight types: document, image, video, audio, acronyms and abbreviations, blog post, forum topic, and project description 1946–1996. Our search focused on one item type—documents. All DEC "documents" have been categorized as one or more of 35 document types. To conduct the search, we used the advanced search filters for document type, date, and keywords. We focused on the document types most likely to be evaluations (e.g., assessment, final evaluation report, special evaluation) and excluded types that clearly were not evaluations (e.g., annual report, loan or grant agreement, program planning document). We also excluded baseline studies and mid-term evaluations, as well as non-HIV/AIDS-related evaluations. We further refined our searches by using the keywords "evaluation," "impact

[1] The recording and PowerPoint slides from the webinar can be found at: http://www.cpc.unc.edu/measure/resources/webinars/measuring-impact-qualitatively.

[2] The specific questions posed, along with written responses, are included in Appendix 1.

evaluation," "performance evaluation," "assessment," "HIV," "AIDS," "mixed methods," "quasi-experimental," and "non-experimental." In particular, we looked for both mixed-methods and predominantly qualitative impact evaluations. We did not specifically seek out gender-sensitive evaluations.

The study team met regularly to discuss findings and exclude documents deemed not relevant. Our initial review (Round 1) resulted in winnowing the number of documents to 224. The subsequent winnowing process centered on reading the executive summary and methods sections of each document (Rounds 2 and 3). In Round 4, with the number of documents reduced from 144 to 104, we grouped the remaining documents as either mixed-methods impact or qualitative-oriented evaluations. We approached these two categories broadly, primarily based on how the authors described the evaluation. Overall, this was not a systematic review of every HIV/AIDS-related evaluation report in the DEC, but rather a purposeful document selection. Table 1 provides a summary of the number of documents included in each search round. A list of the 224 documents from the initial review, along with a tally tracing the search round winnowing process, is included in Appendix 2.

Table 1. Number of documents included in each search round

	Round 1	Round 2	Round 3
2003–2008	76	53	44
2009–2013	133	100	85
2014–present	15	15	15
TOTAL	***224***	***168***	***144***

	Round 4 2003–present	Round 5 2003–present
Mixed-methods impact evaluations	23	15
Qualitative-oriented evaluations	81	17
TOTAL	***104***	***32***

We selected 32 evaluation reports for more in-depth review—15 mixed-methods impact evaluations and 17 qualitatively oriented evaluations (Appendix 3). We sought to achieve balance and variety across the evaluation reports selected using the geographical location of the project evaluated, evaluating organization, project implementer, and year of publication.

Data Analysis

The first step in the analysis was to collectively read and discuss the design for each evaluation, the methodologies and methodological terminology used, the writing in the report, and the rationale behind the evaluation's overall approach. We also considered the following questions:

- How is impact being defined in the evaluation report?
- What do the authors consider evidence of impact (or related concepts, such as success, outcomes, performance, etc.)?
- Do the authors view quantitative and qualitative evidence the same or differently?
- What data sources constitute evidence? If people represent the data sources, which people are being surveyed, interviewed, and so on, and are these people in the position to assert that impact occurred?

One member of the study team entered each report into the ATLAS.ti coding and analysis software package. The initial coding covered the basic characteristics of the 32 evaluation reports (e.g., publication date, country, page length). Subsequent coding involved applying both predetermined and emergent codes. Predetermined codes included labels for specific aspects of each evaluation report (e.g., design, control group, qualitative evaluation methods, discussion of impact). Emergent codes served as a way to flag unique, interesting, or unexpected aspects of each report (e.g., regarding writing style, data collection choices, limitations).

Analysis was a purposeful exercise to mine the evaluation reports for insights. The aim was to use coding and analysis to read each report through favorable, critical, and pragmatic lenses. This effort was twofold: teasing out commonly used qualitative research methods, terms, and writing conventions in evaluations, and considering how to address gender with qualitative research methods in them. The intent was not to assess the evaluating organizations or project implementers, and the analysis does not link organizations with evaluation reports.

FINDINGS

This section of the paper examines the findings from the study. The aim of this analysis in considering the findings from both the document review and the participant research activities, is to map out various methodological strengths, weaknesses, and choices (including the integration of gender) and, in turn, offer insights and suggestions for enhancing the use of qualitative data collection and data analysis methods in evaluative work. Throughout this paper, the recommendation is straightforward: Increased attention to the detail, nuance, and transparency of qualitative research methods in evaluative work goes hand in hand with understanding and assessing impact in qualitative terms. This section includes individual subsections focused on (1) overview of the documents reviewed (2) overview of the participant research, (3) strategies for establishing an evaluation approach, (4) traps to avoid and understand, and (5) addressing gender with qualitative research methods in evaluation. The last subsection includes specific tips for "measuring impact qualitatively" in evaluative work.

Overview of the Documents Reviewed

The documents reviewed consisted of 32 USAID-funded HIV/AIDS evaluation reports. Thirteen were described as impact evaluations; however, most of them had limitations in relation to the impact evaluation design (e.g., limitations with a rigorously defined counterfactual, ex-post design). The 19 non-impact evaluations were categorized primarily as final performance evaluations. Across the 32 evaluations, 28 focused on a single project in a single country, with 21 different countries, 25 evaluating organizations, and 26 project implementers represented. The reports were 21–339 pages long, for an average of 106 pages, including appendices; and 52 pages, excluding them. The average length of 52 pages is not surprising, given that USAID generally sets a page limit of 50 pages for evaluation reports.

Table 2. Summary of the 32 evaluation reports reviewed

Evaluation report types		***Qualitative methods for primary data collection***			
Evaluation reports by data type	**32**	***Key informant interview***	***Focus group discussion***	***In-depth interview***	***Other qualitative method***
Mixed methods (use of this term)	10				
Qualitative and quantitative methods (no use of the term mixed methods)	13				
Primarily qualitative methods	9				
Primarily quantitative methods	0				
Evaluation reports by evaluation type	**32**				
Primary data collection (quantitative)	15				
▪ Impact evaluation	13				
Primary data collection (qualitative)	31	23	19	8	8
▪ Final performance evaluation	19				
Countries represented across the evaluation reports reviewed (21)					
Côte D'Ivoire (2), Ethiopia (3), Ghana (1), Honduras (1), India (3), Kenya (2), Malawi (2), multiple country evaluations (3), Namibia (1), Nepal (1), Rwanda (1), South Africa (4), Swaziland (1), Tanzania (1), Uganda (3), Zambia (2), and Zimbabwe (1)					

Year of the evaluation report, across the evaluation reports reviewed (12)
2004 (1), 2005 (1), 2006 (2), 2007 (1), 2008 (3), 2009 (2), 2010 (1), 2011 (3), 2012 (9), 2013 (4), 2014 (4), and 2015 (1)

Evaluating organizations represented across the evaluation reports reviewed (25)	
1. Business Enterprise, University of Pretoria	13. Management Sciences for Health
2. CAMRIS	14. Management Systems International (2)
3. Care India	15. MEASURE Evaluation (2)
4. CHANGES2	16. MELA PLC
5. Clacherty & Associates Education and Social Developments (Pty) Ltd	17. Mendez England & Associates
6. Development & Training Services (3)	18. MIDEGO, Inc.
7. Engender Health	19. The Mitchell Group (2)
8. FARST Africa	20. QED Group, LLC (3)
9. Feedback Research and Analytics, Inc.	21. Save the Children
10. HORIZONS (2)	22. Social Impact, Inc.
11. Impact Consulting	23. Social Scientific Systems
12. John Snow International	24. Synergy
	25. USAID

Project implementers represented across the evaluation reports reviewed (26)	
1. Care India	15. John Snow International Research & Training Institute, Inc.
2. CHANGES2	16. Johns Hopkins University Center for Communications Programs
3. Chemonics	17. Management Sciences for Health
4. Childline Mpumalanga	18. National Association of Child Care Workers
5. Children in Distress Network	19. Government of Ghana National HIV Prevention Program
6. Development Alternatives, Inc.	21. Pact
7. Elizabeth Glaser Pediatric AIDS Foundation (2)	22. Pathfinder
8. Engender Health (2)	23. Population Services International
9. FHI360	24. Save the Children (2)
10. GOAL	25. USAID
11. HORIZONS (2)	26. Zambia VCT Partnership
12. Humana People to People	
13. IntraHealth	
14. John Snow International (2)	

Overview of the Participant Research

The study team engaged in activities such as attending webinars, mapping out evaluation-focused organizations, learning the history of those who have advocated for impact evaluations, and networking via Twitter and LinkedIn. Engaging in these activities was a form of participant research and facilitated peer-to-peer interaction as well as reflection regarding the field of evaluation. Throughout the participant research, one dominant theme emerged—hesitancy and even fear to use the word "impact." This trepidation seemed to dissipate somewhat if a definition and explanation of intent were included. Often, definitions and explanations aligned with one of two schools of thought:

- The general goals of an impact evaluation are to map causal relationships, measure change over time, and attribute that change. With these goals, impact becomes largely a quantitative metric for comparing projects and their intended outcomes as they relate to broader objectives around sustainable economic and social development.

- More often than not, in impact evaluations, less emphasis is placed on considering impact in qualitative terms—that is, through the perspectives of individual beneficiaries, their specific experiences, and how those experiences might have changed over time. Such a focus can be important in accounting for the subjective aspects of the term "impact."

Strategies for Establishing an Evaluation Approach

In its 2011 policy on evaluation, USAID provides definitions for two types of evaluations: impact evaluation and performance evaluation.

- ***Impact evaluations*** measure change in a development outcome attributable to a defined intervention; impact evaluations are based on models of cause and effect, and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change.
- ***Performance evaluations*** focus on descriptive and normative questions: what a particular project or program has achieved; how it is being implemented; how it is perceived and valued; whether expected results are occurring; and other questions pertinent to program design, management, and operational decision making.

The definitions do not imply that all impact and performance evaluations are required to follow the same design and use identical methods. Indeed, both commonality and lack of commonality in design, method, and approach were constructive findings across the 32 evaluation reports reviewed.

Asserting Mixed Methods

Across the 32 evaluation reports reviewed, 31 indicate the use of qualitative data collection as part of the evaluation; one indicates the use of secondary qualitative data. The two statements below illustrate a subtle but telling terminology variation.[3]

> *The evaluation team used both qualitative and quantitative methods to collect and analyze information relevant to the objectives, the four outcomes of the development hypothesis, and the research questions outlined in the Scope of Work. (P3)*

> *A sequential mixed methods design was used to combine quantitative (survey design) and interpretive qualitative aspects. (P2)*

Across the 32 evaluation reports reviewed, 13 indicate the use of both qualitative and quantitative methods, and 10 assert the use of mixed methods, worded as such. At first glance, the thinking might be to add 13 and 10 and conclude that across the reports reviewed, there are 23 examples of a mixed-methods approach. After all, such an approach would include the use of both qualitative and quantitative data collection and data analysis methods. Such a conclusion is shortsighted, however. Simply including both qualitative and

[3] Although all 32 evaluation reports reviewed are publicly available, the analysis seeks to maintain anonymity to the degree possible. Throughout this paper, the parenthetical reference at the end of quotes represents a method to track the source for each. The redactions (e.g., XXXX) are to hide details that might enable a reader to connect the quote to a specific evaluation.

quantitative data collection and analysis methods does not automatically make the approach a mixed-methods one. "Both qualitative and quantitative" and "mixed" are not synonyms.

Establishing a mixed-methods approach for an evaluation represents a specific strategy. Specifically, a mixed-methods approach requires combining qualitative thought with quantitative thought—conceptually; methodologically; and at the design, data collection, and data analysis phases. Reality is also a factor in whether an evaluation can use a mixed-methods approach. Often evaluations have small teams as well as limited time, budget, evaluator availability, and other constraints. It is prudent to be aware of these types of constraints and ensure that an assertion of "mixed methods" is not simply the use of a phrase, but instead reflective of a mixed-methods approach consisting of in-depth and detailed thought and execution.

Throughout an evaluation, the evaluation team makes choices about specific information needs—both what information is needed and what form of data makes good sense. As a result, the team makes decisions concerning the relevance, feasibility, and usefulness of qualitative as well as quantitative data collection and analysis in meeting the stated information needs. Generally, with qualitative and quantitative data choices and decisions, evaluations tend to avoid the path of simply using one or the other by segregating qualitative and quantitative ways of thinking. Instead, evaluations increasingly integrate qualitative and quantitative data collection methods, and blend qualitative and quantitative analysis and reporting. Part of the integration and blending—as both strategy and approach—facilitates ways to be reflective about not necessarily categorizing data with the words "qualitative" and "quantitative," but rather focus on the data in relation to its insightfulness and utility.

Embracing what "mixed" methods means

With a mixed-methods strategy, it can be valuable to use those words and, in turn, important to really consider the nuances of what "mixed" means.

Why compartmentalize? Data are data

It may sometimes be beneficial to avoid thinking of data as either qualitative or quantitative. Often, data are simply data.

Presenting Design and Method

All 32 evaluation reports reviewed include a section entitled "Methodology." Beyond this common section heading, however, the content, level of detail, and writing quality in their respective methodology sections varies. A section titled "Design" is present in 11 of the reports. Although the lack of a section titled "Design" in 21 of the 32 reports does not necessarily indicate that these evaluations had no design, the reports frequently conflate design and method.

Across the evaluation reports, the terms "design" and "method" often are not well defined and accurately applied. In the first statement below, for example, quasi-experimental is termed a design, whereas in the second statement, it is a method.

A quasi-experimental study design was employed. (P12)

This study employed a quasi-experimental method. (P2)

Across the 32 reports, the most robust methodology sections tend to be those that include specific design statements. For example, the three statements below clearly assert a design statement.

This was a cross-sectional descriptive and analytical process and outcomes-based evaluation employing rapid participatory assessment techniques. (P5)

The evaluation employed integrated mixed methods, with an embedded quasi-experimental cluster design as its foremost feature, comparing the effects of the program on outcomes of interest across matched treatment and comparison communities. (P14)

The evaluation was conducted as a post-test only non-experimental design using quantitative and qualitative methods. (P27)

Know the purpose of a design statement

The scope of work is not the design; rather, evaluations are further designed after receiving the scope of work. A design statement begins the process of articulating the vision guiding an evaluation as well as the research and analysis techniques.

Precise writing around design and method

Clear thought and writing is crucial for distinguishing design and method. Precise articulation and writing go a long way in laying out a rationale for the design and method.

These types of specific design statements are the exception, however, not the norm across the reports. In the absence of a well-crafted design statement, the methodology sections often convey information on evaluation design via lists. The statement below, for example, is the opening paragraph of a methodology section.

The Team performed an intensive desk review of the documents, including those provided by USAID, and data encountered through Internet searches, site visits and discussions with local counterparts (Appendix B). A Team Planning Meeting was held XXXX to draft the evaluation framework, followed by an orientation meeting with USAID on XXXX. The framework was subsequently approved by USAID in an email dated XXXX. Following that meeting, the Team met in XXX with the implementing agencies and other stakeholders as identified by USAID (Appendix B). The Team finalized the evaluation framework in accordance with USAID feedback and guidance, and further tested and refined the discussion guides (Appendix D). (P24)

In the example below, the list includes seven methodologies, albeit several of the items on the list (e.g., meetings, briefings, presentations) are not methodologies. For example, meetings, briefings, and presentations might follow a participatory or consultative methodology, but they are activities rather than methodologies.

State the design and state it well

A list is a poor substitute for articulation of the detail and nuance that surrounds design statements, methodologies, and the logic for the choices made.

The methodology included (1) a team planning meeting between the team and USAID/XXXX; (2) extensive desk review of all project-related documents; (3) interviews with key informants; (4) field visits to three of the seven XXXX districts in XXXX, one XXXX district, and one XXXX district (control districts), where team members met select NGOs and community members and visited DAPCUs, ICTC, ART, and STI Clinics; (5) meetings with the XXXX in XXXX; (6) client briefings with XXXX through in-person meetings and teleconferences; and (7) presentations and discussion of findings with members of the XXXX team, USAID and XXXX. To enhance the quantitative rigor of the evaluation, the team undertook a separate epidemiological study to analyze health outcomes. The evaluation team conducted field visits to verify data collection and to inform subsequent findings qualitatively. Key informant interviews further enhanced the findings of the XXXX Documentation team and provided additional insight. (P4)

The two statements below contain lists that lack detail and nuance; instead, the lists appear to represent the evaluation design and method.

The evaluation methodology included a documents review; key informant interviews with project stakeholders and beneficiaries; focus group interviews with beneficiaries; [and] plenary group dialogue meetings and case study interviews of identified best practices. (P7)

The process evaluation has been based on scrutiny of all available program documentation, various publications on the HIV/AIDS problem and PEPFAR programs, plus some background information and technical reports. (P10)

Across the 32 evaluation reports, there are numerous instances in which the design statements include a protective tone and a disclaimer around the choices made. In the two statements that follow, the protective tone and disclaimer become dominant in the sentence. The detail of the design statement is lost.

A comparative mixed method design [was] structured as a rapid appraisal to fit within the eighteen days of field research time available to the team. (P17)

This evaluation used multiple sources of evidence to obtain a comprehensive and in-depth understanding of [the] complex, diverse and multiple phenomena present, to control the errors implicit in any chosen research method, to support sound analyses, to arrive at practical conclusions, and to make accurate inferences. (P25)

In the first example, the reference to having only a limited number of days of field research time detracts from the main part of the sentence. In the second example, defensive and disclaimer-like words bury the rest of the sentence.

Similar to embedding a disclaimer in a design statement is embedding circular logic pertaining to a scope of work provided to an evaluation team. A scope of work informs and guides the design and method of an evaluation, yet it is not itself the design and the method. Statements such as the two below, although asserting compliance to a scope of work, are ineffective in conveying the design and method for the evaluation.

The evaluation employed a mix of quantitative and qualitative methods to answer the evaluation questions. (P23)

The evaluation used a combination of methods and approaches for collecting and analyzing the information required to achieving the evaluation objectives. (P30)

Engage effectively with scope of work

Establishing design and method is not an exercise of restating a scope of work; rather, it is an exercise of responding to and refining a scope of work.

Design statements specifically, and a methodology section more broadly, benefit from balance. Thoughtful conceptualization and writing demonstrate an understanding of the scope of work but avoid going overboard by simply restating the scope of work or asserting the intent to follow it. In the statements below, for example, the text is somewhat list-like and reads like a restatement of the evaluation scope of work, yet the statements integrate thought, detail, and a response; as a result, they are effective in beginning to convey the design and method for the evaluation.

A team of three experienced consultants worked together to conduct this evaluation, drawing on the datasets of each project and data from the 2010 DHS, the health management information system, Monitoring and Evaluation Management

Systems (MEMS), TRACnet, PBF, IQ Chart and other assessments. Because a baseline study was not conducted, the performance evaluation focused on descriptive and normative questions. (P3)

This performance evaluation took place over a five-week period in XXXX, covering the first four and a half years of portfolio performance, with an emphasis on the last two years. The five-member evaluation team followed a methodology consistent with USAID's January 2011 Evaluation Policy, focusing on descriptive and normative questions, including: what the program had achieved; how it was being implemented; how it was perceived and valued; whether expected results were occurring; and other questions pertinent to program design, management, and operational decision making. (P22)

Embracing Fit-for-Purpose

The 32 evaluation reports present 32 unique evaluation approaches. This variety suggests a certain level of reliance on conceptualizing each evaluation as fit-for-purpose, a strategy to tailor an evaluation design to meet the information needs of the evaluation and various realities surrounding its timing and objective. With the purpose of an evaluation known, the evaluation team can specifically design it to fit that purpose. Across the evaluation reports reviewed, the term "fit-for-purpose" never appears.

Nonetheless, recognizing the fit-for-purpose dimension surrounding evaluation work is important for several reasons. First, evaluations take place in the real world and in the context of real-world constraints. Nearly all of the evaluation reports reviewed include statements such as those noted below.

With a small evaluation team and limited time for field data collection, the evaluation was conducted with only a small sample of project sites, activities and informants. (P28)

… the high level, strategic nature of the evaluation, lack of baseline information for some of the key indicators, and the limited time in which to conduct the evaluation. In practical terms, this meant that the evaluation team took a regional focus rather than conducting a detailed country by country evaluation. (P22)

Budget constraints did not allow for a full comparison study of the XXXX programme. In the absence of such a control group and wishing to make some kind of comparison, data collected for the XXXX evaluation was compared with findings from a control group from the XXXX evaluation where the same questions were asked and where possible, with the control group data from the KAP study in this evaluation where the questions were identical. (P16)

Second, each of the 32 evaluation reports reviewed identify constraints in relation to data availability and data logistics. More often than not, evaluations are interrelatedly fit-for-purpose regarding time and budget, and data. As the statements below indicate, data-related constraints can play a role in methodological options as well as the types of analysis that are and are not possible.

Maintain realistic expectations

Embracing fit-for-purpose as part of the evaluation approach focuses important attention on how an evaluation is actually undertaken and maintaining realistic expectations for it.

Sampling limitations: XXXX maintained a list of all households and community members reached by means of Household Registers that were only available in hard copy. Because these registers were kept in the XXXX office and there were in excess of 150 registers per site, it was not feasible to include the full list of households in the random sampling process. (P14)

The XXXX IPs had not been asked to capture financial data down to a sufficient level of detail to permit in-depth financial analysis (e.g., cost per supervisory visit) within the time allowed for the evaluation. These factors limited the methodologies available and prevented the team from conducting the requested cost-effectiveness analysis. (P3)

Data-related fit-for-purpose also extends beyond data availability and logistics. In the ideal world, evaluators would have access to high-quality secondary data; however, the sources of secondary data (e.g., baseline studies, mid-term studies, interim studies, program monitoring data and surveys, formative research) exist in the real world. Thus, secondary data are sometimes messy, of poor quality, inconsistently collected, and left unmanaged. For example, as noted below, indicator definitions change and projects change indicators.

There were problems with changing indicator definitions, numerators and denominators, questionnaire design, data entry and cleaning, and analysis in the B-E surveys, and inclusion criteria in the BSS surveys that should be avoided in the future. (P11)

The XXXX project has undergone significant changes in activities and PMPs, and the structure of work plans. (P23)

A third factor in the fit-for-purpose dimension surrounding evaluation work relates to programmatic and subject matter considerations. As noted in the statement below, international development evaluations often integrate various types of capacity building. Doing so is central to the broader mandate of international development; yet, the specifics of capacity building potentially have implications for the design and implementation of an evaluation.

For this study there was a strong need from the community stakeholders to have their own local people involved in the survey (to benefit their community members). (P14)

Evaluations generally strive to maintain a balance of methodological and subject matter expertise. In some situations, however, the pendulum of such expertise is not within the control of the evaluation team. The statement below highlights a situation in which a limitation related to the subject matter (HIV testing) essentially moved the evaluation toward a more fit-for-purpose approach.

The main limitation of the study was in the sensitiveness of the study topic. A number of people were reluctant to disclose whether they had taken HIV tests. (P18)

Although none of the reports reviewed explicitly indicate a fit-for-purpose approach and do not use the phrase, all of the evaluations, to varying degrees, took on elements of such an approach. The primary strategy for establishing an evaluation approach described across all of the reports centers on tailoring the evaluation to meet the information needs in the context of reality. These fit-for-purpose elements relate to budget, time, data availability, data logistics, data quality, programming, and topical subject matter. An effective evaluation design considers all of these elements.

Traps to Understand and Avoid

Present throughout this paper is a call for thoughtful inclusion of qualitative data collection and analysis methods in evaluations. The advocacy in this paper revolves around examining the qualitative data collection methods across the 32 evaluation reports reviewed and critical reflection on them as part of making recommendations for improvements.[4] To this end, Denzin (2010) presents an interesting critical assessment. In describing his analysis as an exercise in tracing some of the history surrounding increasing calls for triangulation and mixed-methods research and evaluation, his central assertion is as follows:

> *Persons who are less familiar with the rich traditions of qualitative inquiry are telling others with the same lack of experience how to do qualitative work. (Denizen, 2010, p. 420)*

This type of instruction exists, Denizen suggests, because the call for mixed-methods work has largely come from those with quantitative expertise, not those with qualitative expertise. Across the 32 evaluation reports reviewed, none provides biographical sketches with enough detail to know whether the evaluators are individuals with quantitative or qualitative expertise. Thus, the document review does not provide evidence to support or refute Denizen's assertions; however, to suggest that evaluations—including mixed-methods evaluations—have been and continue to be predominantly commissioned, designed, led, and overseen by individuals with quantitative expertise is not outlandish. In focusing on traps to understand and avoid, it is therefore constructive to consider Denizen's analysis as having stood the test of time and indicative of an overarching trap. This section discusses three traps to understand and avoid: lack of specificity, automatic pilot mode, and mapping quantitative onto qualitative.

Lack of Specificity

Among the 32 evaluation reports reviewed, 31 describe primary data collection using qualitative methods; many of them use the phrase "qualitative data collection" without providing further details. Often, the body of a report includes a reference to "qualitative data collection," and further discussion appears in an appendix, presumably to save pages as part of adhering to page limitations. Whether the reference is fleeting or relegated to an appendix, the phrase "qualitative data collection" frequently appears as a specific "method." This usage represents a common trap to avoid, as the details associated with any qualitative data collection effort are important.

The text box on the next page lists terms used across the 31 reports reviewed for various qualitative data collection methods. With 40 unique terms, there are multiple instances in which it appears the authors coined new terms to describe their methods. Doing so is not necessarily a problem, particularly if the purpose is specificity; however, a single term can potentially fall short in describing a qualitative data collection method. In the body of an evaluation report, the presentation of any qualitative data collection method benefits from a description of what the method is and what it entailed.

[4] Advocacy and instructions for how to undertake qualitative data collection and conduct qualitative data analysis are plentiful. For example, USAID has released Technical Notes entitled *Conducting Mixed Methods Evaluations* (June 2013) and *Impact Evaluations* (September 2013). Appendix 4 provides a list of additional similar literature.

Terms used by report authors for qualitative data collection methods

Capacity assessment tool	In-depth discussion
Case story	In-depth interview
Case study	Insightful occurrences
Case study documentations	Interactions with stakeholders
Client interaction	Interpretive phenomenology
Client interviews	Key actor interview
Community-level group discussion	Key informant interview
Comprehensive analysis session	Observational checklists
Detailed discussions	Observations
Detailed narratives	Open-ended interviews
Direct observations	Opportunistic group discussion
Discussions	Organization interviews
Email survey	Plenary group dialogue meetings
Facility checklist	Regular group analysis
Facility-level group discussion	Site visits
Field visits	Structured group discussions
Fieldwork	Structured interviews
Focus group discussion	Success story
Focus group interview	Telephone discussions
Guided group discussion	Triangulation

About half of the reports indicate that observation was included as a qualitative data collection method. In many instances, the report authors seem to convert the term "observation" to something slightly more specific, such as client interaction, field visits, observation checklists, and site visits. However, simply using these slightly more descriptive terms to denote a qualitative data collection method does not equate to providing the details and nuances concerning the data collection choices made and the logic behind them. Table 3 outlines some of the details and nuances left unanswered when relying only on a term as explanatory text.

Table 3. The value of explanatory text

Qualitative data collection method = observation	
Observation elaborated as:	***Details and nuances often missing from the reports reviewed***
Client interaction	Clients of whom? What clients? Why these clients and not others? What was the goal in interacting with clients? Was the goal achieved? Did the interactions lead to anything unexpected? What are possible reasons and contextual factors for variation between the goal and any unexpected outcomes?
Field visits	Where in the field? Why these field sites and not others? Is "the field" a community, an institution, or something else? What was the goal in making field visits? Who was visited? Was the focus of the visit observation only, or was there an interview or focus group discussion as well?
Observation checklists	Who is being observed? Is the observation of a person or a process? Checklists of what? Why this checklist and not a checklist of something else? How is what was observed part of the report?
Site visits	What sites? Are these sites where the project is being implemented? Are these sites where services are being provided? Why these sites and not others? What individuals were present at these sites, and were there interactions with them?

To a certain degree, many of the 31 evaluation reports appear to treat qualitative data collection methods in a perfunctory way. The message conveyed in many of them is essentially that it was a performance evaluation, so site visits (for example) were conducted. If a reader wonders why site visits were conducted, it would be hard to gain an understanding of the motive from the report. In addition, "site visit" as a stand-alone term does not indicate to a reader whether the visit included observation only or observation and an interview. In this example, the lack of specificity concerning the term "site visit" seems to signal an expectation that every reader of an evaluation report automatically knows what this term means, both in general and for that particular evaluation. In reality, without explanatory text a reader cannot know what choices the evaluation team made, nor understand the relevance and purpose of the site visits.

In evaluation work generally, qualitative data collection methods center on speaking with individuals or groups of individuals. One of the first sets of choices, and one area for specificity, revolves around the type of interview: key informant interview, focus group discussion, or in-depth interview.

Given the commonality of these terms, the assumption often is that these are specific terms to describe qualitative data collection methods. However, these terms do not provide the needed level of detail. For example, consider the three terms and think about what the word choice elicits and the questions that come to mind:

> ***Key informant interview:*** Is there another set of informants who are non-key informants? Is the individual being interviewed a project beneficiary? Does the individual work for the donor funding the project or directly for the project? Does the individual work for the government, a nongovernmental organization, or in the private sector? Is the individual a subject matter expert or a service provider? Does

the individual reside in the project implementation area? Is the individual a stakeholder? If so, what does the term "stakeholder" mean? Is the individual being interviewed a leader in his or her community?

Focus group discussion: With whom and how big a group? Men only, women only, or men and women together? Are those in the group project beneficiaries? Does anyone work for the donor funding the project or directly for the project? Is anyone a subject matter expert or a service provider? Does anyone reside in the project implementation area? Is anyone a stakeholder? If so, what does the term "stakeholder" mean? Is anyone one in the group considered a leader in their community?

In-depth interview: What makes the interview "in depth?" Is the individual being interviewed a project beneficiary? Does the individual work for the donor funding the project or directly for the project? Does the individual work for the government, a nongovernmental organization, or in the private sector? Is the individual a subject matter expert or a service provider? Does the individual reside in the project implementation area? Is the individual a stakeholder? If so, what does the term "stakeholder" mean? Is the individual being interviewed a leader in his or her community?

Across the 31 evaluation reports, more often than not the methodological discussion concerning key informant interviews, focus group discussions, and in-depth interviews appears in a single sentence or short paragraph. The statements below, for example, are the only ones regarding qualitative data collection methods in their respective reports.

Add details, be specific

Simply relying on commonly used terms such as "key informant interview," "focus group discussion," and "in-depth interview" is not sufficient because they lack specificity.

Forty sessions of Focus Group Discussions were held in the four provinces – 10 in each province, with two sessions per XXXX one [of] which comprised … XXXX and one other targeting XXXX. (P9)

Findings from key informant interviews and observations of the final evaluation team, including field visits to project sites in 5 states and 2 districts (see Table at end of Annex). (P11)

The primary data collection method was a household survey, supported by interviews and focus groups with stakeholders of interest, and observations made during site visits. (P14)

In contrast, giving full attention to the methodological choices when undertaking qualitative data collection and providing a thoughtful explanation of these methods are both important. The statement below, for example, is better in providing some specificity.

Based on detailed questions provided in the scope of work as well as questions arising from the document review, the team developed a comprehensive set of questions to be answered during the evaluation, and identified the sources of information to be used for each. From this list, the team designed interview guides for each of the following groups: USAID respondents, XXXX core partners, Government of XXXX officials, other donors, international NGOs, and NGO implementing partners. (Note: the team refers to sub-recipients or implementing agencies as implementing partners throughout the report.) The team identified key information to look for when observing the project sites of implementing partners. The complete set of key evaluation questions and interview guides is at Annex 5. The team interviewed, in addition to USAID and XXXX team members, about 40 individuals from government, other donors and INGOs. The team also interviewed and, in most

cases, observed activities of 31 past and present implementing organizations, representing XXXX networks and XXXX support groups; prevention partners; treatment and care partners; research, management and logistics organizations; and legal support partners. (P33)

This paragraph does a good job of specifying with whom the interviews were conducted. The various affiliations for each individual interviewed are clear, and it is useful to know that an interview guide was prepared for each group. To a certain extent, the strength of this paragraph rests in its exclusion of terms such vague terms as "key informant interview," "focus group discussion," and "in-depth interview." Not using these terms forced the report writers to be specific in describing the full set of interviews undertaken.

Missing from this paragraph, however, is specificity concerning why these particular groups of individuals were interviewed—the report does not present the logic behind the choices made. Providing detail is generally advisable; lack of specificity is a trap to understand and avoid.

Automatic Pilot Mode

In undertaking any evaluation, it is imperative to begin with and regularly circle back to its foundational questions. Specifically, they would be questions and discussions concerning the purpose of the evaluation, the data sources informing it, and the interrelationship of those two elements. Such a strategy helps the evaluators stay grounded in the goals, logic, and plan for the evaluation, and avoid a hurried or unthinking approach.

The risks of working for speed

A hurried evaluation runs the risk of taking the fastest path or easiest choice. Take time to develop a thoughtful approach to the evaluation. Avoid adopting an approach simply because it was used for other evaluations.

Deciding on an appropriate data collection strategy for an evaluation is not automatic, nor is there a default with respect to the sources for or appropriateness of primary or secondary data. Conveying how data collection strategies align with the purpose of the evaluation is imperative, particularly as part of avoiding the appearance of a side purpose or extra data collection being tossed into the mix without a reason.

Among the 31 evaluations that undertook primary qualitative data collection, numerous instances occur in which data collection methods appear to have come about as a side purpose, or have been thrown in at the end. In these instances, the details are lacking on two fronts—the reader learns nothing about the logic or intent of certain aspects of the data collection strategy, or what certain aspects of the data collection strategy yielded. The sentence below is an emblematic example.

Qualitative data were also collected through a total of 16 focus group discussions with XXXX and XXXX. (P12)

The report does not provide any additional information concerning the 16 focus group discussions. The reader does not learn the objectives or results of these 16 focus group discussions; instead, a single sentence represents the qualitative data collection strategy and its results. Single, brief sentences such as this one are common across the evaluation reports reviewed.

It is possible that those 16 focus group discussions yielded interesting information, informed the other data collection strategies and the overall presentation of the evaluation results; conversely, they may have been an afterthought in the context of a larger data collection strategy. The reader cannot determine the full story

from reading the report. Without such detail, a reader is left wondering how these focus group discussions align with the purpose of the evaluation and data collection strategy.

Table 4 presents the types and combinations of qualitative data collection methods used across the 32 evaluation reports.

Table 4. Qualitative data collection methods, by type

Number of times each data collection type used (n = 32)		
Key informant interview		23
Focus group discussion		19
In-depth interview		8
Other (e.g., case study, observation)		8
Number of types of data collection	**Times used**	**Possible combinations of types used**
Zero types of primary qualitative data collection	1	Not applicable
One type of primary qualitative data collection	11	5 times, key informant interview only 5 times, focus group discussion only 1 time, in-depth interview only
Two types of primary qualitative data collection	14	8 times, key informant interview and focus group discussion 3 times, key informant interview and other 1 time, key informant interview and in-depth interview 1 time, in-depth interview and focus group discussion 1 time, in-depth interview and other
Three types of primary qualitative data collection	5	2 times, key informant interview, focus group discussion, in-depth interview 2 times, key informant interview, focus group discussion, other 1 time, key informant interview, in-depth interview, other
Four types of primary qualitative data collection	1	1 time, key informant interview, focus group discussion, in-depth interview, other
TOTAL	**32**	

The data in the table indicate that key informant interviews were the predominant qualitative data collection method across the 32 evaluation reports reviewed. An automatic pilot way of thinking seems to prevail to a certain extent, in which performance evaluation equals key informant interviews. Key informants were generally experts and tended not to be direct project beneficiaries. Interviews with direct project beneficiaries, such as pregnant mothers, people living with HIV or AIDS, or sexually active adolescents, tended to be categorized as in-depth interviews.

In noting a distinction between key informant and in-depth interviews, it is useful to consider two additional reference points—the type of information and perspectives key informants provide, and where key informants tend to be located. Across the reports reviewed, it appears that key informants often provided information and perspectives about other people (direct project beneficiaries) and were located in the capital city (not a

project implementation area). An exercise to weigh the pros and cons of key informant interviews represents an important step in avoiding an automatic pilot way of thinking. Table 5 presents some of these pros and cons when choosing key informants, with each pro also a con and each con also a pro. Ultimately, whether to include key informant interviews in an evaluation is best treated as a case-by-case decision.

Table 5. Pros and cons of key informants

Key informants generally include the following: Project employees, donor organization staff, service providers, community leaders, subject matter experts, or staff working in NGOs or the international development field	
PROS—KEY INFORMANT	**CONS—KEY INFORMANT**
• Knowledgeable about the community and its residents, with a keen understanding of the nuances of the subject matter and issues the project is trying to address • Often closely connected to the project and with insider perspectives; can provide useful insights about how the project works and is being implemented • Often able to provide important historical and contextual perspectives	• Given their knowledge of the community, potentially have pre-established views about the beliefs, practices, and customs of the project beneficiaries in that community • As insiders, potentially have a stake in the results of the evaluation and may consciously or unconsciously focus on project successes and exclude a focus on challenges • Might lack knowledge of the details and day-to-day realities in the present moment

In addition to limited detail about the selection of data collection methods, procedures for data management often did not appear in the evaluation reports reviewed. Across the 31 evaluations that conducted primary qualitative data collection, slightly more than half included a discussion about the process and protocol for recording, storing, managing, and analyzing the interview, focus group discussion, or observation data. The three statements below are emblematic of how such a discussion unfolded.

Plan for data management

The need for transparency exists in relation to recording, storing, managing, and analyzing data. Nothing about these processes should be an ad hoc exercise. There is no automatic pilot button for these activities.

> *FGD and IDI [focus group discussion and in-depth interview] guides were prepared for each target group and translated to XXXX for ease of administration. All discussions were tape recorded after getting verbal consent from the study participants and notes were also taken. The recorded information was transcribed verbatim. (P1)*
>
> *Two local consultants and one international consultant conducted all interviews in XXXX. The interviews were tape-recorded; the most revealing and pertinent quotations were transcribed word for word. The transcribed quotations were analyzed using the Ethnograph software, which allows for coding of useful themes and subthemes and facilitates analysis. This report includes illustrative quotations from those transcriptions. (P29)*
>
> *For the qualitative data, each FGD was recorded and an enumerator was trained to take notes of the dialogue during each focus group and individual interview. Interviews were not individually transcribed but summarized in tables using Microsoft Word. This approach saved resources and allowed the study team to analyze the interview notes using structured data summary tables highlighting the major focus areas of the study. When individual quotes were needed, these were transcribed verbatim. No names were linked to quotes. Verbal consent to record the discussions was obtained from participants before the discussions began. (P13)*

Through a careful reading of these statements, the different data management and analysis choices become apparent. For example, each statement indicates that the evaluation team recorded all interviews and focus group discussions; however, each describes a different approach for transcribing the recording. In the first instance, the evaluation team produced a verbatim transcript. In the second instance, the team produced a transcript of only the pertinent portions. In the third instance, the team mined the recordings for quotes and transcribed them verbatim. It is important to emphasize that although these three approaches are different, none is more correct than the others, and each carries pros and cons.

Common across each of the three statements is transparency—the writing describes the choices. In the second and third examples, however, subjectivity or bias appears to be in play to a greater degree compared to the first example. Questions such as those noted below highlight where and when subjectivity entered into the equation in relation to the second and third statements.

- What constitutes revealing and pertinent quotes?
- How can you be sure that all of the revealing and pertinent quotes were transcribed?
- Who decided when individual quotes were needed?
- What criteria determined how one individual quote was picked over a different one?

A full verbatim transcript (as used in the first instance) potentially carries less subjectivity. When everything said during an interview or focus group discussion is included in a transcript, no one is making choices regarding what to include and exclude. Still, human beings produce transcripts and, as a result, no transcript—whether partial or full—is free of the possibility of subjectivity.

Own unavoidable subjectivity

Be careful in trying to establish "how much" subjectivity might be present or not present across different approaches. Instead, recognize that subjectivity is unavoidable. Be cognizant of and write transparently about this reality throughout an evaluation.

Time and budget constraints are two realities that sometimes make transcription unfeasible. In such instances, notes become a possible path through which to record, store, manage, and analyze data. The two statements below provide examples of instances where an evaluation team used notetaking in their approach to recording, storing, managing, and analyzing data.

> *The evaluators took detailed handwritten notes during the briefings, interviews and discussions which they typed up and analyzed the content. (P28)*
>
> *Team members took detailed notes from interviews and site visits. Information from document review, interviews and site visits were aggregated according to the evaluation questions described above so that team conclusions would be based on data derived from several sources. (P33)*

At first glance, the choice of notetaking, rather than recordings and transcripts, might drive an assessment of the choices made. As with recording and transcripts, notetaking as an approach carries pros and cons. The weakness with these two statements is less about notetaking as the choice and more about the lack of detail and transparency. Missing is detail on the structure of the notes and how those data were managed and analyzed in logical and systematic ways. This type of lack of transparency is common across the 32 evaluation reports reviewed.

As noted above, across the 31 evaluations that conducted primary qualitative data collection, slightly more than half included a discussion about the process and protocol for recording, storing, managing, and analyzing the interview, focus group discussion, or observation data collected, meaning that slightly less than half did not do so. In these instances, single statements such as the one noted below were common.

> *The main part of the impact evaluation consisted of field data collection (9 focus group discussions and 28 household interviews that were held at a total of 10 project sites). (P10)*

Although this statement provides useful numbers, what remains unknown is the data management plan. The concern and trap to avoid is giving the impression that interview and focus group discussion data were recorded, stored, managed, and analyzed only in the heads of the evaluation team members.

Mapping Quantitative onto Qualitative

Table 6 presents two lists of types of words and phrases pertaining to methods and techniques, many of which appear in the 32 evaluation reports reviewed. The left column includes words and phrases common to quantitative-oriented work; the right column incudes those common to qualitative-oriented work. Consideration of the table represents an exercise in reflection concerning word choice.

Table 6. Reflection concerning word choice

Quantitative work		*Qualitative work*
Representative	⟷	Strategically indicative
Instruments	⟷	Tools
Respondents	⟷	Participants
Sample	⟷	Site selection
Questionnaire	⟷	Interview guide
Structured script	⟷	Facilitated conversation
Counting responses	⟷	Grouping comments
Unit of analysis	⟷	Interview or participant type
Categorizing	⟷	Themes and patterns
Indicator estimate	⟷	Common behaviors and practices
Confidence interval	⟷	Spectrum of varied experiences
Statistically significant	⟷	Purposefully targeted
Measure change over time	⟷	Reasons for change over time
Measure impact	⟷	Assess and understand impacts
Generalizable	⟷	Time and place specific

Populating the two columns is not an exact science; rather, making the two lists based on reviewing 32 evaluation reports and engaging in participant research represents an exercise in encouraging discussion and reflection concerning word choice. With this set of 15 juxtapositions and the many additional possibilities, several questions are pertinent. For example, the inclusion of the arrows is to facilitate considering whether the paired words or phrases are interchangeable. Does the interchangeability arrow work one way or both

ways? Is it feasible and logical to map individual words or phrases from one approach (quantitative) onto another approach (qualitative)? What about the other way around? Are there instances in which mapping would be misleading or detrimental regarding the data collection, data analysis, or presentation of findings? These questions aim to encourage consideration of the possible implications of various word and phrase choices, and to advocate for reflection about the degree that quantitative words and phrases potentially carry undue weight and overly influence approaches to qualitative work.

Mindfulness around word choice

Be cognizant and precise about word choices as part of avoiding the trap of mapping quantitative onto qualitative automatically and without consideration.

Based on a review of the 32 evaluation reports, one example in which mapping quantitative onto qualitative is potentially problematic is in relation to the choice of using the word "sample." As a term and a technique, sample can be misleading when used in relation to qualitative data collection and analysis. Sampling implies statistical precision in relation to selecting (or sampling) a subset of geographic locations from the full set of geographic locations (or a subset of the population from the full population). Representative sampling thus facilitates a way to extrapolate from the data subset to the full set. With qualitative data collection, it is feasible to use terms and techniques, such as convenience sampling, purposive sampling, or snowball sampling, to signal lack of statistical precision. There are also advantages in acknowledging that, technically speaking, when using qualitative data collection methods, sampling is not feasible. Instead, there is a site selection process and a participant recruitment process. Each process is purposeful and strategic, with a set of criteria guiding the decision making. Transparently explaining a site selection process and a participant recruitment process is important. Simply invoking the word "sample" does not represent transparency or an explanation. Statements such as the one below, for example, are common across the reports reviewed. In this statement, it appears that both use of "sample" and stating the sample size is meant to represent a transparent explanation regarding the participant recruitment process.

> *A total of 181 participants made up the qualitative sample (Table 3). (P15)*

In recreating the shell of Table 3 from this report (Table 7), it appears that some transparency and explanation is provided. However, the focus seems to be more on the end result of the sample and less a description of the participant recruitment process. Across the 32 evaluation reports reviewed, the focus on the result of qualitative "sampling" often is greater than that on the details concerning the site selection process and recruitment process.

Table 7. Recreation of Table 3

Qualitative data source	Organization	Number of participants	Number of females	Number of males
Staff interviews				
OVC focus groups				
HIV support group				
3 x caregiver focus groups				
5 x caregiver home visits				
Key informant interviews				

Based on a review of the 32 evaluation reports, one example in which mapping quantitative onto qualitative is potentially advantageous is unit of analysis. The phrase "unit of analysis" refers to who or what is being analyzed (e.g., person, persons, place, practice, belief); associated techniques would center on making clear the perspective or point of view of the data. In this context, attentiveness to the unit of analysis in qualitative data collection and analysis is potentially advantageous in two ways.

1. If qualitative data collection and analysis aims to complement a survey and indicator estimates, it is important to understand the unit of analysis for each individual indicator. For example, certain indicators collected through a population-based household survey can have a different unit of analysis compared to those collected through a beneficiary- or facilities-based survey. More specifically, a household as a unit of analysis represents a group of people, whereas a beneficiary as a unit analysis may only represent one person, and a facility as a unit of analysis represents an institution potentially providing services. As units of analysis, these examples are not neatly congruent. If interviews and focus group discussions are to be conducted to complement a survey and indicator estimates, it might be useful to align the unit of analysis for the interviews and focus group discussions to the various indicators (e.g., household, beneficiary, facility).

2. Generally, three terms describe interview types. Each interview type tends to have a different unit of analysis; that is, the perspective or point of view for each is different. It is valuable to be attentive to these differences when analyzing and reporting findings from interview data. The interview types are as follows:

    - In-depth interviews—Participants tend to speak about themselves as individual persons and in relation to their experiences as well as the experience of their households. The individual person is the unit of analysis. The participants are conveying things they know happened to them or their households.

    - Key informant interviews—Participants tend to speak about other people, what they believe is the experience of those other people, and the community more broadly (as opposed to speaking about themselves as individuals and what they actually experience). The individual person is not the unit of analysis—there is no unit of analysis because the participant is conveying things they believe to be true.

    - Focus group discussions—Participants tend to speak about both themselves and other people. As result, the unit of analysis is mixed, in that participants convey things they know happened to them or their households and things they believe to be true about other people or even about society more broadly.

Across the 32 evaluation reports, the use of the phrase "unit of analysis" does not feature prominently. Readers of a report usually can deduce the unit of analysis, but discussion is lacking. In the statement below, for example, the explanatory text makes a unit of analysis problem apparent in listing each grouping of people interviewed. Each group is relevant to the data collection strategy; however, each offers insights from different perspectives. The explanatory text does not touch on this variation in unit of analysis.

The XXXX Project involved several groups whose activities significantly affect the final results of the interventions. At the core of the project, although unreachable, is the target population who are the beneficiaries of project interventions - People at risk of AIDS, People Living with AIDS (PLWA), Orphans and vulnerable children (OVC) and their families. Second, we have the implementation partners whose activities are key to achieving results. Third, we have the health managers at local, district and regional levels. Fourth, we have the country level managers directly involved in guiding project interventions and operations. And finally, we have the XXXX management and staff who have implemented the enhancement capacity efforts and the USAID officers who monitor and guide the entire process. (P25)

In comparison, the statement below explicitly uses the phrase "unit of analysis" to signal the different perspectives across the different groups of people interviewed.

Three main units of analysis were selected: first, national level leaders who had received training and other inputs from XXXX; second, the district level, including the XXXX unit of local government and local level civil society organizations; and third, HIV/AIDS affected Households, as identified from People with HIV/AIDS (PHA) registration lists. (P17)

Decisions about mapping the words, phrases, and associated techniques common to quantitative-oriented data collection and analysis onto qualitative-oriented data collection and analysis benefits from a case-by-case approach. In each case, it is important to recognize all factors that influence the choices evaluators make and consider the possibility that quantitative words and phrases sometimes align well to qualitative work, and sometimes overly affect the thinking and approaches surrounding qualitative work. A key consideration revolves around the extent to which the upsurge in mixed-methods studies may have resulted in an exercise of "mapping quantitative onto qualitative" in unchecked and perhaps problematic ways.

Addressing Gender with Qualitative Research Methods in Evaluations

Gender plays a considerable role in the HIV epidemic. Across the 32 evaluation reports reviewed, each project carried a focus on gender to some degree. Seventeen of these reports reviewed included an evaluation or discussion of project activities that explicitly addressed a gendered dimension of HIV/AIDS (e.g., gender-based violence, male participation in prevention of mother-to-child transmission [PMTCT], power imbalances in sexual relationships). However, none of the projects was specifically designed as a gender-integrated or gender-transformative project, [5] and none of the evaluations appear to have undertaken a specifically gender-sensitive approach.

Qualitative research methods can be valuable in addressing gender in evaluations. Whether described as engendering an evaluation or taking a gender-sensitive approach, the intent is for all phases of an evaluation to account for the ways and degree that project activities and outcomes have potentially been shaped by gender.

Gender "measured" qualitatively

Gender is a complex construct deeply influenced by power dynamics and systemic inequities. Qualitative research methods enable focus on the how and why, and the intricacies and pervasiveness of that power.

[5] With the rollout of a Gender Equality and Female Empowerment Policy in 2012 and the Automated Directives System (ADS) Chapter 205 in 2017, USAID instituted a requirement that all of its projects integrate gender. Across the 32 evaluation reports reviewed, 23 were published before or in 2012; 9 were published in or after 2013. Thus, determining which of the projects being evaluated adhered to USAID's gender-related requirements is not feasible, nor was it the focus of this review.

Toward Gender-Sensitive Evaluations

Understanding and committing to definitions plays an important role in developing a gender-sensitive evaluation approach. The starting point is recognizing that gender and sex are not the same, and that addressing gender can be part of any project or evaluation. Other related definitions include the following:

- **Biological sex** is our anatomy as female, male, or intersex. It includes our internal and external sex organs, chromosomes, and hormones (USAID, 2012).
- **Gender** is a socioculturally defined set of economic, social, and political roles, responsibilities, rights, entitlements, and obligations associated with being female and male, as well as the power relations between and among women and men, boys and girls (USAID, 2012).
- **Intersectionality.** The definition and expectations of what it means to be a woman or girl and a man or boy, and sanctions for not adhering to those expectations, vary across cultures and over time, and often intersect with other factors such as age, class, race, religion, and sexual orientation (World Health Organization, 2009).
- An individual who is **gender nonconforming** may have a different gender identify or gender expression than their biological sex. Some countries, such as Nepal, have a legally classified third gender (USAID, 2014b).
- The **gender continuum**, developed by the Interagency Gender Working Group, is a conceptual framework helpful in guiding the design and implementation of programs. The framework is grounded in the assertion that programs are either gender blind or gender aware in the extent to which gender norms, relations, and inequalities are addressed. Gender awareness consists of three categories along a spectrum (IGWG, 2017).
    - ***Gender-exploitative*** refers to intentionally or unintentionally strengthening or exploiting gender inequalities, or taking an approach that exacerbates inequalities.
    - ***Gender-accommodating*** refers to working around gender differences and inequalities. Such an approach may result in short-term benefits and realization of outcomes; however, this approach does not work to reduce gender inequality or address the gender systems that contribute to differences and inequalities.
    - ***Gender-transformative*** refers to striving to examine, question, and change underlying conditions to reach health and equity objectives.
- **Gender-sensitive evaluation** is an approach that commits to recognizing gender as a social and cultural construct as part of ensuring that an evaluation considers both the existence and effects of gendered roles, inequities, and structures. Undertaking a gender-sensitive evaluation requires examination of gender needs, interests, and dynamics, as well as the disparities in opportunities, resources, and power organized by gender (USAID, 2014a).

Summative Writing on Gender

Sex disaggregated data provide added value

Differences seen in data by sex provide insights into potential gender differentials and gendered impacts (e.g., fewer HIV-positive women accessing HIV treatment services than HIV-positive men).

In following the definitions of sex and gender, data collection and disaggregation are each distinguished by sex, not gender. The aim is to collect and analyze data by biological designation (sex), not based on a set of culturally defined roles, responsibilities, rights, and entitlements (gender). Across the 32 evaluation reports reviewed, a full understanding of and commitment to the difference between sex and gender appears to be lacking more often than not. For example, in the statements below, the first one incorrectly refers to data disaggregated by gender, whereas the second correctly refers to data disaggregated by sex.

> *Data analysis is an integral component of the entire evaluation process. It is envisioned that the Evaluation Team, after conferring with USAID/XXXX will develop and present a work plan and a design matrix. This plan should detail how both quantitative and qualitative data will be used to do the evaluation. Data will be disaggregated by gender in all applicable areas and into the three zones where IPs worked, both whenever possible. (P3)*

> *Gender mainstreaming is one of the key approaches used by XXXX in all its programming. At a program level there are gender focal people who do internal mainstreaming and ensure that at the programs/interventions level the gender considerations are addressed. M&E data is sex disaggregated and the data is [sic] routinely reviewed to ensure that there are no barriers created through the different interventions. (P7)*

The word "gender" appears often across the 32 evaluation reports, largely in relation to inclusion of a summary of the project's gender component. It is noteworthy that the correct definitions of sex and gender were applied unevenly. For example, the first statement below demonstrated a correct understanding of the definition of gender, whereas the second confused sex and gender in writing: "implemented with both genders."

> *The Theory of Gender and Power posits that women have difficulties seeking self-protection from HIV because of gender-based power imbalances. Their efforts in engaging in safer sex are often influenced by their feelings about their partners, socialization of women as sexually passive, regulation as to how women should express their sexuality, and econo-power factors. In a country like XXXX where marriage and children are still important social achievements, women may strive to stay in a risky relationship to avoid the social costs of being unmarried and childless. Therefore, the intervention included a strong gender component such as incorporating cultural values, gender dynamics, and attitudes that may leave one vulnerable to HIV as a result of such norms. (P12)*

> *The XXXX program gender equity in HIV/AIDS prevention activities could also prove valuable in other gender-oriented health and development programs, especially in reproductive health, safe motherhood, education, food security, civil society, and gender-based violence. Some activities, such as youth clubs, open days, Hope Kit, and XXXX campaigns, are implemented with both genders. These are sometimes further broken down by age and developmental stage, such as young adolescent girls (Girls Congresses, XXXX) or adult men and women (Garage Parties). Some are appropriately directed more toward women, e.g., Have a Healthy Baby PMTCT activities. Some activities directly address gender roles, including the gender packet in the Hope Kit and African Transformations; others are interwoven in the overall activity. (P26)*

Several of the evaluation reports included a brief assessment of the project's gender integration strategies, as well as recommendations for strengthening these activities, as in the following example:

> *A review of the Health Service Providers' Training Module during the final evaluation indicates that while there are some useful and necessary topics included in the contents (Social and Ethical issues in HIV/AIDS and Documentation and follow-up), the module can be further enhanced by including: (a) a gender perspective – e.g. what are the barriers to women's treatment seeking for RTI/STI, how can these be addressed, gender power relations in STIs and including these points in counseling, (b) how to evaluate referrals for medicine availability and quality of care issues, and (c) reasons why condoms are not used and how to address these… (P11)*

These type of statements, however, do not constitute a gender-sensitive evaluation. Evaluations may address how well gender-sensitive strategies addressed gender norms or gendered behaviors, but such references do not make an evaluation gender sensitive.

Recommendations for Gender-Sensitive Evaluation

Regardless of whether the project under evaluation has a strong focus on gender, a gender-sensitive evaluation is possible, and is increasingly the gold standard. Gender-sensitive projects or those that have gender-transforming objectives logically would evaluate and articulate these aspects of the project in the final report. However, it is equally important for projects that do not integrate gender to have gender-sensitive evaluations. In fact, projects that are not gender sensitive (or gender blind or exploitative) can potentially unintentionally reinforce harmful gender norms. For example, a PMTCT program that does not address women's fear of abandonment if testing positive for HIV risks low uptake by women. An HIV prevention program that promotes condom use by playing up male sexual prowess or dominance reinforces the harmful gender norm that men control women's sexuality, and specifically condom use.

Articulate a gender-sensitive approach

Without an articulated gender-sensitive approach outlined in a report, it can appear that the evaluation design did not fully consider all of the specific information needs and vulnerabilities of the stakeholders of the project being evaluated.

Assessing the gender components of a project is an important element in the write-up of a gender-sensitive evaluation. A gender-sensitive evaluation must consider the information needs and vulnerabilities of women, men, gender nonconforming people, and vulnerable subgroups, as well as describe their roles in conceptualizing and implementing an evaluation, and in interpreting and using data from evaluation activities. Equally important as the *what* part of the assessment is the *how* part. A gender-sensitive project accounts for the differing needs of men, women, boys, and girls.

Developing an evaluation approach begins with identifying the evaluation questions. Engaging stakeholders in the evaluation approach and design helps ensure the integration of gender and the overall appropriateness and quality of the questions. Each stakeholder group may have specific questions about the activities or outcomes of the project, which may differ in important ways from those of the project implementer or donor. In addition, as part of the process to define the evaluation questions, one overall aim is to tailor the questions, and the methods to answer them, to address the specific gender context in which the project under evaluation took

Gender and evaluation questions

Evaluation questions in a gender-sensitive evaluation will include whether there were any differential outcomes by sex and age.

place. Consideration of context helps ensure that an evaluation is fit-for-purpose. Evaluation questions typically focus on a project's strategic objectives and intermediate results, including establishing achievements by the project or intervention. A gender-sensitive evaluation goes beyond this focus by applying a gender lens to the more standard evaluation questions and including context-specific gender questions.

A gender-sensitive evaluation necessitates recognizing context-specific vulnerabilities. Often, gender-sensitive evaluations take an inclusive and fluid approach to gender while also recognizing and addressing intersectionality. As a socioculturally defined set of roles, responsibilities, rights, entitlements, and obligations, gender as an identity and marker of difference intersects with other markers of difference and vulnerabilities, such as age, disability, ethnicity, race, religion, and sex. Recognizing intersecting identities and vulnerabilities is also important when developing a data collection plan, including consideration of both who is recruited as interview and focus group discussion participants as well as who will conduct the interviews and focus group discussions.

Gender and data collection methods

Consider the sex of the interview and focus group discussion participants, the sex of who will conduct the interviews or focus group discussions, and what will work best.

Gender expert on the team

Not every evaluation team can include a gender expert. If possible, a local gender expert would be a valuable inclusion, both overall and in relation to data collection tools, site selection, and participant recruitment.

In sum, conducting a gender-sensitive evaluation is different from evaluating a gender-integrated program. A gender-sensitive evaluation is suitable for any type of project regardless of whether that project integrated gender. The international development community increasingly recognizes the benefits of evaluations that adopt gender-sensitive approaches.[6]

Limitations of the Documents Reviewed

Part of the 2011 USAID evaluation policy was a call to improve the quality of evaluations, including the provision of more guidance and oversight from USAID. The push for improvement emerged in a variety of ways. One example was the 2012 U.S. Government Accountability Office (GAO) report on PEPFAR, which was critical of evaluation quality (GAO, 2012). Based on the 2011 policy, the breakdown of the 32 USAID-funded HIV/AIDS-related evaluation reports reviewed is 23 from 2003–2012 and 9 from 2013–present. Logically, the quality of evaluations possibly improved following implementation of the new policy; however, this study did not seek to make such a comparison.

Across the 32 reports reviewed, 13 are described as impact evaluations. It is important to note that the focus of the document review and the study overall has not been a detailed examination to determine whether the 13 selected impact evaluations do or do not truly qualify as rigorous. Instead, the study team applied a broad definition of impact evaluation, largely aligned to how the authors of the report described an evaluation.

[6] Advocacy and instructions for how to undertake gender-sensitive evaluations are plentiful. For example, USAID has released a How-To-Note entitled *Engendering Evaluation at USAID* (October 2016). Appendix 5 provides a list of additional similar literature.

CONCLUSIONS AND RECOMMENDATIONS

The aim of the Measuring Impact Qualitatively study has been to provide strategies for evaluators undertaking impact-focused evaluations. The review of the 32 evaluation reports and the study overall reveal that the planning and implementation associated with the inclusion of qualitative work in evaluations often takes hold in quick and perfunctory ways. As a study, Measuring Impact Qualitatively has endeavored to be thought provoking and advocate a challenge to the status quo.

The title—Measuring Impact Qualitatively—is largely a play on words. Thus, rather than discussing measurement per se, the study has worked to broaden the conversation and draw out insights about evaluative work in terms of conceptualizing, understanding, and assessing impact. As result, advocacy to "measure impact qualitatively" is less about establishing a new approach and much more about enhancements and recognizing that evaluations have long sought to conceptualize, understand, and assess impact in qualitative terms.

The conclusion that emerges from this study is straightforward. The number of evaluations commissioned by USAID is steadily increasing, as is USAID's interest in improving the quality of evaluations. An important point of action amidst this steady increase involves recognizing the ways that "measuring impact qualitatively" has long been part of most evaluations, and opening up discussion about possible improvements for "measuring impact qualitatively" within end line performance evaluations.

The recommendation that emerges from this study also is straightforward. The way forward involves ensuring that (1) the use of qualitative research methods is thoughtful and attentive to detail, explanatory text, and transparency; and (2) evaluations appropriately consider gender. What is needed is not complicated; this paper presents logical, common sense suggestions based on reading documents and thinking through choices, strengths, and weaknesses.

Table 8 summarizes the conclusions and recommendations from the Measuring Impact Qualitatively study by providing tips for measuring impact qualitatively.

Table 8. Tips for measuring impact qualitatively

Strategies for establishing an evaluation approach
▪ Ensure that an assertion of mixed methods is not simply the use of a phrase but instead reflective of a mixed-methods approach consisting of in-depth and detailed thought and execution. ▪ Do not conflate design and method. Specifying an evaluation design—either separate from or in a methodology section—can be a constructive way to convey a concise overview of the evaluation. ▪ Avoid equating methods to lists of activities, disclaimers, protective statements, or circular logic. Consider and share the details, nuances, and logic for the methodological choices made. ▪ Evaluations contain fit-for-purpose elements; explain these elements. Fit-for-purpose elements relate to budget, time, data availability, data logistics, data quality, programming, and subject matter.
Traps to understand and avoid
▪ The phrase "qualitative data collection" is not specific. Consider and write out the specifics surrounding the data collection strategy and approach. ▪ "Key informant interview," "focus group discussion," and" in-depth interview" are not specific terms. Ensure that readers know what is meant by key informant. Describe how and why a focus group discussion was convened. Explain what occurs that makes an in-depth interview different than an interview. ▪ There is no evaluation automatic pilot button to switch on. Discuss and make plans around qualitative research methods options both early and throughout the evaluation. Make evaluation-specific decisions rather than risk thinking of key informant interviews as the only option. ▪ Maintain transparency in recording, storing, managing, and analyzing data from interviews and focus group discussions. Keeping everything in the evaluators' heads is not a good approach. ▪ Be precise about word choices and avoid simply mapping quantitative onto qualitative.
Addressing gender with qualitative research methods in evaluations
▪ A gender-sensitive evaluation is suitable for any project and increasingly is the gold standard. ▪ Know and commit to definitions in relation to sex, gender, and intersectionality. ▪ In designing and planning for an evaluation, the gender continuum is a useful tool and framework. ▪ Plan in relation to the sex of those who will be interviewed and those who will conduct interviews. ▪ Data are to be disaggregated by sex (not gender). ▪ Design the evaluation and use methods to give attention to what is complex about gender.

REFERENCES

Denizen, N. K. (2010). Moments, mixed methods and paradigm dialogs. *Qualitative Inquiry,* 16 (6), 419-427. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1077800410364608.

Interagency Gender Working Group (IGWG). (2017). *The gender integration continuum: Training session user's guide*. Washington, DC: Population Reference Bureau. Retrieved from https://www.igwg.org/wp-content/uploads/2017/12/17-418-GenderContTraining-2017-12-12-1633_FINAL.pdf.

International Cooperation Administration. (1961). *The act for international development: A program for the decade of development.* Washington, DC: International Cooperation Administration. Retrieved from http://pdf.usaid.gov/pdf_docs/pnaax630.pdf.

National Conference on International and Social Development. (1961). *A decade of development.* A report on the Eighth National Conference on International Economic and Social Development, Washington, DC, June 15–16, 1961. Retrieved from http://pdf.usaid.gov/pdf_docs/pcaaa615.pdf.

Reynolds, H., & Curtis, S. (2017, March 24). Whither the impact evaluation? [Blog post]. Retrieved from https://measureevaluation.wordpress.com/2017/03/24/whither-the-impact-evaluation/.

U.S. Agency for International Development (USAID). (2011). *Evaluation: Learning from experience.* USAID Evaluation Policy. Updated October 2016. Washington, DC: USAID. Retrieved from https://www.usaid.gov/evaluation/policy.

U.S. Agency for International Development (USAID). (2012). *Gender equality and female empowerment policy.* Washington, DC: USAID. Retrieved from https://www.usaid.gov/sites/default/files/documents/1865/GenderEqualityPolicy_0.pdf.

U.S. Agency for International Development (USAID). (2014a). *Gender-sensitive evaluation and monitoring: Best and promising practices in engendering evaluations.* Washington, DC: USAID. Retrieved from https://dec.usaid.gov/dec/content/Detail_Presto.aspx?ctID=ODVhZjk4NWQtM2YyMi00YjRmLTkxNjktZTcxMjM2NDBmY2Uy&rID=MTkyNzE5.

U.S. Agency for International Development (USAID). (2014b). *LGBT vision for action: Promoting and supporting the inclusion of lesbian, gay, bisexual, and transgender individuals.* Washington, DC: USAID. Retrieved from http://www.usaid.gov/sites/default/files/documents/1874/LGBT%20Vision.pdf.

U.S. Agency for International Development (USAID). (2016). *Strengthening evidence-based development: Five years of better evaluation practice at USAID 2011–2016.* Washington, DC: USAID. Retrieved from https://www.usaid.gov/documents/1870/strengthening-evidence-based-development-five-years-better-evaluation-practice-usaid.

U.S. Government Accountability Office (GAO). (2012, May). *President's Emergency Plan for AIDS Relief: Agencies can enhance evaluation quality, planning, and dissemination.* Washington, DC: GAO. Retrieved from http://www.gao.gov/assets/600/591284.pdf.

World Health Organization (WHO). (2009). *Integrating gender into HIV/AIDS programmes in the health sector.* Geneva, Switzerland: WHO. Retrieved from http://www.who.int/gender/documents/gender_hiv_guidelines_en.pdf.

APPENDIX 1. ANSWERS TO THE WEBINAR QUESTIONS

The webinar took place on October 29, 2015, and was given by Susan Pietrzyk, the lead researcher for this study. Overall, 585 people registered for the webinar, and 308 attended.[7] The transcript of the webinar shows well over 100 questions. It was not possible to address all of the questions during the webinar. Indeed, the volume and thoughtfulness of the questions overwhelmed me. After 30 minutes of trying to answer the questions, I ended the webinar with the promise to provide a written answer to every question. This appendix is an attempt to fulfill my promise. In addition, taking the time to write an answer for each of these questions proved to be a useful exercise as I worked on writing this paper.

Not surprisingly, because these were questions posed during a webinar, the writing is informal in style. In turn, I chose to respond in an informal style and generally adopted the terms used in the question in my answers. I have grouped the questions thematically as follows: clarifying questions about the document review, big picture questions, conceptualization of impact, expertise and evaluation methods, qualitative research and evaluation methods, next steps, and requests for reference.

Clarifying Questions About the Document Review

Question #1: I'm a bit confused about what documents were included; are you classifying annual reports and the like as evaluations?

No, I did not classify annual reports and the like as evaluations. When I combed through hundreds of evaluation reports on the DEC, it was useful when reports had descriptors in the title, such as annual report, meeting proceedings, quarterly report, etc., because this became a quick way to eliminate reports from consideration for this study.

Question #2: Did you review appendices?

Across the 32 evaluation reports reviewed, most have numerous appendices. The average page length is 106 pages, including appendices, and 52 pages, excluding appendices. Generally, the 32 evaluation reports were required to adhere to a USAID 50-page limit—thus, the likely reason for, on average, 50 pages of appendices.

I looked at the appendices but I did not review them to the same extent I reviewed the body of the report. I thought a lot about this decision. To my mind, I believe that one should not assume or accept that a reader is going to read appendices; therefore, any information the reader needs to read should be in the body of the report. Unfortunately, detailed discussion of qualitative thought and method is something that often ends up relegated to an appendix. I do not agree with this approach. For example, across the 32 evaluation reports, I think there were too many instances where the body of the report had a sentence or two essentially saying, "We did some interviews and focus groups, if you want to know more about this the details are in an appendix." I am of course exaggerating, but still, I find that disconcerting when the details surrounding the qualitative thought and method are not included in the body of the evaluation report. Imagine if an evaluation report included a sentence that essentially said, "We did a household survey, if you want to know more about this, the details are in an appendix." It would be unheard of to do that, but somehow it is ok to relegate the details of the qualitative thought and method to an appendix.

[7] The recording and PowerPoint slides from the webinar can be found here: http://www.cpc.unc.edu/measure/resources/webinars/measuring-impact-qualitatively.

Question #3: Thank you! Maybe in the list you provide, you can give a rating on which are best.

I would not be comfortable rating the 32 evaluation reports reviewed in this study. I would be comfortable in saying that I found strengths and weaknesses in all 32 reports. My objective with this study is to tease out what I see as strengths and weaknesses with the hope that my assessment will help others in terms of the numerous choices and decisions made when undertaking an evaluation.

Question #4: Did you assess the quality of the qualitative methods? Did they go beyond summative findings to look at linkages, etc.?

Yes, in reviewing 32 evaluation reports I assessed the quality of the qualitative methods to a degree. In some instances, if forced to take a stand, I would say that I found the quality to be not so good. That said, I purposefully have not gone down a path of "this is right and this is wrong," as if there is only a single approach to qualitative methods. I have specific opinions about qualitative methods, but other people have great (and different) opinions about qualitative methods.

I tend to think about the review and the assessment as looking for strengths and weaknesses. When I find elements of an evaluation that I see as weak, I tend to give the benefit of doubt. I try to frame the assessment in terms of thinking that any given choice might have been for a specific reason, but if not forced into such a choice or now with the benefit of time and hindsight, other options might have included x, y, or z.

I am not exactly sure I understand summative findings in comparison to looking at linkages, etc. However, I suspect that some of the question relates to the quality of the writing (of the evaluation report). To me, one key objective of an evaluation is to provide summative findings (e.g., an executive summary). I would think that well-written summative findings would be attentive to the nuances, layers, linkages, etc., just as detailed findings (the body of the report) would also be attentive to digging deep. In addition, as I think is implied in the question being posed, the data collection strategies (whether qualitative, quantitative, or mixed) need to be robust and flexible enough to ensure the data can actually draw out what the linkages are, recognizing that, going into the evaluation, those linkages might not be known.

Question #5: Did you look at PEs that were quantitative?

The short answer is yes, I reviewed performance evaluations (PEs) that used quantitative methods. Of the 32 evaluation reports selected for review, about 15 used quantitative methods.

The longer answer is imagine yourself pulling hundreds of evaluation reports from the DEC (2003 to the present). One of the first things you would notice is that USAID does not have uniform titling conventions, nor is there much consistency in word choice around "evaluation" and "assessment." Likely, perfect uniformity would not be possible. Again, think of all the possible titles and word choices for an evaluation. For example, performance evaluation, end-of-project evaluation, final evaluation, impact evaluation, project evaluation, plus more, and plus all of those with "assessment" instead of "evaluation." Certainly, since 2012, when the new USAID evaluation policy took effect, there has been an effort to denote what is a performance evaluation and what is an impact evaluation. However, the waters are still quite murky.

In any event, my point is that I tried not to go overboard with thinking I have to categorize each of these 32 evaluation reports as this or that. My objective was to select 32 evaluation reports that could help shed light on how impact is being conceptualized and examined in evaluation work and how qualitative methods, quantitative methods, both methods, and mixed methods factor into things.

Question #6: Tx for this lesson ... from your experience on this study, when did you realize that your research will have an impact on QR?

When I spent a few days to read hundreds of USAID evaluation reports, I was quick to think that everyone involved in international development as well as monitoring and evaluation should spend a few days on such a reading exercise. Most of what I am advocating is not rocket science. I am making logical, common sense suggestions based on reading documents and thinking through strengths and weaknesses. Further, I believe that my common-sense suggestions and the common-sense suggestions of others would go a long way, as noted above, to improve the standards and quality related to qualitative inquiry, as well as aid in efforts to apply qualitative methods more systematically to assess impact.

Big Picture Questions

Question #7: What is the difference between "experience" and experiential evidence?

In general, I advocate being attentive to the thought and decision-making process associated with conceptualizing and approaching evidence. I will elaborate in relation to the question posed.

First, experience in and of itself is not evidence. What can be considered evidence (experiential evidence specifically) is the combination of documenting and understanding an experience, and putting that experience within the context (what, where, when, how, and why) in which the experience took place. Knowing what an individual person (e.g., project beneficiary, stakeholder) experienced is an important aspect of accumulating and understanding evidence. However, an individual experience is just that—an individual experience. The context intertwined with any individual experience is just as important. Indeed, many internal and external factors can and do shape and make the experience what it is. Thus, it is to say that (experiential) evidence = experience + context.

Second, just the term "evidence" does not really convey all that much. I like to know more details regarding the use of the term "evidence"—the why and the how in particular. That is, the type of evidence, why and how it qualifies as evidence, why and how data were collected, why and how the data went from data to be established as evidence, and so forth. In this sense, experiential evidence is a type of evidence. The phrase gives an indication that the data stem from an effort to understand the experience of an individual (likely by using qualitative research methods and open-ended questions and conversation).

For example, during an interview a question is posed to an individual (such as, "What has been your experience in visiting clinic x?"). An answer given about their experience blended with the context of that answer I would consider experiential evidence. With evidence that draws from experiences, I also think it is important to be fully cognizant of who the interview is with. For example, say an interview is conducted with a nurse at clinic x. The nurse is asked a comparable question ("What types of experiences do patients have at clinic x?") The nurse's response is potentially relevant and useful; however, what the nurse feels is the patient's experience is not necessarily the same as what the patient feels about the experience.

Question #8: Would you please say something about the intimate relation between theory (not what USAID means by theory of change) and methodology?

I love this question and, to me, this is an under-discussed topic. Theory is important. Methodology is important. The balance of theory and methodology is extremely important. My approach, I would say, is such that I read and think about theory more in comparison to methodology, and my favorite and go-to theorists are Michel Foucault (sex and power), Chandra Mohanty (feminisms), and Joan Scott (experiential evidence).

In qualitative work, context is crucial, and it is generally best to think about context in holistic as well as interrelated and intersectional ways. By context, I mean quite a few things, such as dynamics, mitigating factors, systemic barriers, constructed forms of difference, structural barriers, geopolitics, belief systems, customs and ways of being, and this list could get quite long. My point is that without some sort of engagement with theoretical ideas, any methodology runs the risk of forgetting about context and may well end up rather hollow and isolated, likely with a lack of attention to complex realities and little nuance.

Question #9: My confusion around "qualitative" evaluations is with the word "qualitative." In this webinar, here are four ways the word is used: Qualitative method of analysis; Qualitative data collection (methods); Qualitative data; Qualitative expertise and Qualitative work. These seem to be different things, although there may be some overlap. I tend to think about "qualitative" as with "qualitative data," within which there are these categories: nominal, categorical, and numeric (integral or decimal). For me a "method" that uses qualitative data is a qualitative method.

The webinar PowerPoint slides and the report purposively include extensive use of various "qualitative" words and phrases. Part of finalizing the report will include shoring this up to ensure consistency; however, I am a firm believer that there are no singular "qualitative" words or phrases, so my list is not going to get shorter. Indeed, I do not see the phrases noted in the comment as interchangeable. My intent in using these various phrases is to signal different things. To my mind, it is quite important to use precise phrases and avoid laziness with words. Certainly, all-encompassing phrases are sometimes necessary, and in those situations, my preference is "qualitative study" or "qualitative work." I like these two because they signal that qualitative focus is occurring at each phase of an evaluative activity (design, data collection, analysis, and writing). I like the second of the two because it reminds a reader that whether it be qualitative thinking, qualitative data collection, or what have you, it is a lot of work. I tend to avoid the term "qualitative research" because evaluation and research are not synonymous, and generally USAID projects and evaluations do not exactly conduct research.

In only reading this comment and not hearing it during the webinar, I am not sure of the intent or any possible embedded questions. The comment does remind me that I feel that choice, precision, consistency, etc., in relation to words and phrases is an under-discussed topic, and the reality is that there is too much weak writing across a lot of USAID evaluations and reports.

For example, I find it odd and maybe even a bit confrontational when a scope of work or a report title includes "evaluation," preceded by a type of method. More specifically, I find these terms problematic: qualitative evaluation, mixed-methods evaluation, quantitative evaluation. I think the most important things to focus on in a title are the appropriate evaluative descriptors (e.g., baseline, mid-term, end line, final) along with the country, project, and topic. The reader will learn of the evaluation methods by reading the report.

Categorizing evaluations by method type in the title seems a dangerous path of giving space for people to think one type of method is better or worse than another.

I do not understand the last part of this comment; in particular, I do not understand categorizing qualitative data as nominal, categorical, and numeric. I see those categorizing terms as the words of a statistician and I really like to avoid simply mapping the words and methods of statistics onto qualitative thinking and qualitative work.

Question #10: I wanted to ask Susan about her slide where she suggests that complexity was a focus in the 90s, but that's been replaced by more structured/quantitatively oriented paradigm. I am actually seeing the opposite—it seems like with a focus on systems and complexity science, we are in a bit of a "complexity renaissance." I'd be interested in Susan's perspective.

I used the words "complex" and "complexity" as descriptive adjectives. More specifically, I indicated that by the early 1990s, the complexity and severity of the HIV epidemic was more recognized and acknowledged. I was not and I cannot imagine that would I ever refer to complexity as a paradigm, field of study, science, and so forth.

The paradigm shift that interests me, and what I tend to think somewhat critically about, are the ways that PEPFAR, for better and worse, played a major role in ushering in a new force where data, ideas, intervention, approaches, etc., have to be quantitatively proven and then quantitatively proven a second time to be considered worthy of attention. Many within PEPFAR, USAID, other donors, and beyond have, in my opinion, become overly obsessed with quantitative metrics to the point of counter productivity.

Question #11: Would you comment more on how gender issues relate to qualitative assessment of impact?

As a beginning note, I am of the mind that every member of an evaluation team should have at least some understanding of gender as well as gender in the context of the location of the project being evaluated. Further, I believe that every evaluation team should specifically have a gender expert, and gender issues relate to every evaluation.

The starting point with this question is to me recognition that impact is a subjective concept. To me, this is fundamentally immensely important to recognize, and given the impact evaluation trend, has increasingly become problematically forgotten. Put another way, although impact can be a quantitative measurement of change over time, it is important to remember that impact is something that people feel. Impact is messy and complicated. In the context of feeling impact, then, each individual conceptualizes and experiences impact in different ways. Each different conceptualization and experience relates to and intersects with a great many factors, including biological sex, gender norms, and gendered dynamics. As such, whether impacts are negative or positive can also relate to and intersect with sex and gender factors.

An additional point with this question is to consider quantitative metrics disaggregated by sex. Analysis of this nature is imperative in an evaluation. Equally, it is important to remember the limits of quantitative metrics, and the ways that a qualitative assessment of impact, inclusive of a gender perspective, can explore the details and nuances surrounding how individuals conceptualize and experience impact differently.

Conceptualization of Impact

Question #12: Isn't it difficult to measure the extent of impact using qualitative methods? Considering qualitative studies has [a] limited no. of sample[s] as compared to quantitative?

The title of this webinar and the activity overall (Measuring Impact Qualitatively) is a play on words. The point I am making is that, technically speaking, impact cannot be measured because "impact" is a subjective concept. Impact certainly cannot be measured using qualitative methods because measurements are, by definition, quantitative metrics. In tandem then, my point is that in the current international development climate, increasingly everyone wants to "measure" this subjective concept called "impact." Thus, my goal is to broaden the conversation about all of the possible ways to conceptualize, understand, and assess impacts.

It is true that a bunch of people have created space for impact evaluations in international development by taking random control trial methodologies and adapting them to the context of development projects. I think this is really great. However, the point remains that impact is a subjective term, and for this very reason, conceptualizing, understanding, and assessing impact qualitatively, I feel, is quite important.

I would note that, to my mind, trying to compare sample sizes for quantitative studies and qualitative studies should be avoided. Quantitative studies will always win, and more importantly, qualitative studies really do not have samples. The term "sample" is a statistical technique, and it is applicable to surveys and other types of quantitative studies. It makes no sense to me to simply map the term sample onto a qualitative study. Instead, in qualitative studies, sites for data collection are purposefully selected and study participants are also purposefully selected. Thus, in a qualitative study, there is a site selection process (not a sample selection) and there is a participant recruitment process (not a sample drawn).

Question #13: Impact evaluations allow us to attribute causality through statistical methods.

Indeed, this seems a good sentence to describe what has emerged, to some, as the definition of an impact evaluation. I would note, however, that, "causality," like impact, is a subjective term, and there are real challenges with measuring subjective terms.

Question #14: I think that is venturing into dangerous territory to use the term "impact" loosely. As stated, there are a lot of definitions of impact. My understanding is impact implies an empirical method using an experimental design or quasi-experimental design.

I disagree, or at least I would advocate that impact is a noun or a verb relating to describing and assessing influence. Impact is also a subjective concept. Not all impacts are felt the same way; positive impact and negative impact cannot be universally defined, and so forth. The understanding noted is correct; however, a certain set of people decided that "impact" implies an empirical method using an experimental design or quasi-experimental design. Why do one set of people get to co-opt a word? I mean that as a serious question. In certain ways, every evaluation, regardless of method, is about understanding impact. Every evaluation wants to investigate success, process, challenges, did something work, did the intervention help people, did the project make things worse, and so forth. This is to say that in certain ways, every project evaluation boils down to one question: What impacts did the project have? Why do only method and only one set of evaluators get to redefine a word that has been around for centuries and then also be the only people who get to use the word?

Question #15: I think the problem is that the dictionary defines "impact" more broadly, so the majority of people use the word that way. Evaluators need to define it for their audiences whenever they use it. The responsibility for clarifying the use of the term is with the evaluator, not with the audience.

Good points. I would note that the dictionary is a credible source. In addition, as I noted in response to question 14 just above, I would advocate that all evaluators be allowed to use the word "impact." Any use should include explanation surrounding how the word is being used.

Question #16: I think that the meaning of impact that we use or fear today comes from the Center for Global Development paper in 2005, "the Evaluation Gap: When we ever learn"; after that and 3iE, the term impact has become confused.

I agree on all points; albeit I have strong opinions and would say the word "impact" has been problematically co-opted. Here is the link to this paper:

http://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation

Question #17: Are you suggesting that impact evaluations can be exclusively qualitative?

Not exactly. There is a specific type of impact evaluation that only uses quantitative methods. This is a sound approach and, like any evaluation, carries both pros and cons. I am suggesting those interested in this specific type of impact evaluation do not own the word "impact." Thus, if others would be allowed to use the word "impact," then yes, an evaluation can use primary qualitative methods to understand and assess impact.

Question #18: I didn't understand clearly how an impact of X intervention can be measured from either focus group discussion [or] in-depth interview?

Please note the responses above. The impact of X intervention cannot be quantitatively measured from focus group discussions or interviews; however, evaluators, researchers, practitioners, policymakers, etc., can learn a lot about the impact of X intervention from focus group discussions or interviews. In fact, I would suggest that what can be learned about the impact of X intervention will be much deeper, context specific, and more nuanced from focus group discussions or interviews compared to any quantitative metric. Just to be clear, I am not saying data from focus group discussions or interviews are better than survey or other quantitative data; rather, I am saying data from focus group discussions or interviews are different in useful ways.

Question #19: How do we measure the impact from point A to B qualitatively of a particular intervention? Thank you.

Please note the responses above. It is not possible to quantitatively measure the impact of a particular intervention from point A to point B qualitatively.

I would note that point A and point B are single points in time. The measurement methods used in an impact evaluation leave out the ups and downs and nuance that might have occurred in the years between point A and point B. Michael Woolcock of the World Bank explains this well in a graphic on page 13 of his article entitled Using Case Studies to Explore the External Validity of "Complex" Development Interventions, WIDER Working Paper, No. 2013/096. Here is the link:

https://www.wider.unu.edu/publication/using-case-studies-explore-external-validity-%E2%80%98complex%E2%80%99-development-interventions-0

Question #20: How can we deal with the issue of attributing the impact to a program when using only qualitative methods? Can this be done at all?

Please note the responses above. In addition, I think it is fair to ask, as many impact evaluators ask during the design phrase, how are we going to deal with the issue of attributing the impact to a program when using only quantitative methods? Can this be done at all? How do we know if we truly have a viable control group that has received no development programming? How do we truly know if the impact seen in the treatment group is not attributable to a program different from the one we are focused on?

Question #21: How can we work around the issue of attributing the change (impact) to the program with (just?) qualitative methods?

Please note the responses above. In addition, there are either no answers to this question or many answers to this question. It depends on the objectives of the evaluation. If the objective is to explore and investigate how program beneficiaries and other stakeholders conceptualized, understood, and, experienced change over time, including the positive and negative impacts of a program that was implemented, then qualitative methods can be a good way to undertake such an investigation. Data analysis and report writing can focus on describing some of the ways that the beneficiary experiences might be attributable to the program. If the objective is to use a specialized methodology for the purposes of coming up with singular quantitative metrics to measure and attribute change, then no, this cannot be done with qualitative methods.

Question #22: I believe you are suggesting here that there needs to be more attention to standards and quality related to qualitative inquiry, as well as applying qualitative methods more systematically to assess impact. If yes, can you list the top three areas of investment needed in improving standards and quality?

This is a very exacting question, and exacting questions make me nervous, particularly if the expectation is a definitive and singular answer. That said, yes, I reviewed hundreds of USAID evaluations and studied a purposively selected set of 32 HIV/AIDS-related evaluations very carefully. What struck me is that so many evaluations (regardless of method), perhaps even all evaluations, work to assess impact and yes, I found the quality with which qualitative methods are used to be a mixed bag.

A phrase like "areas of investment" is also very exacting and this, too, makes me nervous. Or perhaps I might suggest that "areas of investment" is more a political question. That said, here are three summative points that, in my opinion, would help to improve the standards and quality related to qualitative inquiry, and a fourth point about applying qualitative methods more systematically to assess impact.

1. Avoid simply spitting out like a robot that XX FGD, XX KII, XX IDI will be conducted. Instead, really take some time to think through what people to speak with and why those people.
2. Avoid the self-congratulatory evaluation methodology. Sure, USAID and project staff can help provide details about a project, but these people really are not evaluative key informants.
3. If a well-written report is one goal, allow a realistic amount of time to produce a well-written report, noting that two or three weeks is nowhere near a realistic amount of time.
4. Admit that impact evaluations are not always possible due to budget, time, and project design constraints, and allow those conducting qualitative studies to use the word "impact."

Question #23: We're implementing a psychological rehabilitation of victims in conflict situations, how can we measure the impact?

I am thinking of two ways to approach this question. If the objective is to use a specialized methodology for the purposes of coming up with singular quantitative metrics to measure impact, then a control group of victims in conflict situations who did not receive the psychological rehabilitation programming would be needed. That, to me, starts to introduce some questions about ethics. If, however, the objective is to explore and investigate how a group of victims in conflict situations who received a psychological rehabilitation program feel about the services they received, including the positive and negative impacts of a program, then qualitative methods can be a good way to undertake such an investigation. Data analysis and report writing can focus on describing and assessing some of the ways that the beneficiaries' experiences might be attributable to the program.

Expertise and Evaluation Methods

Question #24: There was an assumption in the presentation that one person cannot be both a qualitative analysis expert AND a quantitative analysis expert. Do you find that to always be the case?

Question #24a: I agree with the respondent, there are people who possess both expertise—quant and qual. I think the more important issue is that an evaluation team should include more than one expert.

If forced to give only a yes or no response, then I would say that no, one person cannot have expertise in both quantitative and qualitative methods. However, this is a complicated question and it warrants more than a yes or no response. A lot depends on the specific evaluation, the specific objectives of the evaluation, and the topical subject matter. A lot also depends on how "expertise" is being defined, and what level of expertise is needed. At the end of the day, I feel this question and staffing decisions tend to benefit from following a case-by-case approach in tandem with a person-by-person approach. The goal is a good and well-balanced fit.

In my experience, many people claim expertise in the use of quantitative, qualitative, and mixed methods; equally, many people have great appreciation for the value of quantitative, qualitative, and mixed methods. I think this is great, and there are lots of confident, experienced, savvy, and skilled people who conduct evaluations. That said, I feel that more often than not, evaluators are slanted toward one end of the spectrum. Put another way, when it comes to quantitative and qualitative methods, most people, again, more often than not, hold a predominant expertise and a secondary expertise.

Question #25: Thanks for the presentation, Susan. I have seen people present before and after qualitative evaluation studies. Would you like to comment on this?

As I understand the question, it seems to relate to all evaluations whether quantitative, qualitative, both, or mixed. To me, the question is: Who all should be involved in the design of and up-front discussion about any given evaluation?

If part of the evaluation approach is to be inclusive, transparent, and participatory, and to solicit input from stakeholders, then up-front discussion (including a presentation) would be important. Equally, a presentation of the results would also be important. However, not all evaluations take on a participatory approach to the same degree. There are good reasons for a high degree of using a participatory approach, just as there are

good reasons to have very little participatory approach. I would suggest that evaluation presentations before implementation as well as presentations after implementation are something to be decided on a case-by-case basis, and largely relate to the extent that a participatory approach is being pursued. I do not see how the single factor of being a quantitative-only or a qualitative-only evaluation would be a driving factor in a case-by-case decision about before and after presentations.

Question #26: I think that the real strength of this study is the mapping of quant and qual. Definitely encourage publication on that.

Thank you for the encouragement. I appreciate that.

I am not sure I understand the suggestion of a publication on the mapping of quantitative and qualitative. That said, I will use this opportunity to restate something I hope was clear in the webinar (and is in the report). I am leery of instances where quantitative methods, techniques, words, etc., come to be mapped onto a qualitative study. I would also not be in favor of the reverse; although, thinking about it, I cannot imagine any serious statistician or survey or quantitative expert ever wanting to map a qualitative method, technique, or word onto a survey, statistical technique, or quantitative study.

In the spirit of never say never, I do believe in exceptions. Below I have noted examples of when, to me, it makes no sense and may even be detrimental to map quantitative words onto qualitative work, as well an example where I feel it is advantageous to map a quantitative word onto qualitative work.

- Qualitative work does not include sampling; instead, there is a site selection process.
- Qualitative work does not include instruments; instead, qualitative work makes use of tools, such as interview guides, talking points, observation checklists, etc.
- Qualitative work does not include respondents, which I feel is a rather cold and closed word; instead, qualitative work includes participants, which I feel is a welcoming and participatory word.
- "Unit of analysis" (a traditionally quantitative phrase) can be useful to think about in qualitative work. In particular, use of this phrase is important in understanding that focus group discussions, key informant interviews, and individual interviews each represent a different unit of analysis.

Question #27: I would be interested to hear thoughts about "qualitative indicators," which many in public health don't believe in.

Note the response to question #26 just above. My thought is that indicators are by definition quantitative metrics as well as the outcome of quantitative data collection and quantitative analysis. "Indicator" is a word that rests in the domain of the quantitative world. I see no reason to map the word "indicator" onto the domain of the qualitative world. Sure, the quantitative and qualitative worlds can be mixed and that is really important, but indicators are by definition quantitative. As such, to me, there is no such thing as a qualitative indicator, and I would not be one to advocate for the use of the term "qualitative indicator."

That said, every quantitative indicator is unique, including its design, intent, calculation, etc. I can imagine that often, indicator experts as well as evaluators discuss the ways that some indicators are more qualitative leaning in comparison to others. One example of a qualitative-leaning indicator that comes to mind is an indicator where the calculation is a simple and one-step count of responses to a single pre-coded question. To me, I would think that qualitative methods were used to establish the pre-coded answers and, in a certain way,

qualitative information is present with such indicators; however, once respondents hear a closed question and once the information is counted and only counted, the information is no longer qualitative. Put another way, I have no objective to taking qualitative information and counting it, but I do object to qualitative information that is counted still being called qualitative information. My preference would be to take the count of the different responses to a single pre-coded question and use those data to aid in the design of a qualitative study.

Question #28: Good point. Performance evaluations are not necessarily only qualitative. They are often mixed methods, including "mini" surveys, direct observation, and secondary data analysis.

This comment makes me think about the phrase "evaluation methods" as well as all of the descriptors that can go before evaluation (e.g., performance evaluation, process evaluation, impact evaluation, final qualitative evaluation, baseline quantitative evaluation, and many more). During my research for this study, I was struck by the ways decisions about methods, and evaluations in general, are largely based on practicalities and fit-for-purpose thinking. With this realization about fit-for-purpose, several thoughts crystalized, such as the following:

- Each evaluation is unique, with a unique purpose.
- There is no master list (e.g., AA Evaluation = ZZ Methods, BB Evaluation = YY Methods).
- An evaluation scope of work is either clear, not clear, or somewhere in between.
- Each evaluation is either under budgeted, adequately budgeted, or somewhere in between.
- Each evaluation has too short a timeline, an adequate timeline, or somewhere in between.
- Secondary data are either available or not available.
- The quality of secondary data is not within the control of an evaluation team.

I could add more to the list above, but my point is that, as this comment seems to indicate, whenever an evaluation is undertaken, a certain set of people make decisions about data collection and data analysis methods, and a wide range of factors influence those decisions. What has been noteworthy to me in scanning hundreds of USAID evaluations and closely reading a purposefully chosen set of 32 HIV/AIDS-related evaluations is the frequency with which evaluation methods (particularly qualitative methods) appear to have been selected in formulaic and perfunctory ways.

Question #29: I think we are in an age of implementation research and process documentation. Would you agree that this increasingly requires more qualitative methods?

It seems to me that the terms "implementation research" and "process documentation" might be new phrases and part of a new age where these buzzwords are being acknowledged by different actors; however, implementation research and process documentation have each been around for a very long time; it's just that the work might have been called something else. It also seems to me that it is not the need for qualitative methods that is increasing; rather, what is increasing is better recognition by a wider set of people that there is need for qualitative methods.

In terms of implementation research and process documentation as currently pursued, I would answer in the way I answered other questions—I would replace "evaluation" with "implementation research" or "process documentation." Thus, my answer is, it depends. Those undertaking implementation research or process

documentation make decisions regarding data collection and data analysis methods that align to their goals. The question is well taken, however, particularly in the sense that it is hard to imagine implementation research and process documentation without in-depth and open-ended conversations.

I think it might be worth considering that implementation research and process documentation, in a way, each represent individual methods in and of themselves. When I think about implementation research and process documentation, one of the first things that comes to my mind is slick infographics. For example, numerous software is available specifically for process documentation. Of course, the software cannot yield anything if there are no data; however, I cannot help but think that, for better or for worse, it is attention to the infographic that drives process documentation more than attention to how the data are collected.

Qualitative Research and Evaluation Methods

Question #30: Could you provide some examples of units of analysis?

"Unit of analysis" is, to my mind, a term more commonly used in survey and quantitative research. The term refers to who or what is being analyzed (person, persons, place, thing, practice, belief, etc.). Even though "unit of analysis" is traditionally a survey and quantitative term, I think it converts well into the qualitative world. In particular, in collecting and analyzing qualitative data, I think it is quite important to be attentive to unit-of-analysis differences in two ways:

1. If qualitative work is aiming to complement a survey, it is important to understand the unit of analysis for the survey. For example, a population-based household survey and a beneficiary-based survey have a different unit of analysis. If interviews and focus group discussions are going to be conducted to complement a survey, it might be useful to align the unit of analysis for the interviews and focus group discussions to the survey (e.g., household or beneficiary).
2. Generally, three terms describe interview types. Each interview type tends to have a different unit of analysis, by which I mean the perspective or point view for each is different. To my mind, it is quite important to be attentive to these differences when analyzing and reporting findings from interview data. The three interview types are as follows:
    - In-depth interviews—Participants tend to speak about themselves as an individual person and in relation to their experiences as well as the experience of the household. The individual person then is the unit of analysis. The participants are conveying things that they know happened to them or their household.
    - Key informant interviews—Participants tend to speak about other people, what they believe is the experience of other people, and the community more broadly (as opposed to speaking about themselves as an individual and what they actually experience). The individual person then is not exactly the unit of analysis. In certain way, there is no unit of analysis, because the participants are conveying things they believe to be true.
    - Focus group discussion—Participants tend to speak about both themselves and other people. The data then consist of the unit of analysis being mixed.

Question #31: This is a comment rather than a question. I appreciate your point about really needing to call out methods used and being specific about who they were used with. Seems obvious, but your point is a good reminder of how important it is to be specific about how one went about doing research.

Well said, I could not agree more. My point above about unit of analysis and interview type I think is a good example of the importance of being specific. In addition, I made similar comments in response to question #6 when I noted that the majority of suggestions I am making are common sense more than they are rocket science. I would also note that I think the amount of time it takes to write out (and write out well) how one went about doing research is often underestimated. As I suggested in question #22, two to three weeks, a common amount of time allocated for an evaluation report, is usually nowhere near enough time.

Question #32: I understand what was said about the IDIs being under-represented and the beneficiary voice as critical. But aren't FGDs with beneficiaries equally as strong in collecting participant/beneficiary perspectives of the program?

I feel there is no universal answer to this question. With each evaluation, it is important to consider the pros and cons of various interview types, and then make choices accordingly. Often the choices are not a matter of what is right or wrong; rather, different contexts warrant different choices, and different people make different choices.

In many instances, focus group discussions with beneficiaries can be a productive and useful interview type. In some instances, given factors such as time, safety, logistics, etc., focus group discussions may be the only option. For example, in highly remote areas, it may not be feasible (or safe) to travel house to house to conduct interviews with individual beneficiaries. It may be safer (in relative terms) for a group of people to meet in a mutually agreed-on location for a focus group discussion.

If all options are available, I feel the cons of focus group discussions outweigh the pros. Not everyone will have the same opinion, but here are some of the reservations I have about focus group discussions:

- Focus group discussion participants tend to shift between speaking about their personal experience, the experience of their next-door neighbor, the experience of a distant (likely nonexistent) neighbor, and societal-level problems. To me, what is being said (the data) is in the world of being apples and oranges; put another way, this creates a unit-of-analysis problem. What an individual person conveys as his or her experience is a different type of information (data or evidence), compared to what an individual person thinks is the experience of someone else.
- More often than not, only a couple of people meaningfully contribute to focus group discussions. The others say nothing or, worse, are disruptive. Relatedly, focus group discussion participants are likely to let what others say influence what they say. I find that often, there is a dominant voice among the group, and the nondominant individuals are swayed by the dominant voice and sometimes simply restate what the dominant person is saying.
- Of course, all of these points can be flipped into positives. An argument can be made that, when in groups, people are more at ease and more comfortable in talking, so they say more. I am not one to all out dismiss focus group discussions, but I stand strongly by my first bullet—the unit of analysis is murky with focus group discussions

I recently read an interesting article, which asserted that in-depth interviews cost 20–36 percent less to reach thematic saturation, compared to focus group discussions. Here is the citation:

Namey, E., Guest, G., McKenna, K. & Chen M. (2016 April). Evaluating bang for the buck: A cost-effectiveness comparison between individual interviews and focus groups based on thematic saturation levels. *American Journal of Evaluation*. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1098214016630406?journalCode=ajec.

Question #33: Which is more important: in-depth, FGDs, or key information interviews? I prefer the first one, especially when there are conflict[s] of results, is that right? My question is because I get different answers about same problem, mostly from key information.

My response would be that it depends. With each evaluation, it is important to clearly establish the goals for the evaluation and then make choices about interview types that will best help to achieve the goals. I would also note that I am a firm believer in the logic of taking ownership of your choices, by which I mean always explain why a particular choice was made, and be specific in that explanation.

Question #34: What about participant observation?

The short answer is yes, observation can be an important research method or technique within an evaluation. I also would note that observation is a skill that requires thought and planning and, like most skills, one becomes a better observer with time and experience.

The longer answer is that, technically speaking, "participant observation" and "observation" are two different things. With the former, one would have to actually participate in what is being observed. For example, if the study is focused on experiences with HIV testing at clinic, to undertake participant observation would mean that the researcher would be tested for HIV at the clinic (as in, they would participate). Generally, I would say participant observation is more a long-term (ethnographic or anthropological) research method. In evaluation work, which is often a few weeks, maybe a month or six weeks of fieldwork, it is observation (not participant observation) that can be valuable. That is, for example, visit a clinic to interview those who work in the clinic, but also observe what the clinic looks like and what takes place. Generally, the evaluation team would develop some sort of observation checklist as part of designing the suite of data collection tools.

Question #35: Were there any examples of how direct observation can be used to understand the impact programs have on beneficiaries?

The short answer is no. I did not find any evaluation reports where I thought the use of direct observation as a data collection method was used particularly well as part of understanding the impact programs have on beneficiaries.

I would, however, note the following. I was surprised that across the 32 HIV/AIDS-related evaluation reports that I reviewed that only a few of them listed "direct observation" as a data collection method. To my mind, two aspects of this discovery seem noteworthy:

- I suspect direct observation is common in most evaluations—that is, unless evaluators purposefully close their eyes when undertaking an evaluation. What is less common is for evaluators to make the decision to formally and methodically document and analyze what they observe. I think this is a shame, as I tend to think observations can be quite telling, particularly with evaluations where part of the intervention or project involves the provision of health services in health facilities. I would

hazard a guess that lack of time and the need to follow specific report structures are each a big part of why documented direct observations are rarely interwoven into the analysis and evaluation reports.

- As much as I feel that direct observation is constructive, it also has its cons. It is important to note that direct observation, in my opinion, is not very strong as a single method—by which I mean that any direct observation, as a form of data or as a form of evidence, likely would need to be in conjunction with other data or evidence. The juxtaposition could be complementary or contradictory, but the juxtaposition needs to be there. For an evaluator to assert, "I observed X and therefore X is true," to me, is a statement resting on very thin ice.

Question #36: To what extent can we generalize findings of the qualitative research?

The short answer is: not at all.

The longer answer is to say that, in my opinion, no well thought-out qualitative study sets out to generalize the findings. The aim of qualitative work is to contextualize, offer possible explanations, provide insight, raise new and expanded questions, understand perceptions, look at behaviors, and so on. To generalize and offer a definitive or singular answer or analytic point is not the goal. This is not at all to say that generalizations are unimportant; it is, however, a reminder about what generalizations are and are not. When collected and analyzed in a statistically representative way, numbers can facilitate generalizations. Qualitative work does what numbers cannot do, which is to look beyond the numbers.

Next Steps

Question #37: Impact evaluation has not been given a lot of attention even at policy level in a good number of institutions. My suggestion is that we need to build this presentation with some possible next steps, [which] would help experts do more analysis based on the in-depth review done here.

The first part of this comment seems related to my response to question #22, with question #26 also relevant. What comes my mind is that a specific methodological approach to impact evaluations has been given a lot of attention. There are many who are ardently arguing for increased use of a specific type of impact evaluation, and others (such as me) have some reservations about a universal adoption of a specific methodological approach to impact evaluations. In my mind, I do not want to fight about who is "right" or what is "better." However, I do feel it is important to advocate for and work to improve they ways that impact can be examined and assessed in a great many ways, including through the use of qualitative methods.

This webinar presentation and the subsequent paper is one small study in a giant ocean of studies (and dare I say, a giant quantitative ocean). That said, the high registration numbers (585) and attendance (309) indicate interest in the range of ways impact can be examined and assessed. Hopefully, those who attended the webinar continued the conversations afterward, and hopefully with the report, the conversations will continue further.

Question #38: What suggestions do you have for us as a community to come to consensus on definitions and agree to provide a concrete description of the research methods used in sufficient detail?

One challenge with this question is use of the word "community" and sliding into thinking that there is a singular evaluation community or a singular qualitative community, and so forth—as if there is one right way and all the rest are wrong. However, that 585 registered for this webinar and 308 attended seems interesting—an interesting indication that there appears to be a big group of people who are wanting to raise the kinds of questions I am raising. To answer the question more directly, my suggestion would be the simple cliché of practice what you preach.

Question #39: From what you presented, it seems your report is more of a critique. Is there any source we can use to improve qualitative eval and mixed methods? Or are there any plans to take that next step after the analysis done in this report?

Indeed, this study includes critical examination. I ardently stand by the importance of critical examination. To my mind, to know how to do something well, one also needs to be able to recognize when something has not been done well, and really know the specifics of the shortcomings. In addition, through engaging in critical examination, the aim of this study has been to assess and provide insights concerning strengths and weaknesses in a purposefully selected set of evaluation reports.

I do not exactly understand "take that next step." This study was a single activity under the MEASURE Evaluation project, which is a project with a sizable library of resources and tools. I hope this paper will be welcomed into that library. I also hope evaluators and the broader international development community will read and use this paper. I see this paper as a resource and a tool to help refine and enhance how qualitative work is undertaken.

Sources per the first part of the question are included in the next section and the subsequent appendices.

Request for Reference

Question #40: Did you discover any excellent examples of true mixed methods analysis?

Question #40a: Ok, so give us some "very good" examples of mixed methods eval.

Question #40b: Can you identify a good resource on qualitative research methodology?

I understand the desire to ask these types of questions and appreciate the importance of examples. I am, however, going to provide an answer potentially seen as a non-answer, or even an annoying answer. I feel that the best advice I can give is that if you want good examples of mixed-methods studies, then you should look for them. Specifically look in journals, via googling, the DEC, donor webpages, etc. The exercise of looking, reading, and comparing will go a long way in helping you make your own assessments about what is good and what is not. Simply taking my word about what is good seems a bit of a short cut. A good mixed-methods study is one that is thoughtfully designed, customized, and specifically executed, not a study that simply and exactly follows the approach of another study.

Question 41: Was there a strong qualitative study design that, based on your assessment, was able to adequately assess the intervention's impact?

No. However, I say no partly because it seems to me that a certain set of people have co-opted the word "impact" and implied there is only one way to assess it. As result, I feel as though trying to assess an intervention's impact through qualitative work is a no-win situation. For too many people, the review or response to the report would be one short sentence—this is not an impact evaluation.

That said, across the 32 reports there are certainly examples of steps taken to thoughtfully design a qualitative study. As I noted in question #3, I would not be comfortable in rating the 32 evaluation reports, but I would advocate that pulling a bunch of evaluation reports from the DEC and reading them is a great way to understand what is being done well and what is being done lazily. In addition, across the 32 reports there are certainly examples where the analysis has done well in providing useful insights regarding an intervention's impact. Again, it is not hard to spot good examples (albeit, I think there are more bad examples), and some of the elements that make analysis good include nuance, specificity, thoughtfulness, thoroughness, transparency, and overall ability to back up what is being conveyed.

Question #42: I have been using the Rosetta Stone of Logical Frameworks, which I've found useful:

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC8QFjAAahUKEwidueqP7-fIAhUQ6WMKHSzbDY0&url=http%3A%2F%2Fwww.mande.co.uk%2Fdocs%2FRosettastone.doc&usg=AFQjCNFMkum0lKDpbHWD5IVJnw-Qr0KOiQ&sig2=0U7A1K8bbGl3PId8dxcQcw.

Great. Thank you. Indeed, it is quite fascinating to see the ways terminology varies across different donor agencies. I was particularly intrigued by the column labeled "ultimate impact." Seeing this term, I think, speaks well to my comments about how everyone does and should be able to use the term "impact." I guess now, though, we are in a world where there are Impact Evaluations (uppercase "I" and "E" because the control group aspect is methodologically sound), impact evaluations (lowercase "i" and "e" because the control group aspect is methodologically fuzzy), and also ultimate impact (not how "ultimate impact" is different from "non-ultimate impact").

APPENDIX 2. LIST OF 224 DOCUMENTS CONSIDERED FOR REVIEW

Legend

Y = Yes, continue to next search round

N = No, do not continue to next search round

QO = Qualitative-oriented evaluation

IE = Impact evaluation

Title		Year	Search round			
			2	3	4	5
1	Evaluation of the Training in Reproductive Health (TRH III) Project	2003	Y	Y	QO	N
2	Applying lessons learned to child survival and maternal care services	2003	Y	N	N	N
3	Community health partnerships (CHAPS) project final evaluation report	2003	Y	Y	N	N
4	Enabling the expansion & sustainability of integrated RCH [reproductive & child health]/infectious diseases outreach services in Ranchi, Jharkhand, [India]: Final evaluation	2003	N	N	N	N
5	Evaluation of PRIME [improving the performance of primary providers in reproductive health] II	2003	N	N	N	N
6	Evaluation of the Training in Reproductive Health (TRH III) Project	2003	N	N	N	N
7	Final evaluation HIV/AIDS impact mitigation through Mobilizing Affected Communities Project Nepal	2003	Y	Y	QO	N
8	Final evaluation report: Varzob, [Tajikistan] Reproductive Health Improvement Project	2003	N	N	N	N
9	Impact assessment of HIV/AIDS on the municipalities of Ongwediva, Oshakati, Swakopmund, Walvis Bay and Windhoek	2003	N	N	N	N
10	Improving RH [reproductive health] quality: PRIME [Improving the performances of primary providers in reproductive health] II -- mid-term assessment	2003	N	N	N	N
11	Malawi final evaluation the Partnership for Equity, Access and Quality (PEAQ) Project	2003	Y	Y	N	N
12	USAID/Cambodia: Strategic objective close out report – improved reproductive and child health	2003	N	N	N	N
13	Care and support portfolio final evaluation	2004	Y	Y	QO	N
14	Africare Child Survival Project final evaluation	2004	Y	N	N	N
15	Burkina Faso evaluation of the logistics system of ARVS	2004	Y	Y	QO	N
16	Child Survival XVI Project	2004	Y	N	N	N
17	Evaluation of community education interventions in sexual and reproductive health services in urban-marginal areas of La Paz, Bolivia	2004	N	N	N	N
18	Evaluation of the Indonesian Aksi Stop AIDS ASA Program	2004	Y	Y	QO	N

Title		Year	Search round			
			2	3	4	5
19	Maximizing resources to meet client needs evaluation of comprehensive HIV/AIDS care and support model	2004	Y	Y	IE	Y
20	PLACE in South Africa evaluation of a successful community-based AIDS prevention program	2004	Y	Y	QO	N
21	Report of the final evaluation of Umoyo network Malawi	2004	Y	Y	QO	N
22	USAID birthspacing programmatic review: An assessment of country-level programs, communications, and training materials	2004	N	N	N	N
23	Voluntary Counseling and Testing (VCT) Program final evaluation	2004	Y	Y	QO	N
24	Westmorland adolescents' reproductive health: An evaluation of Youth. now's peer educator/peer link intervention	2004	Y	Y	N	N
25	Assessment of USAID reproductive health and family planning activities in the Eastern European and Eurasian region, with special reference to Armenia, Kazakhstan, and Romania	2005	N	N	N	N
26	Assessment: CARE Tumaini program – provides home-based care for people living with HIV/AIDS and support for orphans and vulnerable children	2005	N	N	N	N
27	Evaluation of the CIES adolescent program	2005	N	N	N	N
28	Expanded community-based distribution project impact analysis report Zimbabwe	2005	Y	Y	N	N
29	Expanded HIV/AIDS prevention program and VCT strengthening in Zambia	2005	Y	Y	IE	Y
30	Final evaluation of the AMKENI project	2005	Y	Y	QO	N
31	Improving the health of Guatemala's most vulnerable population: Migrant and resident women and children in the Boca Coast region of Southwestern Guatemala: Final evaluation report	2005	N	N	N	N
32	Strategic appraisal: USAID/Albania health strategy and portfolio – achievements, opportunities and a way forward	2005	N	N	N	N
33	The USAID/India urban health program: An evaluation of activities to date and recommendations for the future	2005	N	N	N	N
34	USAID/India: Strategic objective close-out report – reduced fertility and improved reproductive health in North India: FY 1994-2002	2005	N	N	N	N
35	Assessment of the USAID/Bangladesh component of DELIVER Project: A success to build on	2006	N	N	N	N
36	Child survival-community health initiative for the Districts of Kouroussa and Mandianga Guinea	2006	Y	Y	N	N
37	District health partnership child survival program	2006	Y	Y	N	N
38	Evaluation of a pilot program to provide psychosocial support to children affected by HIV/AIDS	2006	Y	Y	QO	N
39	Evaluation of AWARENESS project	2006	Y	Y	N	N
40	Evaluation of Jhpiego's VCT training activities in the Caribbean Region	2006	Y	Y	QO	N
41	Evaluation of the Avert Society project final report	2006	Y	Y	QO	N
42	FY 2002-2006 Title II DAP final evaluation	2006	Y	N	N	N

Title		Year	Search round			
			2	3	4	5
43	IMPACT Ethiopia final evaluation report	2006	Y	Y	QO	N
44	Impact evaluation of Students Alliance for Female Education (SAFE) clubs and peer education programs	2006	Y	Y	IE	Y
45	Improved child survival in Malawi through community-based interventions and strengthening of the health delivery infrastructure	2006	Y	Y	N	N
46	Integrating child survival Haiti final evaluation	2006	Y	Y	N	N
47	Lessons learned from the Benin HIV-AIDS prevention program	2006	Y	Y	QO	N
48	Malawi DELIVER: End of project evaluation	2006	N	N	N	N
49	Ndwedwe child survival project	2006	Y	N	N	N
50	PATH support to Ukraine to implement the national TB program	2006	Y	Y	N	N
51	Reproductive and Child Health, Nutrition and HIV/AIDS Program (RACHANA) – final evaluation	2006	Y	Y	QO	Y
52	Rwanda Umucyo illumination child survival project final evaluation report	2006	Y	N	N	N
53	Strategic objective no. 519-003: "Health of Salvadorans, primarily women, youth and children, improved" 1997-2005 – close-out report	2006	N	N	N	N
54	USAID/Zambia SO3 close-out report, 1998-2004	2006	N	N	N	N
55	A strategic assessment of three integrated health projects in Cambodia	2007	N	N	N	N
56	AIM project evaluation final report	2007	Y	Y	IE	Y
57	End of program evaluation report – Youth Education for Life Skills (YES)	2007	Y	N	N	N
58	Evaluation of the FANTA Project	2007	Y	Y	N	N
59	Evaluation of the Jamaica adolescent reproductive health activity 2000–2007	2007	N	N	N	N
60	Final evaluation – CRS institutional capacity building	2007	Y	N	N	N
61	Final evaluation report – Reducing child morbidity and strengthening health care systems	2007	Y	Y	QO	N
62	Final evaluation report – Zarafshan partnership for scaling-up innovation approaches for rural Tajikistan	2007	Y	Y	N	N
63	From behavior change communication to strategic behavioral communication on HIV in Kenya 1999-2006	2007	Y	Y	N	N
64	Improving the health of mothers and children of rural Jinotega Nicaragua	2007	Y	Y	N	N
65	Uganda community-based growth promotion: Program review	2007	N	N	N	N
66	An evaluation of WWFs USAID and Johnson & Johnson supported projects	2008	Y	N	N	N
67	Assessment of the BRIDGE project	2008	Y	Y	QO	Y
68	Ethiopia positive change – children, communities and care end-of-project evaluation	2008	Y	Y	QO	N
69	Evaluation of capacity building in two faith-based HIV/AIDS projects in Ethiopia	2008	Y	Y	QO	N

Title		Year	Search round			
			2	3	4	5
70	External end of project evaluation health center renovation coordination project	2008	Y	Y	QO	N
71	External end of project evaluation urban agriculture program for HIV/AIDS affected women and children	2008	Y	Y	QO	Y
72	Final Evaluation I-LIFE Malawi development assistance program	2008	Y	Y	N	N
73	Presidents emergency plan for AIDS relief program evaluation	2008	Y	Y	N	N
74	School as a Workplace in Kenya Evaluation of the Teachers Matter HIV/AIDS Project	2008	Y	Y	IE	Y
75	Sexual behavior and STI HIV status an evaluation of the effect of interview mode on reporting	2008	Y	Y	QO	N
76	Training journalists to report on HIV/AIDS final evaluation of a global program	2008	Y	Y	N	N
77	End-of-project evaluation of palliative care services	2009	N	N	N	N
78	A concurrent evaluation of Phase II of the NRHM BCC campaign	2009	N	N	N	N
79	Assessment of the USAID health policy initiative IQC and Task Order 1	2009	N	N	N	N
80	Community education and sensitization as an OVC care and support strategy: Evaluation of the Integrated AIDS Program – Thika in Kenya	2009	Y	Y	IE	N
81	Community-based psychosocial intervention for HIV-affected children and their caregivers	2009	Y	Y	IE	N
82	Early Marriage Evaluation Study (EMES)	2009	N	N	N	N
83	Effects of a community-focused approach supporting the most vulnerable children: Evaluation of SAWAKA Jali Watoto Program in Kagera, Tanzania	2009	Y	Y	IE	N
84	Effects of programmes supporting orphans and vulnerable children: Key findings, emerging issues, and future directions from evaluations of four projects in Kenya and Tanzania	2009	Y	Y	IE	Y
85	End-of-project evaluation of the ACE program: Final evaluation report	2009	Y	Y	QO	Y
86	Ethiopia: Evaluation of the mothers' support group strategy	2009	N	N	N	N
87	Evaluation of the COMPRI-A social marketing program by USAID in the Islamic Republic of Afghanistan	2009	N	N	N	N
88	Extending service delivery (ESD) participatory assessment	2009	N	N	N	N
89	Final assessment and evaluation report: Tuberculosis program implementation in the Central Asian republics: Kazakhstan – Kyrgyzstan – Tajikistan – Turkmenistan – Uzbekistan	2009	N	N	N	N
90	Final evaluation: The private sector program in Ethiopia – assessing the foundation for a continuum focused on public-private health sector collaboration in addressing HIV/AIDS and tuberculosis in Ethiopia	2009	Y	Y	N	N
91	Final evaluation of the project for expanding the role of networks of people living with HIV/AIDS	2009	Y	Y	QO	N
92	Impact evaluation of the Kingston priorities for local AIDS control efforts intervention	2009	N	N	N	N

Title		Year	Search round			
			2	3	4	5
93	PEPFAR public health evaluation – care and support – Phase I Kenya	2009	Y	Y	QO	N
94	PEPFAR public health evaluation – care and support – Phase I Uganda	2009	Y	Y	QO	N
95	Preventing the medical transmission of HIV in Zambia: Final evaluation report	2009	Y	Y	QO	N
96	Promoting a home-based program model for supporting children affected by HIV/AIDS: Evaluation of Tumaini Project in Iringa Region, Tanzania	2009	Y	Y	IE	N
97	The difference interventions for guardians can make: Evaluation of the Kilifi Orphans and Vulnerable Children Project in Kenya	2009	Y	Y	IE	N
98	Abstinence and risk avoidance for youth project report (ARK): end of project evaluation for Haiti, Kenya and Tanzania	2010	Y	Y	QO	N
99	ARC TWC evaluation: Guyana and Tanzania comparative report	2010	Y	Y	QO	N
100	Assessment of three USAID/Peru health projects implemented by the Ministry of Health: Vigia; Coverage with quality; Improved health for populations at high risk	2010	N	N	N	N
101	Breaking barriers project: Kenya, Uganda and Zambia – end-term evaluation report	2010	Y	Y	QO	N
102	Drivers of behavior change – qualitative research findings in Jordan	2010	Y	N	N	N
103	End of project evaluation report: Tackling girls' and young women's vulnerability to HIV and AIDS in Tanzania	2010	Y	Y	QO	N
104	End-of-project evaluation report: Faith-based regional initiative for orphans and vulnerable children (FABRIC)	2010	Y	Y	QO	N
105	Evaluation of Advancing Surveillance, Policies, Prevention, Care and Support to Fight HIV/AIDS in Nepal (ASHA) Project	2010	Y	Y	QO	Y
106	Evaluation of USAID/Nepal's key social marketing and franchising project: AIDS, reproductive health, and child survival (N-MARC): Final report	2010	Y	Y	N	N
107	Final evaluation report of the BELONG (Better Education and Life Opportunities for Vulnerable Children through Networking and Organizational Growth)	2010	Y	Y	QO	N
108	Final evaluation report: Toliara Region Expanded Impact Project (TREIP)	2010	N	N	N	N
109	Formative evaluation: Presidential initiative on AIDS strategy for communication to youth	2010	Y	Y	N	N
110	Formative evaluation of quality of care initiatives by Ministry of Health – Uganda report: Final report	2010	N	N	N	N
111	Global health: Trends in U.S. spending for global HIV/AIDS and other health assistance in fiscal years 2001–2008	2010	N	N	N	N
112	Monitoring and evaluation of the emergency plan progress (MEEPP): end-of-project evaluation	2010	Y	N	N	N
113	PEPFAR public health evaluation – Care and support: Phase 2 Kenya	2010	Y	Y	IE	N
114	RAPIDS impact evaluation final report 2005-2009 key findings	2010	Y	Y	QO	N

Title		Year	Search round			
			2	3	4	5
115	USAID/Nigeria: Centre for Development and Population Activities (CEDPA) positive living project: Final evaluation	2010	Y	Y	QO	N
116	USAID/SENEGAL HIV/AIDS and tuberculosis programs: Interim assessment	2010	N	N	N	N
117	USAID/Zambia gender-based violence programming evaluation	2010	Y	N	N	N
118	Avert end of project evaluation	2011	Y	Y	QO	N
119	Decent care values in palliative care services: A formative evaluation	2011	Y	Y	N	N
120	End of project evaluation: Sustainable reintegration of orphans and vulnerable children into family and community life	2011	Y	Y	QO	N
121	GOAL Uganda: Scaling up innovative HIV/AIDS interventions through strong partnerships amongst conflict-affected populations of Pader (Agago) District, Northern Uganda, 2008-2011: End project evaluation report	2011	Y	Y	IE	Y
122	HIV/AIDS portfolio evaluation USAID/Cambodia: Final report	2011	Y	Y	QO	N
123	Lessons learned from the USAID/Senegal community health program (CHP): Final evaluation	2011	N	N	N	N
124	Research and evaluation report: Spread of PMTCT and ART better care practices through collaborative learning in Tanzania	2011	Y	Y	N	N
125	Research and evaluation report: The partnership for quality improvement to improve PMTCT and ART services in Tanzania – assessment of results, capacity, and potential for institutionalization	2011	Y	Y	N	N
126	Samastha project, USAID/India: Final evaluation	2011	Y	Y	QO	N
127	Taking knowledge for health the extra mile: An evaluation of the Malawi knowledge for health project (K4Health)	2011	Y	Y	IE	Y
128	USAID South Africa PEPFAR treatment program final evaluation	2011	Y	Y	QO	N
129	USAID/Benin: Evaluation of family health activities	2011	Y	Y	N	N
130	USAID/Honduras: HIV/AIDS prevention programs evaluation	2011	Y	Y	QO	Y
131	A global development alliance to combat HIV/AIDS in the agribusiness and mining sectors in Zambia: End of project evaluation	2012	Y	Y	N	N
132	AIDS prevention and control (APAC) evaluation, USAID/India	2012	Y	Y	IE	Y
133	An assessment of the Sudan HIV and AIDS program (SHAP)	2012	N	N	N	N
134	Childline Mpumalanga OVC community development programme: Evaluation report	2012	Y	Y	IE	Y
135	Circles of support (COS) OVC project: Evaluation report	2012	Y	Y	N	N
136	End of project evaluation of the USAID/Zimbabwe Family AIDS Initiative (FAI) project	2012	Y	Y	QO	Y
137	Enhancing strategic information project: End of project evaluation	2012	Y	Y	QO	N
138	Evaluation of the 100% condom use program in An Giang: Results and lessons learned for sustainability and scale-up	2012	Y	Y	N	N

Title		Year	Search round			
			2	3	4	5
139	Exploring the effects of holistic capacity building: An evaluation of NuPITA's technical assistance on NPI partners' service delivery and sustainability	2012	Y	Y	IE	Y
140	Exploring the impact of the Community Based Care for Orphans and Vulnerable Children (CBCO) Program	2012	N	N	N	N
141	Final evaluation: THANZI ("good health") (TB health activities in Zambia)	2012	Y	N	N	N
142	Final evaluation report on the roads to a healthy future (ROADS II) project in East, Central and Southern Africa	2012	Y	Y	QO	N
143	Impact evaluation of the Bangladesh Smiling Sun Franchise Program (BSSFP)	2012	N	N	N	N
144	Leadership initiative for public health in East Africa: Outcomes assessment	2012	N	N	N	N
145	Malawi Community REACH Program final evaluation report	2012	Y	Y	QO	N
146	Market-based partnerships in health (MBPH): Final evaluation, USAID/India	2012	N	N	N	N
147	May'khethele OVC programme: External outcomes evaluation report	2012	Y	Y	IE	Y
148	Monitoring and evaluation services for the Civil Society Fund study on the effectiveness of OVC interventions towards improvement in food and nutrition security and economic strengthening among OVC	2012	Y	Y	N	N
149	Multi-year Assistance Program (MYAP), cooperative agreement #FFP-A-00-08-00080-00: Final evaluation report	2012	Y	N	N	N
150	Networks of Hope program: Evaluation report	2012	Y	Y	QO	N
151	Performance evaluation and assessment of USAID/Kenya Nutrition and HIV Program (NHP)	2012	Y	Y	QO	N
152	Performance evaluation of the civilian assistance program (CTAP) - Afghanistan	2012	Y	N	N	N
153	Performance evaluation of the FamiSalud project	2012	N	N	N	N
154	Performance evaluation of USAID support to palliative care in South Africa	2012	Y	Y	QO	N
155	Planning for life Phase 2: Evaluation report	2012	N	N	N	N
156	SMBCI project final performance evaluation: Final report	2012	Y	Y	N	N
157	South Africa Umbrella Grants Management (UGM) performance evaluation	2012	Y	Y	QO	N
158	Strengthening the capacity for country ownership: End-of-project evaluation	2012	Y	Y	QO	Y
159	Summary report of a qualitative research study on intermittent preventive treatment during pregnancy for malaria in rural Zambia	2012	Y	N	N	N
160	The impact of the ISIBINDI program on vulnerable youth: Evaluation report	2012	Y	Y	IE	Y
161	USAID/Ethiopia: End of project evaluation for the urban health extension program	2012	Y	Y	QO	Y
162	USAID/Office of HIV and AIDS: Project SEARCH end of project evaluation supporting evaluation and research to combat HIV	2012	Y	Y	N	N

Title		Year	Search round			
			2	3	4	5
163	USAID/Philippines: External evaluation of the tuberculosis portfolio (2006–2011)	2012	Y	Y	QO	N
164	USAID/Rwanda HIV/AIDS Clinical Services Project (HCSP) end of project evaluation	2012	Y	Y	QO	Y
165	USAID/Senegal: Social marketing program performance evaluation	2012	N	N	N	N
166	USAID/South Africa umbrella grants management project end of project partner evaluation: Anglican Aids and Healthcare Trust (AAHT) Vana Vetu Program	2012	Y	Y	QO	N
167	USAID/South Africa umbrella grants management project end-of-project partner evaluation: Greater Nelspruit Rape Intervention Programme (GRIP)	2012	Y	N	N	N
168	USAID/South Africa umbrella grants management project end of project partner evaluation: Hands at Work	2012	Y	Y	N	N
169	USAID/South Africa umbrella grants management project end of project partner evaluation: Heartbeat Centre for Community Development	2012	Y	Y	QO	N
170	USAID/South Africa umbrella grants management project end of project partner evaluation: Noah (Nurturing Orphans of AIDS for Humanity)	2012	Y	Y	QO	N
171	USAID/South Africa umbrella grants management project end of project partner evaluation: Senzakwenzeke community development	2012	Y	Y	QO	N
172	USAID/South Africa umbrella grants management project end of project partner evaluation: Woz'obona/Sekhukhune Educare Project	2012	Y	Y	QO	N
173	USAID-supported TB control activities in Mexico: Final evaluation	2012	Y	N	N	N
174	Assessment of Over-the-Counter HIV Rapid Test Kits in Namibia	2013	N	N	N	N
175	Community health volunteer program functionality and performance in Madagascar: A synthesis of qualitative and quantitative assessments	2013	N	N	N	N
176	Constituency dialogues and citizen engagement in Cambodia: Findings from a mixed methods impact evaluation	2013	Y	N	N	N
177	Encouraging men's participation in HIV and AIDS prevention and HIV testing services: Evaluation of the Men As Partners (MAP) approach in Côte d'Ivoire	2013	Y	Y	QO	Y
178	End of project evaluation: Expanding the access to and scope of palliative care for people living with HIV/AIDS (PLHIV) and their families – final report	2013	Y	Y	QO	N
179	End of project evaluation: USAID's "Children First" orphans and vulnerable children (OVC) project in Zimbabwe	2013	Y	Y	QO	N
180	End of project evaluation of the PSI social marketing project in Madagascar	2013	Y	Y	N	N
181	End-of-project performance evaluation of the Ukraine HIV/AIDS service capacity project: Final report	2013	Y	Y	QO	N
182	Evaluating the impact of the Wula Nafaa Natural Resources Management Program in Senegal on the socio-economic status of the population: A quasi-experimental design analysis	2013	Y	N	N	N

Title		Year	Search round			
			2	3	4	5
183	Evaluation of the TCE program in Mpumalanga and Limpopo: Evaluation report	2013	Y	Y	IE	Y
184	Evaluation report: USAID/Uganda faith-based HIV and AIDS program	2013	Y	Y	QO	N
185	Final evaluation for the Nehnwaa child survival project: Census-based impact-oriented methodology for community-based primary health care in Nimba County, Liberia	2013	Y	Y	N	N
186	Final evaluation of the health communications partnership project (HCP II): Final report	2013	Y	Y	N	N
187	Final evaluation report: Projet AIDE performance evaluation 2009-2013	2013	Y	Y	N	N
188	Final performance evaluation: AIDS Population and Health Integrated Assistance, People-Centered, Local Leadership, Universal Access, and Sustainability (APHIAplus) Nairobi/coast project	2013	Y	Y	QO	Y
189	Final performance evaluation of the School-Community Partnership Serving Orphan and Vulnerable Children Affected by HIV/AIDS (SCOPSO) project	2013	Y	Y	N	N
190	Final performance evaluation of the teacher education and professional development project in Kenya	2013	Y	N	N	N
191	Health and life skills project (HELP) among house girls in Nairobi: An evaluation of the effect of HELP on house girls' vulnerability towards STI/HIV and unintended pregnancy	2013	Y	Y	N	N
192	HIV among female sex workers and men who have sex with men in Swaziland: A combined report of quantitative and qualitative studies	2013	N	N	N	N
193	Improving quality of HIV services and health worker performance in Tandahimba District, Tanzania: An evaluation	2013	Y	Y	QO	N
194	Integrating family planning into fistula services: An evaluation and case study	2013	Y	N	N	N
195	Johns Hopkins University (JHU) HIV communication program project performance evaluation	2013	Y	Y	QO	N
196	Measuring HIV stigma and discrimination among health facility staff: Brief questionnaire	2013	N	N	N	N
197	Operationalising structural interventions for HIV prevention: Lessons from Zambia	2013	N	N	N	N
198	Performance Evaluation – USAID/Zimbabwe Promoting Recovery in Zimbabwe (PRIZE) Project	2013	Y	Y	N	N
199	PharmAccess International Tanzania Police, Prisons, and Immigration (TPPI) Project: Final evaluation report	2013	Y	Y	QO	N
200	Rapid assessment of pediatric HIV treatment in Nigeria	2013	N	N	N	N
201	Strategic assessment to define a comprehensive response to HIV in Iringa, Tanzania: Summary of findings	2013	N	N	N	N
202	Strengthening local capacity to combat HIV/AIDS in Zambia: End of project performance evaluation	2013	Y	Y	N	N
203	USAID | DELIVER project performance evaluation Malawi 2007-2013	2013	N	N	N	N
204	USAID tuberculosis task order 2015 performance evaluation	2013	Y	N	N	N

Title		Year	Search round			
			2	3	4	5
205	USAID/Brazil: Performance evaluation of the tuberculosis portfolio, 2001-2012	2013	Y	N	N	N
206	USAID/Democratic Republic of the Congo: integrated HIV/AIDS program (ProVIC)	2013	N	N	N	N
207	USAID/South Africa umbrella grants management project end of project partner evaluation: Project Concern International	2013	Y	Y	QO	N
208	USAID/Tanzania: Channeling men's positive involvement in a national HIV response project (CHAMPION): End of project evaluation	2013	Y	Y	QO	Y
209	USAID/Uganda AFFORD: Health marketing initiative project evaluation: Improving the lives of Ugandans	2013	N	N	N	N
210	A performance evaluation of the national HIV prevention program for FSW and MSM in Ghana	2014	Y	Y	QO	Y
211	Central Asia HIV/AIDS program evaluation	2014	Y	Y	QO	Y
212	Communications Support for Health (CSH) Program. Safe love outcome evaluation report	2014	Y	Y	N	N
213	Eliminating pediatric AIDS in Swaziland project evaluation report	2014	Y	Y	QO	Y
214	End-of-project evaluation of the Sustainable Action Against HIV and AIDS in the Community Project (SAHACOM)	2014	Y	Y	QO	N
215	Georgia HIV prevention project: End of project report	2014	Y	Y	QO	N
216	Impact evaluation of the Mayer Hashi program of long-acting and permanent methods of contraception in Bangladesh	2014	Y	Y	IE	N
217	Impact evaluation of the project "Strengthening Sustainable Orphans and Vulnerable Children (OVC) Care and Support in Côte d'Ivoire" in the urban context of Abidjan	2014	Y	Y	QO	N
218	Mozambique program assessment: Community care for vulnerable children in an integrated vulnerable children and home-based care program	2014	Y	Y	QO	N
219	Performance evaluation of USAID testing and counseling projects in Tanzania: Final performance evaluation of Angaza Zaidi and universal HIV and AIDS intervention for counseling and testing	2014	Y	Y	QO	N
220	Performance evaluation of USAID/Uganda capacity program	2014	Y	Y	QO	N
221	SYMMACS: The Systematic Monitoring of the Male Circumcision Scale-up in Eastern and Southern Africa. Final report of results from Kenya, South Africa, Tanzania and Zimbabwe	2014	Y	Y	QO	N
222	The Impact of community caregivers in Côte d'Ivoire	2014	Y	Y	IE	Y
223	USAID/Zambia: Zambia-Led Prevention Initiative (ZPI) end-of-project performance evaluation	2014	Y	Y	QO	N
224	Endline evaluation of the USAID TransACTION prevention and care services for at-risk mobile populations programs in Ethiopia	2015	Y	Y	IE	Y

APPENDIX 3. LIST OF THE 32 USAID-FUNDED HIV/AIDS-RELATED EVALUATION REPORTS REVIEWED

Date	Evaluation Title
2004	Maximizing resources to meet client needs evaluation of comprehensive HIV/AIDS care and support model
2005	Expanded HIV/AIDS prevention program and VCT strengthening in Zambia
2006	Impact evaluation of Students Alliance for Female Education (SAFE) clubs and peer education programs
2006	Reproductive and Child Health, Nutrition and HIV/AIDS Program (RACHANA) – final evaluation
2007	AIM project evaluation final report
2008	School as a workplace in Kenya evaluation of the Teachers Matter HIV/AIDS Project
2008	BRIDGE Project final evaluation
2008	External end of project evaluation of the Urban Agriculture Program for HIV/AIDS affected women and children
2009	Effects of programs supporting orphans and vulnerable children: Key findings, emerging issues, and future directions from evaluations of four projects in Kenya and Tanzania
2009	End-of-project evaluation of the ACE program: Final evaluation report
2010	Evaluation of Advancing Surveillance, Policies, Prevention, Care and Support to Fight HIV/AIDS in Nepal (ASHA) Project
2011	GOAL Uganda: Scaling up innovative HIV/AIDS interventions through strong partnerships amongst conflict-affected populations of Pader (Agago) District, Northern Uganda, 2008-2011: End project evaluation report
2011	Taking knowledge for health the extra mile: An evaluation of the Malawi Knowledge for Health Project (K4Health)
2011	USAID/Honduras: HIV/AIDS prevention programs evaluation
2012	Exploring the effects of holistic capacity building: An evaluation of NuPITA's technical assistance on NPI partners' service delivery and sustainability
2012	Childline Mpumalanga: OVC community development programme: Evaluation report
2012	USAID/Ethiopia: End of project evaluation for the urban health extension program
2012	Strengthening the capacity for country ownership: End-of-project evaluation
2012	The impact of the ISIBINDI programme on vulnerable youth: Evaluation report
2012	USAID/Rwanda HIV/AIDS Clinical Services Project (HCSP) end-of-project evaluation
2012	AIDS prevention and control (APAC) evaluation, USAID/India
2012	May'khethele OVC programme: External outcomes evaluation report
2012	End-of-project evaluation of the USAID/Zimbabwe Family AIDS Initiative (FAI) Project
2013	Final performance evaluation: AIDS, Population and Health Integrated Assistance, People-Centered, Local Leadership, Universal Access, and Sustainability (APHIAplus)
2013	USAID/Tanzania: Channeling Men's Positive Involvement in a national HIV response project (CHAMPION): End of project evaluation
2013	Evaluation of the TCE programme in Mpumalanga and Limpopo: Evaluation report
2013	Encouraging men's participation in HIV and AIDS prevention and HIV testing services: Evaluation of the Men As Partners (MAP) approach in Côte d'Ivoire
2014	A performance evaluation of the national HIV prevention program for FSW and MSM in Ghana
2014	Central Asia HIV/AIDS program evaluation
2014	The impact of community caregivers in Côte d'Ivoire
2014	Eliminating pediatric AIDS in Swaziland project evaluation report
2015	Endline evaluation of the USAID TransACTION prevention and care services for at-risk mobile populations programs in Ethiopia

APPENDIX 4. MIXED-METHODS LITERATURE

Manuals, Guidelines, Toolkits, Briefs, etc.

Bamberger, M. (2012). *Introduction to mixed methods in impact evaluation.* InterAction Impact Evaluation Notes No. 3. Washington, DC, USA: InterAction. Retrieved from http://www.betterevaluation.org/en/resources/guides/intro_mixed-methods_impact-evaluation.

Bricki, N., & Green, J. (2007). *A guide to using qualitative research methodology.* Geneva, Switzerland: Medecins Sans Frontieres. Retrieved from https://evaluation.msf.org/sites/evaluation/files/a_guide_to_using_qualitative_research_methodology.pdf.

Davies, R., & Dart, J. (2015). *The "most significant change" (MSC) technique: A guide to use.* Version 1.0. United Kingdom: Care International; Australia: Oxfam Community Aid Abroad; Australia: Learning to Learn, Government of South Australia; United Kingdom: Oxfam New Zealand, Christian Aid, Exchange; Ibis, Denmark: Mellemfolkeligt Samvirke; United States: Lutheran World Relief. Retrieved from https://www.mande.co.uk/docs/MSCGuide.pdf.

Jones, N., Jones, H., Steer, L., & Datta, A. (2009). *Improving impact evaluation production and use.* London, United Kingdom: Overseas Development Institute. Retrieved from https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/4158.pdf.

Kielmann, K., Cataldo, F., & and Seeley, J. (2012). *Introduction to qualitative research methodology: A training manual.* London, United Kingdom: Department for International Development. Retrieved from https://www.rbfhealth.org/sites/rbf/files/Introduction%20to%20Qualitative%20Research%20Methodology%20-%20A%20Training%20Manual.pdf.

Levine, R., & Savedoff, W. (2015). *The future of aid: Building knowledge collectively.* Policy Paper 050. London, United Kingdom: Center for Global Development. Retrieved from https://www.cgdev.org/sites/default/files/CGD-Policy-Paper-Levine-Savedoff-Future-Aid.pdf.

Mack, N., Woodsong, C., MacQueen, K. M., Guest, G., & Namey, E. (2011). *Qualitative research methods: A data collector's field guide.* Research Triangle Park, NC: FHI 360. Retrieved from https://www.fhi360.org/sites/default/files/media/documents/Qualitative%20Research%20Methods%20-%20A%20Data%20Collector's%20Field%20Guide.pdf.

Savedoff, W., Levine, R., & Birdsall, N. (2006). *When will we ever learn? Improving lives through impact evaluation.* Report of the Evaluation Gap Working Group. London, United Kingdom: Center for Global Development. Retrieved from https://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. *Broadening the range of designs and methods for impact evaluations.* Department for International Development (DFID) Working Paper 38. London, United Kingdom: DFID. Retrieved from https://www.oecd.org/derec/50399683.pdf.

White, H. (2009). *Theory-based impact evaluation: Principles and practice.* Working Paper 3. New Delhi, India: International Initiative for Impact Evaluation. Retrieved from http://www.3ieimpact.org/media/filer_public/2012/05/07/Working_Paper_3.pdf.

Peer-Reviewed Journal Articles

Bamberger, M., Rugh, J., Church, M., & Fort, L. (2004). Shoestring evaluation: Designing impact evaluations under budget, time and data constraints. *The American Journal of Evaluation,* 25 (1), 5-37. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/109821400402500102.

Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. *Qualitative Research,* 15 (2), 219-234. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1468794112468475?journalCode=qrja.

Chenail, R. J. (2011). Ten steps for conceptualizing and conducting qualitative research studies in a pragmatically curious manner. *The Qualitative Report*, 16 (6), 1715-1732. Retrieved from http://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1324&context=tqr.

Denizen, N. K. (2010). Moments, mixed methods and paradigm dialogs. *Qualitative Inquiry,* 16 (6), 419-427. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1077800410364608.

Harrison, E. (2015). Anthropology and impact evaluation: A critical commentary. *Journal of Development Effectiveness,* 7 (2), 146-159. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/19439342.2015.1015436?journalCode=rjde20.

Lub, V. (2015). Validity in qualitative evaluation: Linking purposes, paradigms, and perspectives. *International Journal of Qualitative Methods,* 14 (5), 1-8. Retrieved from http://journals.sagepub.com/doi/pdf/10.1177/1609406915621406.

Maxwell, J. A. (2015). Expanding the history and range of mixed methods research. *Journal of Mixed Methods Research,* 10 (1), 12-27. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1558689815571132.

Torrance, H. (2016). Experimenting with qualitative inquiry. *Qualitative Inquiry,* 23 (1), 69-76. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1077800416649201.

Books

Creswell, J. A. (2015). *Concise introduction to mixed methods research.* London, United Kingdom: Sage Publishing. Retrieved from https://us.sagepub.com/en-us/nam/a-concise-introduction-to-mixed-methods-research/book243856.

Gertler, P. J., Martinez, M., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). *Impact evaluation in practice.* Second Edition. Washington, DC: The World Bank.

Green, J., & Thorogood, N. (2014). *Qualitative methods for health research.* Third Edition. London, United Kingdom: Sage Publishing. Retrieved from https://us.sagepub.com/en-us/nam/qualitative-methods-for-health-research/book239018.

Greene, J. C. (2007). *Mixed methods in social inquiry.* San Francisco, CA: Wiley. Retrieved from http://www.wiley.com/WileyCDA/WileyTitle/productCd-0787983829.html.

Hesse-Biber, S., & Burke Johnson, R. (Eds.). (2015). *The Oxford handbook of multimethod and mixed methods research inquiry.* United Kingdom: Oxford University Press. Retrieved from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199933624.001.0001/oxfordhb-9780199933624.

Posavac, E. (2015). *Program evaluation: Methods and case studies*. Eighth Edition. London, United Kingdom, and New York, NY: Routledge, Taylor & Francis.

Quinn Patton, M. (2015). *Qualitative research & evaluation methods integrating theory and practice*. Fourth Edition. London, United Kingdom: Sage Publishing. Retrieved from https://us.sagepub.com/en-us/nam/qualitative-research-evaluation-methods/book232962.

Richards, L. (2015). *Handling qualitative data: A practical guide*. London, United Kingdom: Sage Publishing. Retrieved from https://us.sagepub.com/en-us/nam/handling-qualitative-data/book241828.

Tashakkori, A., & Teddlie, C. (Eds). (2010). *Handbook of mixed methods in social and behavioral research*. Second Edition. Thousand Oaks, CA: Sage Publishing. Retrieved from https://us.sagepub.com/en-us/nam/sage-handbook-of-mixed-methods-in-social-behavioral-research/book233405.

Taylor, S. J., Bogdan, R., & DeVault, M. (2015). *Introduction to qualitative research methods: A guidebook and resource*. Hoboken, NJ: Wiley. Retrieved from http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118767217,subjectCd-EV00.html.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage Publishing. Retrieved from https://us.sagepub.com/en-us/nam/foundations-of-mixed-methods-research/book226302.

APPENDIX 5. LITERATURE ON GENDER-SENSITIVE EVALUATION

Bamberger, M., & Segone, M. (2011). *How to design and manage equity focused evaluations*. New York, NY: UNICEF. Retrieved from https://www.unicef.org/evaluation/index_60809.html.

Boender, C., Santana, D., Santillán, D., Hardee, K., Greene, M., and Schuler, S. (2004). *The "So What?" report: A look at whether integrating a gender focus into programs makes a difference to outcomes*. Washington, DC: Population Reference Bureau (PRB) for the Interagency Gender Working Group (IGWG).

Espinosa, J. (2013). Moving towards gender-sensitive evaluation? Practices and challenges in international development evaluation. *Evaluation* 19 (2), 171-182. Retrieved from http://journals.sagepub.com/doi/10.1177/1356389013485195.

Gender, Agriculture, & Assets Project (GAAP) (2014). GAAP: *A toolkit on collecting gender and assets data in qualitative and quantitative program evaluation*. Washington, DC: IFPRI and ILRI. Retrieved from http://gaap.ifpri.info/files/2010/12/GAAP_Toolkit_Update_FINAL.pdf.

Glick J. L., Theall K., Andrinopoulos K., and Kendall, C. (2018). For data's sake: Dilemmas in the measurement of gender minorities. *Culture, health & sexuality*. 2018 (X):1-16. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/13691058.2018.1437220?journalCode=tchs20.

Podems, Donna R. (2010). Feminist evaluation and gender approaches: There's a difference? *Journal of MultiDisciplinary Evaluation* 6 (14), 1-17. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/199.

MEASURE Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, North Carolina 27516
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of the MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-17-230
ISBN: 978-1-64232-094-7

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE
Evaluation