
USAID
FROM THE AMERICAN PEOPLE

# AN EVALUATION FRAMEWORK FOR USAID-FUNDED TIP PREVENTION AND VICTIM PROTECTION PROGRAMS

## Final Report

December 2009

This report was produced for the Social Transition Team, Office of Democracy, Governance and Social Transition of the United States Agency for International Development (USAID/E&E/DGST) by Creative Associates International, Inc. and the Aguirre Division of JBS International, Inc. It was prepared under ABE Contract EDH-I-00-05-00029-00. Jo Ann Intili, Ruth Rosenberg, Mirka Tvaruzkova, and Abigail Ladd contributed to this report.

This document is one of a series of publications sponsored by the Social Transition Team in the Bureau for Europe and Eurasia on social transition issues in the Region. You may download the documents in this series from the USAID Development Experience Clearinghouse website (http://dec.usaid.gov), using the Doc IDs for reference. If you would like more information about these and upcoming products, please contact the Social Transition Team at USAID at eesocialtransition@usaid.gov.

The Social Transition Series includes:

**An Evaluation Framework for USAID-funded TIP Prevention and Victim Protection Programs**
December 2009
Doc ID: PN-ADR-431

**Toolkit for Integrating Domestic Violence Activities into Programming in Europe and Eurasia**
July 2009
Doc ID: PN-ADO-921

**The Job that Remains: An Overview of USAID Child Welfare Reform Efforts in Europe and Eurasia**
June 2009
Doc ID: PN-ADO-922

**Best Practices in Trafficking Prevention in Europe and Eurasia**
January 2009
Doc ID: PN-ADN-296

**Methods and Models for Mixing Services for Victims of Domestic Violence and Trafficking in Persons in Europe and Eurasia**
December 2008
Doc ID: PN-ADN-297

**Social Work Education and the Practice Environment in Europe and Eurasia**
December 2008
Doc ID: PN-ADN-298

**Best Practices for Programming to Protect and Assist Victims of Trafficking in Europe and Eurasia**
November 2008
Doc ID: PN-ADN-295

**A Strategic Approach to Characterizing the Status and Progress of Child Welfare Reform in CEE/CIS Countries**
July 2008
Doc ID: PN-ADN-294

**Education Vulnerability Analysis for the E&E Region**
July 2008
Doc ID: PN-ADL-768

**An Analysis and Strategic Approach to Monitoring the Status and Progress of Child Welfare Reform in 21 CEE/CIS Countries**
November 2007
Doc ID: PN-ADL-007

**Examining the Intersection Between Trafficking in Persons and Domestic Violence**
June 2007
Doc ID: PN-ADL-005

**Quantifying the Relationship Between Corruption in Education and Economic Development in the Eastern Europe and Eurasia Region**
May 2007
Doc ID: PN-ADL-006

**Europe & Eurasia Bureau Social Transition Strategy**
June 2006
Doc ID: PD-ACH-406

**Child Welfare and Disability Project Database**
June 2006
Doc ID: PN-ADG-329

**USAID and Child Welfare Reform in Romania Challenges, Successes, and Legacy**
June 2006
Doc ID: PD-ACH-405

**Promising Practices in Community-Based Social Services in CEE/CIS/Baltics**
October 2006
Doc ID: PN-ADF-930

**Emerging Practices in Community-Based Services for Vulnerable Groups: A Study of Social Services Delivery Systems in Europe and Eurasia**
June 2006
Doc ID: PN-ADG-301

**Domestic Violence in Europe and Eurasia**
June 2006
Doc ID: PN-ADG-302

**Gender Issues in Europe and Eurasia**
June 2006
Doc ID: PN-ADG-303

**The Hope of Tomorrow – Integrating Youth into the Transition of Europe and Eurasia Volumes I&II**
May 2006
Doc ID: PN-ADG-304
Doc ID: PN-ADG-305

**The Importance of Social Investments for Economic Growth and Governing in Transitioning States: A Survey of Recent Evidence**
June 2006
Doc ID: PN-ADG-306

# AN EVALUATION FRAMEWORK FOR USAID-FUNDED TIP PREVENTION AND VICTIM PROTECTION PROGRAMS

## Final Report

December 2009

# Table of Contents

## Tables



## Figures

# List of Terms

| Term | Meaning |
|---|---|
| Baseline | Baseline data is the first measurement of a performance indicator. The baseline is used to learn about current levels and patterns of performance before the program intervention. It sets the current condition against which future change can be tracked (Kusek, 2004). |
| Benchmark | Interim indicators or markers of progress toward a goal (Weiss, 1996). |
| Beneficiaries | The individuals, groups, or organizations that benefit from the program intervention (Kusek, 2004). Direct beneficiaries may be those who receive training or services. Indirect beneficiaries may be those treated by trained staff, or the families and colleagues of those who received services. |
| Causality | *Also Causal attribution.* The claim that *x* caused *y*. In evaluation the claim that the program was responsible for the observed effect (Weiss, 1996). |
| Data | Individual facts, opinions, attitudes, or items of information (IOM, 2008). |
| Effect Size | The magnitude of the relationship between two variables. In evaluation, usually the measure of the magnitude of the relationship between program variables and outcomes (Weiss, 1996). |
| Evaluation | The systematic and objective assessment of an on-going or completed project, program or policy, its design, implementation and results. The aim is to determine the extent of implementation, efficiency, effectiveness, impact and/or sustainability, on or for a variety of targets or purposes (IOM, 2008). |
| Evaluation Framework | A map for carrying out an assessment of program work, measuring both the extent to which the program has implemented its planned activities and the extent to which the activities have resulted in achieving the intended objectives. A framework differs from an evaluation plan in that it does not spell out the details of implementing an evaluation for a given program. |
| Evaluation Plan | Articulation of an evaluation strategy with detailed activities for assessing progress, outputs, outcomes and impact, cost-effectiveness and sustainability of a program, intervention, or initiative. It specifies the type of data to be collected, frequency of data collection, data collection forms (and the rationale for them), sampling strategy, analysis strategy and plan, and the kinds of statements that will be able to be made, with what level of certainty and expected error rates. |
| Formative Evaluation | A type of evaluation conducted during the course of program implementation whose primary purpose is to provide information to improve the program under study (Weiss, 1996). |
| Goal | *Also Overall objective.* A single statement of the broader aim of a program, i.e., how the program can contribute to a larger national or international development plan or action (IOM, 2008). |
| Impact | Positive and negative, primary and secondary, long-term effects produced by a program intervention, directly or indirectly, intended or unintended (Kusek, 2004). Achieving impact is a broader, more difficult criterion than producing *outcomes* which is, again, a higher bar than producing *outputs*. |
| Indicators | Quantitative and qualitative factors or variables that provide a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of the program (Kusek, 2004). |
| Inputs | The financial, human, and material resources use for the program intervention (Kusek, 2004). |
| Intervention | Activities funded and implemented in order to have an impact on an identified problem, such as trafficking in persons. |
| Logic Model | A management tool used to improve the design of interventions, most often at the project level. It involves identifying the strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, performance indicators, and the assumptions or risks that may influence success or failure. It thus facilitates planning, execution and evaluation of a program intervention (Kusek, 2004). |
| Monitoring | A continuing function that uses systemic collection of data on specified indicators to provide management and the main stakeholders of an ongoing program intervention with indications of the extent of progress and achievement of program objectives (Kusek, 2004). |
| Objective(s) | The specific, desired program outcomes (Kusek, 2004). Objectives should be tied to the overall goal of the program, and be situated within a realistic appraisal of potential accomplishments given what others have been able to do, or what has been accomplished previously. |
| Operationalize | To translate general program inputs, processes, and goals into specific, measurable benchmarks and performance indicators (Weiss, 1996). |
| Outcomes | The likely or achieved short-term and medium-term effects of an intervention's outputs (Kusek, 2004). Outcomes should reflect the results of program activities and their impact on program goals. However, outcomes may not be broad enough to yield impact on addressing the problem of trafficking overall. |
| Outputs | Direct and measurable results expected from program activities. They should be tangible, visible and measurable outputs of program work. If they are sustainable beyond the activity, they may turn into program outcomes and have impact on the problem of trafficking overall (IOM, 2008). |

| Term | Meaning |
|---|---|
| Performance Indicator(s) | Pre-determined measurements that track specific changes or outputs of a program. Performance indicators are directly linked to measuring progress toward program objectives and are often a combination of monitoring and evaluation (IOM, 2008). |
| Process Evaluation | A study of what goes on while a program is in progress. Process evaluation relates to the phase of the program studied – in this case, program implementation (Weiss, 1996). |
| Program Audit | An independent, objective assurance activity designed to add value and improve a [program's] operations. Performance auditing is concerned with relevance, efficiency, and effectiveness (Kusek, 2004). |
| Selection Bias | The bias resulting from preexisting differences between program participants and the comparison group. Effects found at the conclusion of the evaluation may be due to the fact that different types of people were selected or selected themselves into the program and comparison groups (Weiss, 1996). |
| Stakeholders | Agencies, organizations, groups or individuals who have a direct or indirect interest in program work and outcomes there from, and who are affected positively or negatively by the implementation of activities (IOM, 2008). |
| Summative Evaluation | A study conducted at the end of a program (or of a phase of a program) to determine the extent to which anticipated outcomes were produced. Summative evaluation is intended to provide information about the worth of the program (Weiss, 1996). |
| Theory of Change | The assumptions that link a program's inputs and activities to the attainment of desired ends; it include both implementation theory and program theory (Weiss, 1996). |
| Trafficking in Persons (TIP) | The recruitment, transportation, transfer, harboring or receipt of persons by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or benefits to achieve the consent of a person having control over another person, for the purpose of exploitation. Exploitation shall include, at a minimum, the prostitution of others or other forms of sexual exploitation, forced labor or services, slavery or practices similar to slavery, servitude or the removal of organs (IOM, 2008). |

# Acronyms

| | |
|---|---|
| AEA | American Evaluation Association |
| E&E | Europe and Eurasia |
| GAO | Government Accountability Office |
| ILO | International Labor Organization |
| IOM | International Organization for Migration |
| M&E | Monitoring and Evaluation |
| NGO | Non-governmental Organization |
| RCT | Randomized Control Trial |
| SPSS | Statistical Program for the Social Sciences |
| TIP | Trafficking in Persons |
| TVPRA | Trafficking Victims Protection Reauthorization Act |
| VoT | Victim of Trafficking |
| UNODC | United Nations Office for Drugs and Crime |
| USAID | U.S. Agency for International Development |

# Executive Summary

The purpose of this report is to develop an evaluation framework for USAID prevention and victim protection programs that address trafficking in persons (TIP). An *evaluation framework* is an analytic tool designed to provide technical guidance for meaningful, reliable, and valid evaluation of specific program outcomes and impact. *Prevention* programs support campaigns focusing on public awareness, education, advocacy, income generation, and demand reduction. *Protection and victim assistance* programs provide shelters and targeted services for identified and potential victims of trafficking. Although there is near universal agreement about the fact that we must improve the impact evaluation of anti-TIP programs in order to enhance understanding of what works and what does not, limited information is available on *how* to do this. The purpose of this report, therefore, is to provide concrete guidance on *how* to evaluate anti-TIP programs to those who are designing and implementing such programs.

In preparing this report the authors conducted a review of evaluation frameworks, current evaluation literature and handbooks, and case studies related to anti-TIP initiatives. This report is intended to complement a number of previous reviews of counter-trafficking programs and proposals for indicators by providing a framework for evaluating anti-TIP programs typically implemented with USAID funding.

It should also be noted that the report was written for USAID and the staff of its implementing partner organizations. The report is designed to help them understand what is involved in evaluating anti-TIP program impact and to provide specific suggestions when planning evaluations.

The report is divided into five sections:

1. Foundations of an Evaluation Framework
2. Design Strategies for Evaluating an anti-TIP Program
3. Challenges to and Recommendations for Evaluating anti-TIP Programs
4. Sample Plan for Evaluating a TIP Prevention Program
5. Sample Plan for Evaluating a Victims of Trafficking (VoT) Protection Program

## Foundations of an Evaluation Framework

An evaluation framework serves as a model when developing an anti-TIP program evaluation plan. The framework helps to ensure that the evaluation does not focus solely on whether a program's objectives were achieved, but rather links the interventions to program impact. This allows evaluators to assess in what ways the interventions were an integral part of the achievement of the objectives or in what way the interventions failed to achieve the objectives. The issues that are important for building an effective evaluation framework are:

1. Understanding the purpose of the evaluation;
2. Recognizing the theory of change upon which the program is built; and
3. Developing the logic model.

The evaluators and stakeholders should be aware that impact evaluations are more costly and time-consuming than process evaluations or program audits. In general, the costs of the evaluation increase when studying impact – requiring more data forms, data collectors, collection from a larger sample of respondents, and a larger investment overall in evaluation design, logistics management, analysis and reporting. Impact evaluations generally require use of comparison groups or collection of data in a way that allows quantification of change from baseline data.

How the activities lead to the expected change is called the *theory of change*. The theory of change links a program's inputs and activities to the attainment of desired ends; it articulates both the implementation of the program and the steps that lead to program impact (Weiss, 1996).

Understanding and articulating these steps and connections is critical for evaluating any anti-TIP program. The logic model helps to articulate the theory of change embedded in anti-TIP programs. Components of the logic model are shown in the table below.

| Components of the Logic Model | |
|---|---|
| **Component** | **Definition** |
| Inputs | *Human and financial resources used for the program intervention. In anti-TIP programs, the inputs are often (but not always) the targeted beneficiaries of the program.* |
| Activities | *Actions taken or work performed through which inputs are mobilized to produce outputs.* |
| Performance Indicators | *Qualitative and quantitative measures or variables to be applied to the program activities. Performance indicators are directly linked to measuring progress toward program objectives and are often a combination of monitoring and evaluation. Interim performance indicators are called benchmarks.* |
| Pathways | *Linkages that specify how activities of a program lead to the expected outputs, outcomes, and impact of a program. Pathways specify and map performance indicators through each step of the logic model.* |
| Expected Outputs | *Direct and measurable results expected from program activities. They should be tangible, visible and measurable products of program work. If they are sustainable beyond the activity, they may turn into program outcomes.* |
| Expected Outcomes | *The short-term and medium-term effects of a program's outputs. Outcomes should reflect the results of program activities and their near-term effect on program goals. However, outcomes may not be broad enough to yield impact on addressing the problem of trafficking overall.* |
| Expected Impact | *The long-term effects produced by a program intervention, linked closely to the overall program goal. Such a goal could be as ambitious as reducing and preventing trafficking, but could equally be less ambitious for smaller or shorter term programs.* |
| Assumptions | *Hypotheses about factors or risks which could affect the progress or success of a program intervention. Our underlying beliefs about the program, the stakeholders or beneficiaries.* |
| External Factors | *Factors which are not explicitly in the control of the program but which can have an important effect on the intended outcomes and impact, such as government policies or changes in the trafficking situation in the country.* |

## Design Strategies for Evaluating an Anti-TIP Program

Once the purpose of the evaluation is clear and the theory of change and logic model have been specified, the evaluators will begin their task of specifying the evaluation questions. Evaluation questions drive the design of the instrumentation and data collection methods. Though the overall evaluation question may be general, such as, "Did the program reduce the vulnerability of the victims to trafficking?" the specified questions help identify the evaluation method and what kind of data will be needed to answer these questions.

The evaluation questions help the evaluator understand not only the type of data that will be needed to collect information, but also the evaluation method that will be used to measure program impact. Four evaluation methods are discussed:

1. Classical experimental design;
2. Longitudinal cohort analysis;
3. Longitudinal analysis of the treatment only; and
4. Cross-sectional design.

In a **classical experimental design** the evaluator constructs identical treatment and control groups. The *treatment group* receives the intervention, the *control group* does not. The evaluator must determine whether it is ethical to give the treatment to one group while denying it to the other. If it is determined that the control group will be harmed by not receiving treatment, this evaluation method should not be used. In most evaluations of VoT protection programs, classical experimental design cannot be used due to this reason. This design is best applicable in the evaluation of TIP prevention programs. In awareness raising activities, such as VoT identification training for immigration officials, the intervention may be given to one group and not another without causing harm.

*Consider classical experimental design when the program intervention can be given to one group and denied to another without causing harm to either party. This method is not recommended for most VoT protection programs, but could be used in evaluation TIP prevention programs.*

In **longitudinal cohort analysis**, the evaluator collects longitudinal data on a cohort (or group) of individuals and families representing the treatment and the comparison. A *comparison group* is a chosen group that does not participate in the program intervention, but unlike the control group it is not from the same population as the treatment group. Both groups are followed for the same time periods, and compared internally across time as well as with each other, by virtue of compiling indices of known characteristics to represent key features the evaluator has measured.

*Consider longitudinal cohort analysis when resources are available and data collection can occur before and throughout the program intervention. Program managers can build evaluation knowledge in terms of the level of incidence of trafficking.*

**Longitudinal analysis of the treatment group** only is much less expensive and time consuming than the longitudinal cohort study. This is one half of the longitudinal cohort group analysis. It can inform policy makers and planners how the treatment group performed in a program. However, the treatment group may have special characteristics, such as ethnicity, age, and gender, or other characteristics which are not obvious. As a result, one cannot generalize beyond the treatment group. The statements made would have to be qualified to reflect impact only for potential or actual victims identical to the ones treated.

*Consider longitudinal analysis of the treatment only to understand the impacts of a particular program. This method can generate reliable information about how the treatment group performed given an intervention.*

Finally, **cross-sectional data analysis** provides a snapshot comparison of a treatment and a comparison group at one point in time, usually after the program has started. The comparison group is selected *after* the intervention to match the characteristics of the treatment group before they entered treatment. The difficulties with this approach relate to whether the two groups are indeed similar, and what the differences might be. One benefit of this method is that cross-sectional data is less expensive to collect than longitudinal data. However, this strategy is generally not recommended as a stand-alone method, as it does not provide sufficient confidence for drawing conclusions about the intervention.

*Consider cross-sectional data analysis to make comparisons about treatment and comparison groups after a program intervention, with the understanding that the differences between the two groups may not be attributed to the program intervention solely. This type of analysis is most applicable in cases when data was not collected before the start of the program, and when the budget of the evaluation or program does not include baseline data collection.*

Sampling defines how many respondents have to be recruited in order to yield valid data that can be used to support decision making. Respondents chosen for both quantitative and qualitative anti-TIP

evaluations should be randomly selected to reflect the variety of the intervention population. For studies of anti-TIP activities, the sample size is likely to be fairly large, since the phenomenon being studied is complex and a variety of variables need to be taken into account. For each anti-TIP evaluation, evaluators and stakeholders have to decide how the data collected will yield key findings with what level of certainty. A poorly designed sample can jeopardize the utility of the whole evaluation. Finally, evaluators should adhere to strict ethical behavior when collecting data from vulnerable populations, such as victims of trafficking. Participation in an anti-TIP program or evaluation of the program may jeopardize the security of the victims or those vulnerable to trafficking.

## Challenges to and Recommendations for Evaluating Anti-TIP Programs

USAID's presence in the fight against trafficking worldwide has increased the agency's potential to meet some of the challenges to countering trafficking. With an integrated approach, Missions can increase their ability to evaluate the impact of these programs. Challenges commonly encountered while evaluating anti-TIP programs are as follows:

### *Unclear Evaluation Purpose*

When there is no agreement on what the evaluation purpose is or there is inadequate logic model scaffolding on which to build the evaluation, then the results of the evaluation will be inadequate. Evaluators that are asked to conduct a program evaluation should understand *how* the program has been designed, and what stakeholders have specified as the evaluation purpose. A useful evaluation framework links program work to its intended overall goal. It builds understanding as to what program effects mean, not only in the individual program context, but also in the larger anti-trafficking context in the country or region.

***Recommendation:*** *Hire professional evaluators who carefully review the logic model and develop evaluation questions in conjunction with stakeholders involved in the program design to improve the effectiveness of evaluations. Review the program in context of other work being done to identify common intervention components.*

### *Lack of Time and Funding*

Measuring the impact of interventions requires analysis of change over time, specifically as compared to a baseline, and generating baseline data can require substantial resources. Extensive improvements in the design and evaluation of international development programs have been made in recent years, including improvements in data collection techniques. Alternative evaluation methods can be utilized. For example, analyses can be done of only the participant group (longitudinal analysis of the treatment only), or programs can be evaluated after the program has completed using cross-sectional analysis.

***Recommendation:*** *Be strategic about when and how to do a cost-effective evaluation and design programs with evaluation in mind from the start. For anti-TIP evaluations, understand the constraints of the research methodology and look to techniques that other programs have used in addressing challenges.*

### *Inadequate Data Collection Procedures*

Rather than yielding a blanket statement about whether a program is effective or not, an evaluation framework should hone in on results considered essential for producing the outcomes. Anti-TIP evaluations require data that are reliable, valid, accurate, and that are useful for improving program

functioning and making decisions about allocation of resources and program focus. Indicators should be selected that reflect the actual impact the program was expected to produce.

***Recommendation:*** *Set priorities for information to be collected, based on the logic model; utilize clear, concrete, and authentic indicators to measure what is needed. Insist on consistent data collection techniques and provide training to those collecting the data.*

### *Selection Bias*

To address selection bias, evaluators should acknowledge the constraints of their sample in the evaluation design, and should discuss the characteristics of the sample and how they affect the explanatory power of the evaluation results. The quantitative and qualitative outcomes and impact of the evaluation must be attributed to the sample chosen. The biggest mistake that can be made is to relate the evaluation outcomes to the treatment population at large when the sample is not representative; the results will be invalid and false conclusions can be drawn.

***Recommendation:*** *Recognize selection bias from the beginning of the evaluation and clearly specify the characteristics of the sample and how it may affect the conclusions you will be able to draw about the program outcomes and impact. Discussion of selection bias should always be documented in the written evaluation report, particularly in the evaluation methodology.*

### *Definition of "Trafficking in Persons"*

In evaluating a specific program, the definition of TIP should be stated at the onset of the program and in the program design. As long as the evaluator has an operational definition of TIP, even if it is incomparable, he or she will be able to evaluate the program based on that definition. It is true that programs may then be difficult to compare with varying definitions, but for the types of individual impact evaluations considered in this report, an operational definition of TIP will be sufficient to measure program outputs, outcomes, and impact.

***Recommendation:*** *Identify the operational definition that was used for "trafficking in persons" in the program design; this definition should be used for the evaluation. Though one may not be able to define "trafficking in persons" for all countries and programs, this definition should clarify the phenomenon and overall problem that the intervention is seeking to effect. Once this definition is clear, maintain consistency in defining TIP this way.*

### *Lack of Criteria to Identify Victims of Trafficking*

In an evaluation of a specific anti-TIP program, one should look to the program design to understand the criteria used to identify VoTs. Who is included in the program interventions? How are the program interventions conceptualized and authorized? These are not easy questions to answer, but they will help specify the evaluation questions needed to measure program impact.

***Recommendation:*** *As with the definition of TIP, maintain an operational definition of VoTs, and use this definition to specify the criteria for identification. Based on this operational definition, decide which beneficiaries are to be targeted for the evaluation, and maintain the VoT criteria on hand when developing the performance indicators of impact.*

### *Confidentiality and Protection of Identity*

The evaluators should build the capacity of the local service providers to participate in the data collection process. These organizations may be the most relevant sources of information about the

local patterns of trafficking and the types of populations that are vulnerable to being trafficked in the community. The providers may have distinct definitions of "trafficking" and who victims of trafficking are. As long as these definitions are consistent with the program and evaluation definitions, then the results should be relevant.

***Recommendation:*** *Build the capacity of local organizations and service providers who have close connections with VoTs and vulnerable populations to collect data. Maintain a record of local patterns of trafficking to build a representation of what populations need services and who is at risk.*

### *Demonstrating Impact*

Over time, individual program evaluations should not be the only evaluations a Mission undertakes related to TIP. Long-term evaluation planning should review the range of programs for a Mission and for a sector overall, in terms of the types of and status of programs being implemented, and how evaluations can build knowledge about them. Thus, individual, stand-alone, program evaluations become part of a larger evaluation plan for Mission-wide and sector-wide counter-trafficking initiatives.

***Recommendation:*** *Collect and disseminate lessons learned about effective practices and their relative and absolute impacts for different groups and different ways of being vulnerable to or emerging from trafficking. In the long term, consider Mission-wide or sector-wide evaluation plans to ascertain how different TIP prevention and protection programs work together to affect the incidence of trafficking.*

### *Measuring Vulnerability and Prevention Success*

If vulnerability is the indicator of program success, a promising alternative for measuring vulnerability is constructing a vulnerability index. Similar indices have been created to study economic and environmental phenomena; however, not many indices have been constructed to address social issues.[1] In constructing a vulnerability index, the evaluator chooses the indicators or components of the index, and then collects categorical data (a score) for each of the components.

***Recommendation:*** *Consider constructing a vulnerability index to study changes in behaviors of vulnerability of program participants over time. Such an index, to be operational, would need to be based on solid research looking at a wide range of trafficked individuals, and specific criteria or variables of vulnerability to trafficking.*

### *Measuring the Incidence of Trafficking*

Individual evaluations of program impact should have data collection standards. The Mission should encourage the evaluators to submit research data and any information collected on the incidence of trafficking (whether quantitative or qualitative). Over time, the Mission will be able to consolidate information from various different counter-trafficking initiatives and this could be developed into a TIP database. Though there may be a lack of funding to maintain this database, the presence of a monitoring and evaluation system within the Mission to measure trafficking could greatly improve knowledge at all levels about the status of trafficking in the country and the region.

***Recommendation:*** *For evaluators, develop a database to help consolidate data collected for the impact evaluation (including background data on the incidence of trafficking), and submit this data to the Mission. For Mission staff, consider developing a TIP database of all data collected from various TIP programs, including interventions across other USG agencies, to understand the overall incidence of trafficking over time.*

[1] The United Nations has an Economic Vulnerability Index (EVI) to classify the development of its Member States.

# Introduction

The Trafficking Victims Protection Reauthorization Act (TVPRA) requires annual reporting on the progress of programs to counter trafficking in persons (TIP) funded by the U.S. Government, and USAID requires monitoring and evaluation of its anti-TIP programs in part to inform these reports. *Monitoring* is the act of ongoing program review for the purposes of seeing whether the program is meeting its objectives and delivering the funded activities at the required level to achieve intended results. *Evaluation* is the assessment of results and impact of the work done, with conclusions about effectiveness, either in the interim or at the conclusion of the program.

The purpose of this report is to develop an evaluation framework for USAID prevention and victim protection programs that address TIP. An *evaluation framework* is an analytic tool designed to provide technical guidance for meaningful, reliable, and valid evaluation of specific program outcomes and impact.[2]

## Purpose and Methodology

It has by now become standard practice to claim that it is difficult to pin-point best practices in anti-TIP programming because so few programs are evaluated and that the few evaluations that are carried out yield information about the program outputs (such as counts of trafficking victims provided with assistance, number of laws passed, and number of calls to a hotline) but little information about program outcomes or impact. Although there is near universal agreement about the fact that we must improve the evaluation of anti-TIP programs in order to enhance understanding of what works and what does not, limited information is available on *how* to do this. The purpose of this report, therefore, is to provide concrete guidance on *how* to evaluate typical USAID-funded anti-TIP programs and their impact to those who are designing and implementing such programs.

This report will present best practices in evaluating anti-TIP programs in prevention and victim protection. *Prevention* programs support campaigns focusing on public awareness, education, advocacy, income generation, and demand reduction. *Protection and victim assistance* programs provide shelters and targeted services for identified and potential victims of trafficking (VoTs). Some counter-trafficking programs also include prosecution and law enforcement interventions, which aim to affect social and criminal justice systems through training and technical assistance regarding investigation and witness management. However, USAID does not typically target support for these kinds of programs, and the framework presented here is not specifically designed to address them.

In preparing this report the authors conducted a review of evaluation frameworks, current evaluation literature and handbooks, and case studies related to anti-TIP initiatives. This report is intended to complement a number of previous reviews of counter-trafficking programs and proposals for indicators (see Listed References) by providing a framework for anti-TIP programs typically funded by USAID.

It should also be noted that the report was written for USAID and the staff of its implementing partner organizations. The report is designed to help them understand what is involved in evaluating anti-TIP programs and to provide specific suggestions for how to address them when planning evaluations. However, USAID and its partners should be sure to engage the services of evaluation experts to ensure that the evaluation is conducted in such a way that the findings are valid, reliable and applicable.

[2] See the List of Terms at the beginning of this report for definitions of terms like evaluation, impact, goals and objectives.

## Limitations

This report addresses the need for an evaluation framework and is not intended to provide specific evaluation guidance for each and every counter-trafficking program funded by USAID. Similarly, this report is unable to address all of the challenges in evaluating anti-TIP programs which have been noted by many reviews and reports. Some of these challenges are inherent in the measurement of the extent to which something did not happen (prevention) to measurement of criminal activities and hidden populations. These challenges are addressed in Section 3 and recommendations for how to mitigate them are put forward.

This report does not set out to specify universal indicators and benchmarks for anti-TIP programs; many good resources are already available which provide extensive lists of useful indicators for measuring anti-TIP programs (IOM 2008, Dottridge 2007a), and some of these are included in the Appendices. Rather, this report provides guidance on the decisions to be made throughout the evaluation process and provides illustrative examples of evaluation plans for sample prevention and protection programs. USAID missions or partners should be able to use this framework and sample plans to help them develop evaluation frameworks for their specific programs.

## How This Report is Organized

The report is divided into five sections:

1. ***Foundations of an Evaluation Framework*:** Outlines the purpose of evaluation and its application to linking interventions, outcomes and impact in the TIP context.
2. ***Design Strategies for Evaluating an Anti-TIP Program*:** Provides an overview of sampling, data collection, data analysis and reporting strategies for anti-TIP evaluations.
3. ***Challenges to and Recommendations for Evaluating Anti-TIP Programs:*** Summarizes the major challenges to evaluating counter-trafficking programs, and some recommendations to address these challenges.
4. ***Sample Plan for Evaluating a TIP Prevention Program***: Provides a conceptual approach and evaluation plan for an impact evaluation of a TIP prevention awareness program.
5. ***Sample Plan for Evaluating a Victims of Trafficking (VoT) Protection Program*:** Provides a conceptual approach and evaluation plan for evaluating impact of a victim protection and assistance program.

# 1. Foundations of an Evaluation Framework

An evaluation framework serves as a model when developing an anti-TIP program evaluation plan. The framework helps to ensure that the evaluation does not focus solely on whether a program's objectives were achieved, but rather links the interventions to their intended objective. This allows evaluators to assess in what ways the interventions were an integral part of the achievement of the objectives or in what way the interventions failed to achieve the objectives. Additionally, it is unwise to consider an evaluation framework static. As knowledge of what works and does not work for counter-trafficking programs evolves, the framework must evolve as well.

The issues below are important for building an effective evaluation framework and will be addressed in the following sub-sections:

1. Understanding the purpose of the evaluation;
2. Recognizing the theory of change upon which the program is built; and
3. Developing the logic model.

All of these issues follow in sequence. The first step is clarifying the purpose of the evaluation and what it is intended to assess. The second step, understanding the theory of change, is to recognize the way in which the intervention intends to have an impact. In reference to anti-TIP programs, one must understand the phenomenon of trafficking and how the intervention will change the status quo. The third condition of the framework is to develop the logic model.

## A. Purpose of the Evaluation

The first step in building the evaluation framework is to understand the purpose of the evaluation. Stakeholders and evaluators must agree on what they want to know from the evaluation and what they require from the program results. Typical evaluation purposes are listed in Table 1: process evaluation, program audit, and impact evaluation. Each of these will be discussed further below.

| Table 1. Main Evaluation Purposes | |
|---|---|
| 1. | Process Evaluation – Are there ways to strengthen the results and outcomes through tightening the implementation procedures? |
| 2. | Program Audit – Did the program do what it said it would, the way it said it would? |
| 3. | Impact Evaluation – Did the implementation of the intervention result in the expected results and outcomes, and in what ways? |

Another important distinction to make is whether the evaluation is formative or summative. *Formative evaluations* examine whether program results functioned in a way to yield the outcomes intended, or to suggest what might strengthen or replace them. The evaluator's intention is to help develop and strengthen the program. In Table 1, a process evaluation is an example of a formative evaluation.

*Summative evaluations* determine the extent to which anticipated outcomes were produced. These evaluations are intended to provide information about the worth of the program, and it is the role of the evaluator to render judgment on the program. Program audits and impact evaluations are usually examples of summative evaluations.

### 1. Process Evaluation and Program Audit

A *process evaluation* focuses on the implementation of the program and aims to study what was or was not implemented as was planned (Kusek, 2004). This evaluation strategy is similar to monitoring, but goes further in examining unanticipated changes in program implementation. A process evaluation is useful to show why implementation efforts are or are not on track and can be the basis for countermeasures, if necessary. A process evaluation does not give any information on program outcomes, because it takes place during the implementation of the program.

*Program audits* have been used frequently in evaluating anti-TIP programs, because they are relatively low cost and time effective. Many TIP prevention and VoT protection programs have focused their evaluations solely within the program and with an emphasis on program outputs. This is the lowest cost strategy, of course, for there are no coordination costs, no requirements for "unpacking" which outcomes are attributable to whom, and, since the questions here are focused on whether program activities met the standards required to achieve their objectives, there is no implied comparison. The trade-off has been that typically the outcomes cannot be generalized beyond the program and therefore do not address the larger issue of the program's impact on combating trafficking.

A program may be perceived to be effective if it is producing outputs. However, unless these outputs lead to the achievement of the overall objectives and these objectives can be perceived beyond "program walls," the findings are limited. For this reason, this report will move away from process evaluations and program audits, and will focus on impact evaluations.

### 2. Impact Evaluation

*Impact evaluations* attempt to identify the changes that occurred as a result of program interventions, and establish to what these changes can be attributed (Kusek, 2004). Changes might be attributed to the program itself or other conditions of the program environment.[3] The extent to which a counter-trafficking program is linked to other larger programs, or programs addressing new issues or contexts, may affect the complexity of the evaluation. This is because the analyses necessary to attribute causality to a specific set of activities become more involved.

This report, and particularly Section 2, will focus on technical ways to measure impact in TIP prevention and victim protection programs. The evaluators and stakeholders should be aware that impact evaluations are more costly and time-consuming than process evaluations or program audits. In general, the cost of the evaluation increases – requiring more data forms, data collectors, collection from a larger sample of respondents, and a larger investment overall in evaluation design, logistics management, analysis and reporting. Impact evaluations generally require use of comparison groups or collection of data in a way that allows quantification of change from baseline data. For this to be accomplished, there are at least two points of time in which data have to be collected, and the sample of respondents from which data are collected generally requires going outside the pool of program participants, and often beyond the end of a program.

In order to design an effective impact evaluation of anti-TIP protection or prevention programs, the evaluator must understand and utilize two key pieces of program design: the theory of change and the logic model. The following two sections provide explanations of these design components.

[3] The question of *causality*, or whether the impact on beneficiaries can be attributed solely to the program, is the hardest to tackle when designing a TIP evaluation (Asher, 1976). In the absence of data on counterfactual outcomes (that is, participants that were not exposed to the program), impact evaluations can be rigorous in identifying program effects if they can isolate the effect of the program from other factors and potential selection bias (Khandker et al., 2010).

## B. Theory of Change

Any anti-TIP program is designed with the intention of changing something, of having an impact. How the activities lead to the expected change is called the *theory of change*. The theory of change links a program's inputs and activities to the attainment of desired ends; it articulates both the implementation of the program and the steps that lead to program impact (Weiss, 1996). In order to understand the theory of change, we must consider the key questions in Table 2.

| Table 2. Understanding the Theory of Change |
|---|
| 1. What is the overall goal or objective of the program? |
| 2. Who are the stakeholders involved? |
| 3. For each intervention envisioned what are the immediate, mid-term and long-term expected outcomes? |
| 4. How do these outcomes lead to impact – achievement of the overall goal? |
| 5. In what way is each stakeholder involved in achievement of the overall goal? |

### *1. Linking Interventions to Impact*

The evaluation framework next takes into account how a specific program relates the change and the issues addressed to a desired group of beneficiaries and activities. The framework should hone in on results considered essential for producing the desired outcomes and impact. Some questions to consider when linking the framework to a specific program are summarized in Table 3.

| Table 3. Questions Linking Interventions to Impact |
|---|
| 1. How does this program relate to other work in the field? |
| 2. What data exist on the specific area of program focus? |
| 3. How do the interventions yield the desired results? |
| 4. How, and how much, will these results impact the problem toward the overall goal? |
| 5. Did the interventions target the right beneficiaries? |

*Q1. How does this program relate to other work in the field?*

In a series of interventions, individual programs might only address a part of the problem. Programs may focus on a specific subset of individuals which are not necessarily representative of the entire trafficked or vulnerable population. Or programs may target only one aspect of trafficking, prevention, for example, while not addressing law enforcement or victim protection. In instances where this is the case, the evaluator should either try to make use of findings from other programs that focus on the same issue with other communities or on programs that address other aspects of anti-TIP work, or limit the attribution of impact to the specific group or aspect addressed.

*Q2. What data exist on the specific area of program focus?*

In order to assess impact it is important to have both baseline data and benchmarks for previous achievements. Characteristics of victim and vulnerable populations vary greatly by age, ethnicity, and method of trafficking to name a few, so that one cannot assume that program impact will be the same for different target groups. Marginalized ethnic minorities may react differently to interventions than the general population; victims trafficked into the sex industry may react differently from victims trafficked

into other industries. Prior to collecting data, it is important to understand what is known about the various types of trafficking and the individuals involved, and how the data collected might or might not be representative of the trafficking system as a whole.

*Q3. How do the interventions yield the desired outcomes*?

In order to collect the correct data – data that will portray the power of a given intervention – it is important to understand how the intervention will lead to specific results. This, again, is to understand why an intended goal was or was not achieved. For example, with training interventions, it is not only necessary that the participants will master the key information, but that they believe it is credible, that they aspire to use it, and that they are able to deploy it appropriately and effectively. If these are the pathways to impact, then the standards for evaluation need to reflect them to be successful.

*Q4. How and how much will these results impact the larger problem?*

It is important to envision ahead of time how the program intervention is likely to contribute to combating the larger problem of trafficking. If an organization is primarily dealing with direct assistance for victims of trafficking, and does not address wider issues of development of networks of potential assistance or identification protocols, the impact statement needs to be focused on how the individual assistance leads to integration and empowerment of the individual. The evaluation should not make assumptions about how the assistance creates new resources in the home community or improves identification of victims.

Q5. *Did the interventions target the right beneficiaries?*

The interventions of a program may be well executed, but if they target the wrong beneficiaries then they will not yield the desired impact. If, for example, a protection program is designed to target young women trafficked into prostitution, but the vast majority of trafficking in the country is of young men for labor, then the entire program has been incorrectly targeted.

Stakeholders need to consider the intervention in detail, and make explicit all the assumptions about what the activities are, how they work together, and what is required from them in order for the program to be successful. A key tool for building this is the logic model.

## C. Logic Model

The logic model is the anchor of the evaluation framework. A logic model is a management tool used to map the design of program interventions. It involves identifying strategic elements (inputs, activities, outputs, outcomes, impact) and ways to measure them (performance indicators) and developing causal relationships (pathways). It should also include the assumptions and external factors that may influence success and failure. As already mentioned, the logic model should be developed as part of the original program design.[4] However, in cases where it is not, it is incumbent on the evaluators to understand what the relationships are and to develop a logic model as a basis for the evaluation.

[4] Many evaluations fail because the evaluation indicators are not well aligned with the intervention activities. Successful program design requires agreement on not just what are indicators of trafficking, but also on what are indicators of success – of prevention of trafficking and protection of individuals (IOM, 2008).

### 1. Components of the Logic Model

A logic model is made up of the components listed in Table 4. Most of these definitions are also included in the List of Terms at the beginning of this report.

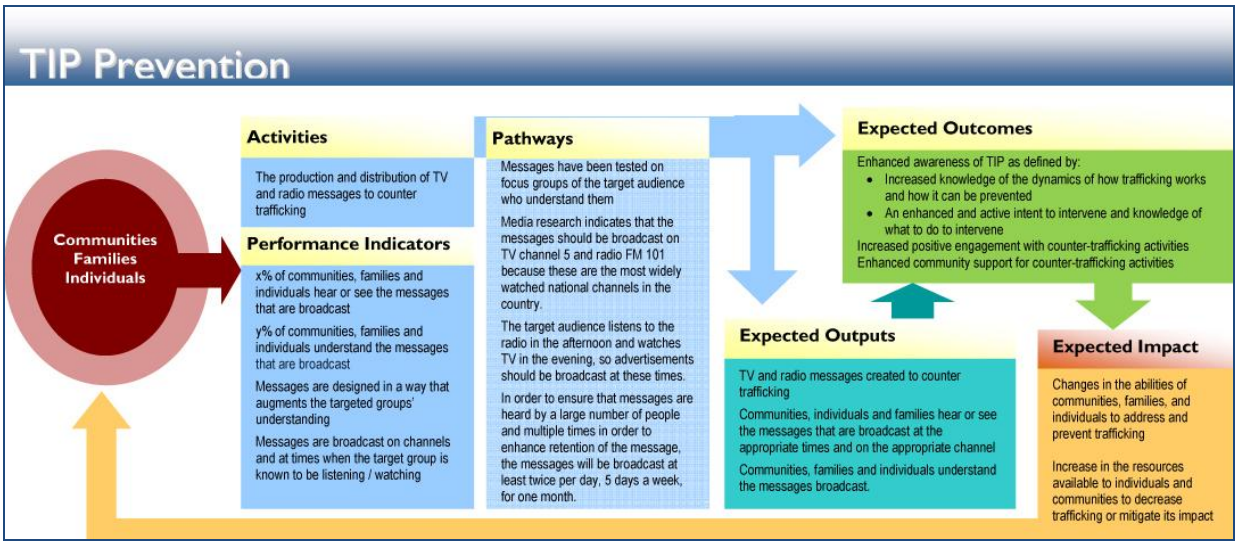
**Table 4. Components of the Logic Model - Definitions**

| Component | Definition |
|---|---|
| Inputs | *Human and financial resources used for the program intervention. In anti-TIP programs, the inputs are often (but not always) the targeted beneficiaries of the program.* |
| Activities | *Actions taken or work performed through which inputs are mobilized to produce outputs.* |
| Performance Indicators | *Qualitative and quantitative measures or variables to be applied to the program activities. Performance indicators are directly linked to measuring progress toward program objectives and are often a combination of monitoring and evaluation. Interim performance indicators are called benchmarks.* |
| Pathways | *Linkages that specify how activities of a program lead to the expected outputs, outcomes, and impact of a program. Pathways specify and map performance indicators through each step of the logic model.* |
| Expected Outputs | *Direct and measurable results expected from program activities. They should be tangible, visible and measurable products of program work. If they are sustainable beyond the activity, they may turn into program outcomes.* |
| Expected Outcomes | *The short-term and medium-term effects of a program's outputs. Outcomes should reflect the results of program activities and their near-term effect on program goals. However, outcomes may not be broad enough to yield impact on addressing the problem of trafficking overall.* |
| Expected Impact | *The long-term effects produced by a program intervention, linked closely to the overall program goal. Such a goal could be as ambitious as reducing and preventing trafficking, but could equally be less ambitious for smaller or shorter term programs.* |
| Assumptions | *Hypotheses about factors or risks which could affect the progress or success of a program intervention. Our underlying beliefs about the program, the stakeholders or beneficiaries.* |
| External Factors | *Factors which are not explicitly in the control of the program but which can have an important effect on the intended outcomes and impact, such as government policies or changes in the trafficking situation in the country.* |

### 2. Depicting the Logic Model

A graphical representation of the logic model for a fictitious TIP prevention program is presented in Figure A. The program intends to increase awareness about counter-trafficking through a media and radio campaign. The intended beneficiary groups of the program are communities, individuals, and families considered at risk of trafficking due to youth, low income and a high interest in migration.

**Figure A Sample Logic Model**




A logic model specifies the relationships between each component of the intervention, anticipated results, and the pathways through which it will contribute to desired outcomes. Each of these relationships, in turn, is analyzed in the program evaluation. Table 5 details each of the components of the logic model from Figure A.

**Table 5. Components of the Logic Model – Examples**

| ***Program: Media and radio awareness campaign to counter trafficking*** | | |
|---|---|---|
| Inputs | → | *Communities, families, individuals* |
| Activities | → | *The production and distribution of TV and radio messages to counter trafficking targeting the intended beneficiaries* |
| Performance Indicators | → | *x% of communities, families and individuals hear or see the messages that are broadcast*<br>*y% of communities, families and individuals understand the messages that are broadcast*<br>*Messages are designed in a way that augments the targeted groups' understanding*<br>*Messages are broadcast on channels and at times when the target group is known to be listening/watching* |
| Pathways | → | *Messages have been tested on focus groups of the target audience and the groups understand them*<br>*Media research indicates that the messages should be broadcast on TV channel 5 and radio FM 101 because these are the most widely followed national channels in the country*<br>*The target audience listens to the radio in the afternoon and watches TV in the evening, so advertisements should be broadcast at these times*<br>*In order to ensure that messages are heard by a large number of people and multiple times in order to enhance retention of the message, the advertisements will be broadcast at least twice per day, 5 days a week, for one month* |

| Table 5. Components of the Logic Model – Examples | |
|---|---|
| ***Program: Media and radio awareness campaign to counter trafficking*** | |
| Expected Outputs → | *TV and radio messages created to counter trafficking*<br>*Communities, individuals and families hear or see the messages that are broadcast at the appropriate times and on the appropriate channel*<br>*Communities, families and individuals understand the messages broadcast* |
| Expected Outcomes → | *x% will experience y*<br>*Enhanced awareness of TIP as defined by:*<br>• *Increased knowledge of the dynamics of how trafficking works and how it can be prevented*<br>• *An enhanced and active intent to intervene and knowledge of what to do to intervene* |
| Expected Impact → | *Changes in the abilities of communities, families, and individuals to address and prevent trafficking, and in the ability of authorities to combat it when it occurs*<br>*Increase in the resources available to individuals and communities to decrease trafficking or mitigate its impact* |
| Assumptions → | *Raising awareness and knowledge about trafficking in various demographic strata of the population will increase awareness for younger populations which are more prone to disregard the dangers*<br>*Successful media campaigns will lower the stigma of trafficking and will alert the public to the various facets of the trafficking system*<br>*However, engrained stigma may prevail among the population concerning trafficking who will disregard the campaign or consider it not their concern* |
| External Factors → | *The approval of government and legal authorities to air messages about trafficking, especially if controversial* |

### *3. Qualitative and Quantitative Performance Indicators*

Declaring that an activity has taken place does not indicate to what extent it has met the program goals or objectives. Qualitative and quantitative performance indicators are created to evaluate how the program has implemented the intended activities and to what extent it has achieved the intended results. Therefore, performance indicators are used for measuring the outputs, outcomes and impact relative to what was planned.[5]

Qualitative performance indicators measure whether an activity can be considered a successful intervention, based on the pathways specified. Taking the example of the TIP awareness program in Figure A, a potential question would be:

> *Is a single broadcast of an anti-trafficking message sufficient to communicate the message to the target audience?*

The answer, according to the logic model, is "no." The pathways indicate that the messages should be broadcast by a medium to which the target group is likely to listen, at a time when they are likely to be able to listen. If nobody listens, because they do not have access to a radio or they do not like the station, then the activity has not met its performance indicator. The same is true if people listen, but

[5] Two particular resources for identifying performance indicators are listed in the References, namely the *Handbook on Performance Indicators for Counter-Trafficking* (IOM, 2008) and *Measuring Responses to Trafficking in Human Beings in the European Union: an Assessment Manual* (Dottridge, 2007).

they are not the ones who could encounter or avoid trafficking. If the right people listen, but they dismiss the messages, or cannot deploy the information, the activity must also be deemed not to have met the criteria specified by the pathways.

Quantitative performance indicators measure how much change has to result for the intervention to be "significant." In the following example, once again from the TIP awareness raising program in Figure A, a possible question asked to establish quantitative performance is the following:

> *What percent of people who listen to the broadcast messages have to remember the message for the intervention to be considered successful?*

One must be careful in setting these performance indicators to ensure that they are neither too overly rigorous nor not rigorous enough. For example, does remembering the message necessarily imply understanding it? And do those who remember it, but who do not claim to feel motivated to use it actually use it in some way that they are not aware of? Is there some strategy to increase the likelihood that the listener will tell someone else, and how much is necessary to remember to potentially effect change?

Determining the quantitative and qualitative performance indicators, both for the intervention itself and for the results expected, is not primarily statistical – these are theoretical and practical questions for the program designer. Answers to them are based in understanding of the TIP problem, its manifestation in the target audience, and what happened with previous interventions of similar description.

# 2. Design Strategies for Evaluating an Anti-TIP Program

The components of an evaluation that significantly contribute to successful analysis of program impact are:

1. Specifying the evaluation questions;
2. Choosing the evaluation method;
3. Selecting the size and composition of the sample; and
4. Collecting and analyzing quantitative and qualitative data.

The following sections describe the options and make recommendations regarding appropriate methods to make these decisions and the foundation of an evaluation strategy that might be constructed. It should be understood that throughout this Section, we are considering both *quantitative and qualitative evaluations of program impact*. A technical guide to definitions of sample type, size, data collection techniques, and analysis are found in Appendix A.

## A. Evaluation Questions

Much of what has been discussed in Section 1 of this report is developed as part of program design, and the evaluators should have this knowledge at the onset of the evaluation. Once the purpose of the evaluation is clear and the theory of change and logic model have been specified, the evaluators will begin their task of specifying the evaluation questions. This is not as straightforward as it may seem, for how one frames the questions determines the nature of the findings. For instance, an evaluator may consider the three following questions:

1. Did the program meet the needs of the victims?
2. Did the program reduce the vulnerability of the victims?
3. Did the program reduce trafficking?

All three are reasonable questions, but extremely different in evaluation purpose and method. The first is a question of participant satisfaction, the second a question of effectiveness of the services over time for the participants, and the third a question that goes beyond the immediate participants to trafficking as a system. Evaluation questions drive the design framework – the rigorousness of the evaluation, the data to be collected, the methods to be used to collect them, and the statements that can be made as a result.

### *1. The Unit of Analysis*

Differences in the nature of the questions above relate to what evaluators call differences in the *unit of analysis*. Referring back to the three questions, the following are implications for the data in each instance:

1. Data relevant to the first question would need to capture participants' perceptions of program services. Data would be collected just from participants, not others outside of the program, and just about their own perceptions of whether the program met their needs.
2. Data collected for the second question would need to capture changes in the specific conditions related to the status and vulnerability of VoTs; a self-report alone would not be sufficient. Data here would be collected from the participants and their perceptions, but also from the results of training or job creation programs, from resources created in the community to reduce trafficking, and from secondary sources who can describe victim vulnerability, such as social workers or shelter personnel.

3. Data about whether a program reduced trafficking overall would need to capture information about the TIP system as well as the specific conditions influencing an individual's victimization. This third question poses a much higher bar for showing impact – it is much more difficult to measure and requires a longer timeframe for data collection. Data here would need to be collected at various times before, during, and after the program intervention, and from various sources.

### 2. *Specifying the Evaluation Question*

Evaluation questions drive the design of the instrumentation and data collection methods. The overall evaluation question may be general, such as, "Did the program reduce the vulnerability of the victims to trafficking?" The specified questions, however, help identify the appropriate evaluation methods and what kind of data will be needed to answer these questions. Table 6 provides the types of questions that should be considered.

**Table 6. Evaluation Questions in Depth**

| **Overall Question** | ***Did the program reduce the vulnerability of the victims to trafficking?*** |
|---|---|
| **Expanded Overall Question** | *The overall question in terms of "to what extent, for whom, etc." This can include quantitative measures (shelter interest increased by x percent) or qualitative measures (victims' accounts of program impact on them, their self-esteem)* |
| **Program Importance** | *The strengths and weaknesses of the program elements and which are the most important in reducing vulnerability to trafficking* |
| **Program Contribution** | *The importance of the program in the context of other work being done and how this program is making a difference in countering trafficking* |
| **Program Effects on Others** | *The program effect not only on the intended beneficiaries (VoTs) but also their families, communities, and service providers* |
| **Program Effects on Resources** | *The program contribution to resources for counter-trafficking initiatives (hotlines, shelters), but also to VoTs themselves and their families and communities* |

These are not intended to be the final questions. However, given that evaluation is expensive, it is important to be clear about what "the real questions" are, what information will be useful, how it will be used, and what kinds of findings are necessary to make future decisions.

***Recommendation:*** *Expand the evaluation questions to specify the data that will need to be collected to gather the right information. Understand the unit of analysis and the level of specification for determining impact. The purpose of the evaluation and the evaluation questions should be designed, discussed and approved by the evaluators and the program stakeholders alike to ensure that the evaluation results are relevant and evaluate the intended (or unintended) goals of the program.*

## B. Evaluation Methods

The evaluation questions help the evaluator understand not only the type of data that will be needed to collect information, but also the evaluation methods that will be used to measure program impact. Considering the Expanded Overall Question in Table 6, we need to think how we would be able to measure quantitatively that interest in shelters increased, or how we could conclude that VoTs feel that they have higher self esteem from personal accounts (qualitative data). Four evaluation methods are

discussed in this section: classical experimental design, longitudinal cohort analysis, longitudinal analysis of the treatment only, and cross-sectional design.

All of these designs are built on the construction of treatment, control or comparison groups. A *treatment group* is the group that participates in the program intervention. The *control group* is a group chosen randomly from the same population as the treatment group but that does not participate in the program. A *comparison group* is a chosen group that does not participate in the program intervention, but unlike the control group it is *not* from the same population as the treatment group. One of the biggest challenges with having a comparison group is the possibility of *selection bias*, where preexisting differences between the treatment and comparison groups may influence the evaluation outcomes. Selection bias is discussed further in Section 3, Challenges and Recommendations.

### *1. Classical Experimental Design*

In a classical or quasi-classical experimental design the evaluator constructs a treatment and control group, and randomly assigns (or matches) members to each group. This type of experiment, also called randomized control trial (RCT), is the best fit for interventions that clearly distinguish the presence or absence of the treatment.[6] Before applying treatment the evaluator must determine whether it is ethical to give the treatment to one group while denying it to the other (the control). If it is determined that the control group will be harmed by not receiving treatment, this evaluation method should not be used. In most evaluations of VoT protection programs, RCT cannot be used due to this reason.

RCT can only be employed in experiments where it is unclear which method (treatment or control) is better. In medical trials, for example, RCT is stopped when the treatment is shown to be "proven" one way or another (for example, when a vaccine is proven to work). It would be unethical to continue denying the control group medication, if it is clear that the treatment (vaccine) works. In trafficking, especially reintegration services, when one knows ahead of time that "not providing assistance" could directly harm the control group, it would be ethically unsound to recommend this as a strategy.

RCT is best applicable in the evaluation of awareness programs in counter-trafficking. Specifically in awareness raising activities, such as VoT identification training for immigration officials, the intervention may be given easily to one group and not another. In successful RCT design, the evaluators would give training to one group of officials and study if the training increases their ability to identify VoTs and traffickers in comparison to a control group from the same population of officers that did not receive the awareness training.

***Recommendation:*** *Consider classical experimental design (RCT) when the program intervention can be given to one group (the treatment) and denied to the control group without causing harm to either party. RCT is not recommended for most VoT protection programs, since it is unethical to deny VoTs assistance services for the purpose of the experiment. However, RCT is appropriate for certain awareness raising activities, such as training of officials, or other prevention activities.*

[6] In education studies, for example, analyses of curriculum impact are sometimes done with RCT. In these cases there is an intervention (participation in a curriculum offered), and treatment and control students vary by gender and age. Gender and age are controlled as part of the random assignment. This case is appropriate for RCT: the treatment is clear; the participants are definite; and the variables to match them on are clearly understood and can be controlled.

### 2. *Longitudinal Cohort Analysis*

In longitudinal cohort analysis, the evaluator collects longitudinal data on a cohort (or group) of individuals and families representing the treatment and the comparison. A cohort sample is a specific type of sample, with the same individuals providing data about their experience from the start to the finish of the evaluation. Both groups – treatment and comparison – are independently representative of the populations from which they are drawn.

Both groups are followed for the same time periods, and compared internally across time as well as with each other, by virtue of compiling indices of known characteristics to represent key features one has measured. In counter-trafficking programs, longitudinal analysis provides the benefit of being able to understand what happens to members of both groups over time, and the change they experience that brings them in or out of vulnerability or victimization from trafficking. It should be noted, however, that this evaluation method is costly and time-intensive.

A key decision issue is, "How long do the two groups need to be followed?" This, in fact, is more of a technical TIP question than a statistical question. From a purely statistical point of view, groups should be followed long enough so that the effects of the intervention can be borne out – i.e., that results are demonstrable, and that results are sustainable. From a practical standpoint, it very much depends on the nature of the interventions and the need to manage scarce resources. At a minimum, we would recommend one year.[7]

***Recommendation:*** *Consider longitudinal cohort analysis when resources are available and data collection can occur before and throughout the program intervention. This analysis is recommended because it allows the application of a variety of statistical techniques that could be useful to anti-TIP assessment. Program managers can build evaluation knowledge on the level of incidence of trafficking, as well as factors that foster or mitigate it, and (from the relative extent of change in the treatment and comparison group) the value of specific interventions. Victim prevention and awareness programs are good candidates for this type of analysis, as are certain victim protection programs. However, VoTs may not want to be part of a long-term evaluation. It is also necessary to consider if the comparison group will agree to be followed over time.*

### 3. *Longitudinal Analysis of the Treatment Group Only (Self-Comparison)*

Longitudinal analysis of the treatment group only is much less expensive and time consuming than the longitudinal cohort study. Moreover, it can provide the basis for assessing whether a change could occur by chance, based on the strength and nature of observed differences. However, the treatment group may have special characteristics, such as ethnicity, age, and gender, or other characteristics which are not obvious. As a result, one cannot generalize beyond the treatment group. The statements made would have to be qualified to reflect impact only for potential or actual victims identical to the ones treated.

This is one half of the longitudinal cohort group analysis. It can inform policy makers and planners how the treatment group performed in a program. Other data would be needed to see if this happened by chance (e.g., there were other factors operating, such as the state of the economy), that would lead to under or over-representation of the significance of the impact.

[7] For example, a variety of organizations have looked at what "retention rates" mean, especially in job retention (e.g., http://www.urban.org/publications/310360.html), how they vary by industry and over time. The variance across contributing factors is large, and thus this must remain a technical decision, rather than one dictated by statistics.

***Recommendation:*** *Consider longitudinal analysis of the treatment only to understand the impacts of a particular program and especially of direct assistance interventions. When other options are available, this is not recommended as a standalone technique as it does not allow results to be compared to a non-treatment group. However, it can generate reliable information about how the treatment group performed given an intervention, and in the context of how individuals performed prior to participation (through collection of baseline data).*

#### *4. Cross-Sectional Data Analysis*

Cross-sectional data analysis provides a snapshot comparison of a treatment and a comparison group. The evaluator compares data on two independently developed samples (treatment and comparison) after treatment was received (at one point in time) to see if the two groups exhibit the same characteristics, or how they differ. The comparison group is selected *after* the intervention to match the characteristics of the treatment group before they entered treatment. In a sense, this approach replicates classical design, and purports to have a treatment and comparison group that have identical characteristics; however, the data is collected after the treatment has been applied.

The difficulties with this approach relate to whether the two groups are indeed similar, and what the differences might be. Trafficking is a very complex phenomenon that can be reinforced or undercut by psycho-social and social network factors; therefore it is difficult to construct a group (especially after treatment) that can serve as a true comparison. Thus many of the differences which seem like they may have resulted from the treatment could have resulted from differences in the two groups at the outset.

This strategy is generally not recommended as a stand-alone method, as it does not provide sufficient confidence for drawing conclusions about a specific intervention. However, one benefit of this method is that cross-sectional data is less expensive to collect than longitudinal data. In order to be cost-effective, a mixed methods approach may be helpful, where cross-sectional data are supplemented by longitudinal data. For evaluating direct assistance programs, collecting cross-sectional data from a comparison group may be the only way to get information and could supplement longitudinal data from the treatment group.

***Recommendation:*** *Consider cross-sectional data analysis to make comparisons about treatment and comparison groups after a program intervention, with the understanding that the differences between the two groups may not be attributed to the program intervention solely. This type of analysis is most applicable in cases when data was not collected before the start of the program, and when the budget of the evaluation or program does not include baseline data. Cross-sectional data analysis may be combined with longitudinal analysis in certain instances to supplement comparison data.*

## C. Sample Selection and Size

Sampling defines how many respondents have to be recruited in order to yield valid data that can be used to support decision making. The design of the sample depends on the evaluation questions and the evaluation method chosen. A *random sample* is chosen following standard scientific methods, by using a list of the population and selecting subjects based on a number generated by a machine or table. Respondents chosen for an anti-TIP evaluation, both quantitative and qualitative, should be randomly selected to reflect the variety of the intervention population. The sample should also be stratified, so that it takes into account characteristics of the population that may influence outcomes. This usually begins with gender, age, and ethnicity, but for trafficking can include the geographical area, method by which the victim was trafficked, and others.

A respondent group is part of a *purposive sample* when it is not selected randomly. This means respondents were selected to meet some external criterion, perhaps the availability of respondents or their willingness to participate (this is particularly true in anti-TIP evaluations). Purposive samples can be biased based upon the influence of the external criterion. Because the exact nature of the internal biases for these samples are unknown – such that one cannot tell what the bias may affect – no statements generalized from findings developed from such samples can be made.

Without using a systematic and standard procedure for randomly selecting respondents, one cannot hope to eliminate the contamination of the findings by a biased sample. This is particularly important if one intends to make conclusions about overall impact or effective practice. It is important to understand that certain biases may not be completely eliminated. However, if one follows as strictly as possible random selection procedures, one is less open to undermining the findings from the start.

***Recommendation:*** *Select treatment, control, and comparison groups based on the evaluation method used, with preference given to random, stratified samples. Purposive sampling may be applicable for certain evaluations; however, this type of sampling is prone to bias, which can affect the evaluation results.*

Choosing an appropriate sample size relates to the issues discussed earlier about a well specified logic model and evaluation questions. For studies of anti-TIP activities, the sample size requirement is likely to be fairly large, since the phenomenon being studied is complex and a variety of variables need to be taken into account. This is equally true for prevention programs that target large portions of the population – such as awareness programs – and protection initiatives, because of the multi-faceted indicators of victimization, and the diverse objectives the interventions have to achieve. For each anti-TIP evaluation, evaluators and stakeholders have to decide how the data collected will yield key findings with what level of certainty. A poorly designed sample can jeopardize the utility of the whole evaluation.

***Recommendation:*** *Determine sample size based on the program to be analyzed and the level of impact the evaluation is expected to measure. Since trafficking is a complex phenomenon, a fairly large sample will be needed to measure changes in the trafficking phenomenon in a country – one that is representative of the differing types of communities, ethnicities, and circumstances within which TIP functions.*

## D. Data Collection and Analysis

Data collection is the next step in the evaluation design process. Data can include information about attitudes, values, knowledge, behavior, budgetary allocations, and service patterns (Weiss, 1996). It can come from various different sources, and *should not* come solely from program participants. In order to gain a holistic view of the program intervention, the evaluators may want to collect surveys and conduct interviews with participants, but also review program documents and budgets, products created (such as pamphlets or training manuals), and perhaps even observations of places or organizations providing services, such as shelters or border control stations. When available, the evaluators can also choose to conduct focus groups of participants, organizations, stakeholders, or service providers. The benefits of each data collection method are specified in Appendix A.

The evaluation method and sample design will specify also when data should be collected. For longitudinal analysis, for example, data should be collected at least two times: prior to and after the program intervention. If time and funding is available, data can be collected at the midpoint as well, and can be incorporated into a midterm evaluation. If the evaluation is an impact evaluation, and baseline data are not available, the next feasible option may be to collect data only once and after the program has finished, or ex-post. In monitoring, data may be collected more frequently, perhaps on a monthly,

quarterly, biannual, or annual basis. However, the amount of data to be collected will always be determined by the type of evaluation design envisioned, and the time and funding available.

### *1. Ethics in Data Collection*

Many program interventions in counter-trafficking are extremely sensitive to both participants and researchers, and therefore maintaining strong ethical standards is of utmost importance. It cannot be overemphasized that some of the participants in TIP prevention and victim protection programs may be risking their security by choosing to enter shelters, or even call a hotline. Ethical issues affect all stakeholders in a program intervention: the participants, the researchers, and the sponsoring organization (Kumar, 1999).

Participants in anti-TIP evaluations may range from VoTs to government workers, vulnerable groups, youth, or social and health workers. For each of these groups, the same ethical code should be in place. Four factors are essential: consent, sensitivity, "do no harm," and confidentiality. First, in every discipline it is unethical to collect information without the consent of the participant (Kumar, 1996). Second, the researcher must be sensitive to the topic that he or she is asking about, and this is especially relevant in trafficking. For most people, questions about sexual behavior, drug use, or criminal activity are considered private. Yet all three of these are topics that counter-trafficking programs are often centered around. Therefore, the researcher must be sensitive to the anxiety of participants in answering these questions or providing personal information. The researcher must also maintain the confidentiality of the respondents in order to ensure continued trust and participation in the study.

Ruth Rosenberg speaks extensively about "do no harm" in victim protection services (2008). Similarly, evaluators should aim to "do no harm" in involving participants in program evaluations. If a participant's involvement is likely to cause harm, the evaluator and the participants must be aware of this and may consider alternate measures. The safety of the participant, and the researcher, is the most important.

Ethics in data collection on the part of the researcher includes avoiding selection bias, the ethics (in some instances) of depriving treatment to create a control group, using information for the appropriate means, and reporting findings in an ethical way. Finally, some time should be spent on the ethics of the sponsoring organization, or the stakeholder of the program evaluation, who is bound to ethics of research as well. Sponsoring organizations may choose to impose direct or indirect controls on the methodology, data collection, or publication of certain data. If the sponsoring organization imposes restrictions that may stand in the way of obtaining or disseminating accurate information, this is unethical (Kumar, 1999).

***Recommendation:*** *Adhere to strict ethical behavior when collecting data from vulnerable populations, such as victims of trafficking. Understand that participation in an anti-TIP program or evaluation of the program may jeopardize the security of the victim or vulnerable group. Use care to prioritize the safety of all members of the evaluation, including the program implementers, evaluators, participants, and other stakeholders.*

### *2. Data Analysis and Report Writing*

Data analysis and report writing are an essential component of the evaluation process; however, the details of data processing, statistical analysis, presentation and interpretation should be left to evaluation and statistical specialists, and are outside the scope of this report. Further information and step-by-step guides for processing, analyzing and presenting data are included in Kusek's *Ten Steps to a Results-Based Monitoring and Evaluation System* (2004), or Kumar's *Research Methodology* (1999) chapters on Processing Data (Chapter 15), Displaying Data (Chapter 16), and Writing a Research Report (Chapter 17). Other works cited in the References can also provide useful tips to analyzing and presenting data.

## E. A Successful anti-TIP Evaluation Plan

In section 1, we discussed the importance of understanding counter-trafficking programs in context with other interventions being done in the same area and at the same time. We also discussed that there are certain decisions that need to be made in order to create a solid logic model. Developing a cost-effective anti-TIP evaluation involves being sure to ask (and answer) key questions for the stakeholders involved. Program implementers, funders, policy makers, and beneficiaries all have their own concerns, and the evaluators, often having limited resources, must negotiate the questions to be asked and answered early into the evaluation process.

Keeping in mind the evaluation framework from section 1 and adding the design recommendations presented in Section 2, Table 7 outlines a successful evaluation plan of a model anti-TIP impact program, which aims to measure change by collecting annual household surveys of vulnerable populations. Though this may not be part of a specific prevention or protection program, it is an evaluation that the Mission can fund to show the impact of several Mission initiatives on the trafficking system as whole. It is included here simply to show the various ways in which the techniques from Section 2 can be applied.

The evaluation in Table 7 employs a mixed-model approach, with quantitative and qualitative data collected to show the impact of various interventions on vulnerability to trafficking. The sample is well represented within a longitudinal sample of vulnerable groups, where some members of the community are likely to be TIP-vulnerable, and others less so.

**Table 7. Anti-TIP Evaluation Plan**

**Evaluation Question:**

How do various interventions (both prevention and protection) funded by the Mission impact household vulnerability to trafficking?

**Evaluation Design:**

- Evaluation Method: Longitudinal cohort analysis, which includes:
  - Comparison group: Household longitudinal survey (annual) on trafficking vulnerability and actual VoTs in hot areas where services have been offered.
  - Treatment group: Sample of households in selected prevention or protection programs to be evaluated.
  - Special case studies: Analysis of family members and key infrastructure representatives to provide a baseline for the description of the trafficking system, and changes in it over time to expand understanding of the meaning of findings related to each TIP intervention areas.
- Sample Selection: Random sample of vulnerable communities (with vulnerability determined based on households' scores on a vulnerability index). The unit of analysis is the household.
- Sample Size: The sample size ranges from 100-150 households each for comparison and treatment groups (this is representative).

**Data Collection and Analysis:**

- Data collection methods: Household surveys, interviews, background statistics, intake data, confirmation with formal justice system statistics when possible.
- Analysis and Reporting: Program intervention effectiveness, specified with statistical significance and effect size.

In addition to the evaluation techniques mentioned in Table 7, evaluators should effectively *manage* all steps of the evaluation process. This includes ensuring active and sustained involvement of program managers, technical specialists in TIP, evaluation specialists in TIP, and other key stakeholders in the evaluation. Management of the process should also include training of data collectors to ensure consistency in data collection methodologies as well as the ethical collection of data.

# 3. Challenges to and Recommendations for Evaluating Anti-TIP Programs

The following section summarizes the challenges to evaluating impact of TIP prevention and VoT protection programs and provides recommendations for addressing these challenges. It must be stressed that many of these challenges have been present since the inception of counter-trafficking programs, and are inherent to trafficking as a system. Some, such as the fact that victims are a hidden population, are not likely to be solved in the near future. Some involve ethical issues as well; if victims do not consider themselves to be victims or do not want to be found, there are only so many program interventions that can be done to attempt to change their minds.

This being said, USAID's presence in the fight against trafficking worldwide has increased the Agency's potential to meet some of these challenges. Many USAID Missions already fund counter-trafficking initiatives or programs which have an effect on trafficking, such as campaigns to combat violence against women, to increase income-earning opportunities for the poor and vulnerable, to expand girls' education, and to promote anticorruption efforts and legislative reform.[8] With an integrated approach, Missions can increase their ability to evaluate the impact of these programs.

## A. Unclear Evaluation Purpose

When there is no agreement on what the evaluation purpose is or there is inadequate logic model scaffolding on which to build the evaluation, the results of the evaluation will be inadequate. Evaluators that are asked to conduct a program evaluation should understand *how* the program has been designed, and what stakeholders have specified as the evaluation purpose. Program design may often be unclear, and it is up to the evaluators to clarify with the program implementers and stakeholders what the purpose of the evaluation (and perhaps even the program) is. If this is not done, evaluators may misinterpret the purpose of the evaluation and deliver a product that is not deemed relevant or representative by the stakeholders. Similarly, if the purpose of the evaluation is not clearly specified by program implementers or stakeholders, the evaluation may deliver results that are wholly irrelevant.

As stated in the sub-section "Linking Interventions to Impact" in Section 1, a useful evaluation framework links program work to its intended overall goal. It builds understanding for what program effects mean, not only in the individual program context, but also in the larger anti-trafficking context in the country or region (how much progress overall is being made through what types of interventions and their sustainability).

***Recommendation:*** *Hire professional evaluators who carefully review the logic model and develop evaluation questions in conjunction with stakeholders involved in the program design to improve the effectiveness of evaluations. Review the program in context of other work being done to identify common intervention components, what is known about them in the context in which they are implemented, and what evaluations have been done before.*

## B. Lack of Time and Funding

Evaluators come up against two major constraints in designing their evaluations: time and funding. Stakeholders want to find out the most about their program, in the least time possible and, of course, for the lowest cost. Measuring the impact of interventions requires analysis of change over time,

[8] See *Trafficking in Persons: the USAID Strategy for Response* (USAID, 2003).

specifically as compared to a baseline, and generating baseline data can require substantial resources. The evaluation technique used directly affects the cost and time of an evaluation. For example, if the evaluators decide to track a participant sample over time, from baseline to finish, with a comparison sample of individuals who did *not* participate in the intervention, both time and cost will be very high. The evaluators will have to be present for most of the lifetime of the project (which can take years), and quality control will cost considerable time and money.

For many impact evaluations in international development, and especially in trafficking, this type of analysis may not be feasible. One can only imagine the challenges of tracking a TIP shelter program over time, and attempting to keep up with former VoTs, who may have overcome their past and do not want to continue reliving what once happened. Comparison groups are almost impossible to capture in anti-TIP evaluations, and this will be shown further in the next two sections on sample evaluation plans.

Extensive improvements in the design and evaluation of international development programs have been made in recent years, including improvements in data collection techniques and forums to share information across thematic areas and countries.[9] Therefore, alternative evaluation methods can be utilized. For example, analyses can be done of only the treatment group (longitudinal analysis of the treatment only), or programs can be evaluated after the program has completed (cross-sectional analysis).

***Recommendation:*** *Be strategic about when and how to do a cost-effective evaluation and design programs with evaluation in mind from the start. For anti-TIP evaluations, understand the constraints of the research methodology and look to techniques that other programs have used in addressing challenges (discussed more in the Section 2, Evaluation Design). Consider using a mix of quantitative and qualitative techniques (a mixed-methods approach) and consider using performance indicators and proxy indicators that capture various program components. Consider evaluating the context of trafficking before the program intervention begins (an ex-ante evaluation). If this is not feasible, gather some baseline data at the onset of the project.*

## C. Inadequate Data Collection Procedures

If the evaluation design and implementation are plagued by missing data or poor analysis, the findings will be irrelevant and inadequate. Even with the best evaluation plan, if much of the sample (or the sample's data) is dropped or misconstrued, or the evaluation instruments do not accurately capture the sample's characteristics, bias is introduced rendering the findings unreliable.[10]

Rather than yielding a blanket statement about whether a program is effective or not, an evaluation framework should hone in on results considered essential for producing the desired outcomes. Anti-TIP evaluations require data that are reliable, valid, accurate, and that are useful for improving program functioning and making decisions about allocation of resources and program focus. Indicators should be selected which reflect the actual impact the program was expected to produce.

While it may be impossible for any evaluator to collect data on the complete population of victims of trafficking or vulnerable populations, this does not mean that one should give up on having consistent data collection and interpretation techniques. If a program has high dropout rates, maintaining a good

[9] Many of the publications listed in the References, such as *Handbook on Impact Evaluation* (Khanker et al., 2010) and *RealWorld Evaluation* (Bamberger et al., 2006) are great tools in considering alternative evaluation strategies in less-than-perfect situations. *Real World Evaluation*, for example, has a whole chapter on time and funding constraints.

[10] For more on the problem of attrition of the sample, see pages 226-228 in Weiss (1996).

case record may provide the evaluator with valuable information, especially when participants may not be available for interviews after the intervention. This also includes teaching the local service providers the importance of keeping consistent and valid data over time.

***Recommendation:*** *Set priorities for information to be collected, based on the logic model; utilize clear, concrete, and authentic indicators to measure what is needed. Insist on consistent data collection techniques and provide training to those collecting the data. Maintain data and case studies throughout the evaluation period. If the evaluation is over more than one year, collect information and build profiles based on the local patterns of trafficking.*

## D. Selection Bias

Selection bias can be introduced through the methods used to select participants for the evaluation. If chosen improperly, the sample of participants chosen for the program evaluation may be significantly different from the larger population receiving services. Examples of selection bias include:

- *Inflating findings:* The evaluation focuses on participants who voluntarily want to be included in the program evaluation, and are active participants in the program. Since the culture of participation often leads to positive support, there is generally little variation in findings when participants are asked for their opinions of services, with all saying the services were successful and no changes could possibly be envisioned. However, this is not always the case.
- *Differentially applicable findings:* The evaluation focuses only on victims in a single geographic area. This area may represent only one ethnicity or other sub-group characteristic; therefore it may be difficult to generalize the data drawn from a single sub-group of participants, when others may have fared differently.
- *Deflating findings:* The evaluation under-represents the power of the project by trying to assess change of the trafficking system as a whole. A trafficking system is a complex phenomenon that takes time to change, and it may be difficult to capture changes that have actually taken place. In order to see change in a system, the evaluation needs to clearly situate itself within the map of the trafficking problem in the area; and its components have to have characteristics that are designed to specifically impact it.

In evaluations of anti-TIP programs, particularly victim protection programs, selection bias is difficult to avoid. One cannot collect data from people who want to remain anonymous or who have not been identified as VoTs. To address this, evaluators should acknowledge the constraints of their sample in the evaluation design, and should discuss the characteristics of the sample and how they affect the explanatory power of the evaluation results. The quantitative and qualitative outcomes and impact of the evaluation must be attributed to the sample chosen. The biggest mistake that can be made is to relate the evaluation outcomes to the treatment population at large when the sample is not representative; the results will be invalid and false conclusions may be drawn.

***Recommendation:*** *Recognize selection bias from the beginning of the evaluation and clearly specify the characteristics of the sample and how it may affect the conclusions you will be able to draw about the program outcomes and impact. Discussion of selection bias should always be documented in the written evaluation report, particularly in the evaluation methodology. If findings are compromised by selection bias, attribute evaluation results to the sample chosen only, never to the treatment population at large. Though parallels may be drawn to the treatment population, these are not significant and can lead to misconstrued interpretations of the program impact.*

## E. Definition of "Trafficking in Persons"

Variations in the definition of TIP hinder the comparability and consolidation of data across countries and organizations. Persons identified as irregular migrants in one country may be registered as victims of trafficking in another. In many countries, cases of labor exploitation meeting the international definition of trafficking may be filed as cases of fraud rather than trafficking. These differences make transnational data collection difficult but also make it difficult to track changes over time as the understanding of trafficking may change within a country.

USAID acknowledges the United Nations' definition of trafficking in persons, based on Article 3 of the United Nations Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children (USAID, 2003). This UN Protocol came into effect in December 2003 and has 117 countries as signatories and 118 countries as parties to the Protocol (IOM, 2008). The 2003 USAID Strategy for Response in Trafficking for Persons recognizes this definition of TIP, which, however, does not mean that each individual program or country will. If a precise definition of TIP is missing in the program to be evaluated, the UN definition should be used to maintain consistency.

In evaluating a specific program, the definition of TIP should not be an issue as long as it is defined at the onset of the program and in the program design. As long as the evaluator has an operational definition of TIP, even if it is in incomparable, he or she will be able to evaluate the program based on that definition. It is true that programs may then be difficult to compare with varying definitions, but for the types of individual impact evaluations considered in this report, an operational definition of TIP will be sufficient to measure program outputs, outcomes, and impact.

***Recommendation:*** *Identify the operational definition that was used for "trafficking in persons" in the program design; this definition should be used for the evaluation. Though one may not be able to define "trafficking in persons" for all countries and programs, this definition should clarify the phenomenon and overall problem that the intervention is seeking to effect. Once this definition is clear, maintain consistency in defining TIP this way throughout the evaluation cycle.*

*Comparing definitions of TIP across programs may be a more difficult issue. In this case, compare definitions and identify if components of each definition are comparable. For example, if both definitions of TIP include the prostitution of individuals, but only one definition includes forced labor and slavery, some comparison can be made across intervention to combat prostitution, but not forced labor.*

## F. Lack of Criteria to Identify Victims of Trafficking

Victims of trafficking vary according to the nature of trafficking (prostitution, slavery, or forced labor) and the factors that led them to be victimized (the characteristics of perpetrators, nature of victimization, and other factors contributing to victimization). Further, the identification of VoTs varies both within and across countries, and over time. As a result, there is a lack of well-developed criteria that identify VoTs. This challenge is closely related to the prior challenge of defining TIP. Often VoTs are portrayed as women or children who are involved in prostitution or forced labor, and these criteria can affect the types of programs that are designed to skew only to these populations (Rosenberg, 2008). This lack of inclusive criteria hinders the ability to identify victims, create consistent statistical databases, and design analytical tools for surveys and estimates.

Identifying criteria for VoTs requires understanding the trafficking system that engages and harbors these activities. Some considerations for developing criteria and understanding victim involvement will

require the program managers and evaluators to think about the trafficking system in different ways. These may include:

- *Trafficking as a business network* – Trafficking is a self-contained business network that lures or impels victims into, out of, and through it. Victims may not only be those that are physically trafficked, but those who are forced to fuel the system or harbor traffickers.
- *Trafficking as a criminal model* – This expands the business model to account for the intersection of trafficking with other criminal activity as a way for enforcing the continuation and success of the enterprise. Victims may be involved in drug dealing or other criminal activities in addition to trafficking.
- *Target purveyor networks* – Sometimes VoTs become traffickers themselves and function as part of the system. These individuals may not be fully aware of how they became enmeshed in the problem, and often see a way out of it by becoming a purveyor of trafficking. The distinction between trafficker and victim may be blurred.

In an evaluation of a specific anti-TIP program one should look to the program design to understand the criteria used to identify VoTs. Who is included in the program interventions? How are the program interventions conceptualized and authorized? These are not easy questions to answer, but they will help specify the evaluation questions needed to measure program impact.

***Recommendation:*** *As with the definition of TIP, maintain an operational definition of VoTs, and use this definition to specify the criteria for identification. Based on this operational definition, decide which beneficiaries are to be targeted for the evaluation, and maintain the VoT criteria on hand when developing the performance indicators of impact. In evaluation design, if the definition and criteria of VoTs are specified at the onset of the evaluation, and evaluation questions are designed with these criteria in mind, then the evaluation results should produce relevant outcomes.*

## G. Confidentiality and Protection of Identity

Service providers may be unwilling to share victim data due to confidentiality concerns. For example, the global database maintained by IOM is not publicly available since assisted victims are in a precarious position and revealing their identity could have a detrimental effect on their safety.

Evaluators should build the capacity of the local service providers to participate in the data collection process. These organizations may be the most relevant sources of information about the local patterns of trafficking and the types of populations that are vulnerable to being trafficked in the community. The providers may have distinct definitions of "trafficking" and who victims of trafficking are. As long as these definitions are consistent with the program and evaluation definitions, then the results should be relevant.

If the service provider has staff that are trained in data collection, let the provider track participant data over time, allowing for more privacy between the subject and the service provider. If the staff are not trained sufficiently, it may be useful to conduct relevant trainings at the beginning of the evaluation. Ideally, and cost-permitting, fully licensed social workers or others who are bound by confidentiality laws should conduct interviews. If this is not possible, once again consider training social workers to international standards and the importance of maintaining confidential records.

***Recommendation:*** *Build the capacity of local organizations and service providers who have close connections with VoTs and vulnerable populations to collect data. Maintain a record of local patterns of trafficking to build a representation of what populations need services and who is at risk.*

Caution is necessary however, as relying on local service providers can result in assumptions being made about vulnerability to trafficking – for example, assuming that trafficking affects only those who are similar in demographics to the victims assisted. There may be many victims who are not identified and assisted because they differ from this demographic, resulting in a vicious cycle of identifying only those victims who meet our expectations, and thus increasing our belief that these demographics signal vulnerability (Rosenberg, 2008).

***Recommendation:*** *Collect any resources available in the community on patterns of trafficking, legal procedures, government and organizational protocols for dealing with VoTs and vulnerable populations, etc. This third-party data will help triangulate data from program participants and service providers.*

Based on research of counter-trafficking programs, one outstanding challenge is that service providers and governments do not have reliable or consistent data about VoTs and vulnerable populations. This will be a challenge as long as governments do not commit themselves to fighting trafficking head on, protecting those that are victims and fully prosecuting those that run and fund the system. This last recommendation anticipates the challenges of collecting data and will be addressed more in-depth in the next sections. In brief, the recommendation here is to collect as much third-party data as possible to be able to triangulate information collected from program participants and service providers.

## H. Demonstrating Impact

Individual programs may address only part of the overall trafficking problem, creating a challenge for demonstrating impact. This is especially true for prevention programming, where it can be difficult to determine the impact of any one intervention on the overall prevalence of trafficking. For example, employment programs may be shown to be effective in helping individuals find employment, but linking such a program to the prevalence of trafficking is far more difficult.

Over time, individual program evaluations should not be the only evaluations a Mission undertakes related to TIP. Long-term evaluation planning should review the range of programs for a Mission and for a sector overall, in terms of the types of and status of programs being implemented, and how evaluations can build knowledge about them. Thus, individual, stand-alone, program evaluations become part of a larger evaluation plan for Mission-wide and sector-wide counter-trafficking initiatives.

A further recommendation for addressing impact on trafficking over time would be an agency-wide "Evaluation Initiative." This Initiative could address many technical and TIP-specific challenges of evaluation by developing a common strategy and data stream for benchmarking anti-TIP progress and impact, and by instituting standard and sound protocols for collecting and managing data at the agency level. It is not a specific recommendation of this report, but is included as a potential evaluation model in Appendix B.

***Recommendation:*** *Collect and disseminate lessons learned about effective practices and their relative and absolute impacts for different groups and different ways of being vulnerable to or emerging from trafficking. In the long term, consider Mission-wide or sector-wide evaluation plans to ascertain how different TIP prevention and VoT protection programs work together to affect the incidence of trafficking.*

## I. Measuring Vulnerability and Prevention Success

Measuring the impact of prevention programs based on vulnerability is extremely difficult because one cannot measure something that has not happened. Program implementers tend to design prevention programs to target communities whom they deem to be vulnerable to being trafficked. Often the factors which make up the categorization of "vulnerable to trafficking" are based on information

gathered from the population of assisted VoTs or from assumptions about trafficking (Rosenberg, 2008). However, this categorization may leave out vulnerable groups which are not as visible or that have not received assistance before, and are not represented by assisted VoTs.

If vulnerability is the indicator of program success, a promising alternative for measuring vulnerability is constructing a vulnerability index. Similar indices have been created to study economic and environmental phenomena; however, not many indices have been constructed to address social issues.[11] In constructing a vulnerability index, the evaluator chooses the indicators or components of the index, and then collects categorical data (a score) for each of the components. The score of each component is on a set scale, and the scores derive a composite index for vulnerability. It must be recognized that the variables chosen, and the scales set, are subjective to the choice of the evaluators. This, however, is a problem with any type of empirical work, and links back to the ethics of program evaluation. A detailed example of a vulnerability index is included in Section 4, Sample Plan for Evaluating a TIP Prevention Program.

***Recommendation:*** *Consider constructing a vulnerability index to study changes in behaviors of vulnerability of program participants over time. Such an index, to be operational, would need to be based on solid research looking at a wide range of trafficked individuals, and specific criteria or variables of vulnerability to trafficking. Similar indices have been developed for economic and environmental vulnerability, and the methods used could be extended to social phenomena, such as trafficking.*

## J. Measuring the Incidence of Trafficking

Effective measurement of the incidence of trafficking remains elusive. Victims are a hidden population and may be unaware, unwilling, or unable to acknowledge that they are trafficking victims. As well, victims from one country may be identified in another and that data may never be shared between them (Rosenberg, 2008). In addition to the difficulties in reaching VoTs to collect information using standard sampling techniques, one must add the ethical considerations involved in trying to do so.

Some countries may also have limited capacity for data collection or their governments' commitment to combating trafficking may be insufficient, adding to the difficulty in obtaining sufficiently reliable data needed for estimating trafficking incidence.

Individual evaluations of program impact should have data collection standards, and one should be to submit their data to the USAID Mission in a readable file, such as Excel or SPSS. Further, the Mission should encourage the evaluators to submit research data also, and any information collected on the incidence of trafficking (whether quantitative or qualitative). Over time, the Mission will be able to consolidate information from various counter-trafficking programs and this could be developed into a program database. Though there may be a lack of funding to maintain this database, the presence of a monitoring and evaluation system within the Mission to measure trafficking could greatly improve knowledge at all levels about the status of trafficking in the country and the region.

***Recommendation:*** *For evaluators, develop a database to help consolidate data collected for the impact evaluation (including background data on the incidence of trafficking), and submit this data to the Mission. For Mission staff, consider developing a program database of all data collected from various TIP prevention and*

[11] The United Nations has an Economic Vulnerability Index (EVI) to classify the development of its Member States. Economic and environmental vulnerability of small island developing states has also been studied by such scientists as Lino Briguglio and U. Kaly. For more information on measuring vulnerability, see Lino Briguglio's article: http://www.unep.org/OurPlanet/imgversn/103/17_mea.htm

*protection programs to understand the overall incidence and status of trafficking in the country. Measuring the prevalence of trafficking in a country could also be vastly improved by the development of standard population surveys.*

In recent years, some innovative ways of measuring the incidence of trafficking have been proposed and utilized. As discussed in Rosenberg 2008, the following ideas have shown promise:

1. International Database – While national databases have been in existence for some time, there are significant problems with knowledge sharing. As a result, huge discrepancies can be seen in the data. An example provided in Rosenberg 2008 involved data about Bulgarian VoTs trafficked to the Netherlands. For the years 2003 and 2004, statistics garnered from national organizations in Bulgaria reported a total of four victims trafficked to the Netherlands in 2003 and six victims in 2004. For those same two years, the Dutch National Rappateur on Trafficking reported 48 and 55 victims from Bulgaria respectively. An international database would help to consolidate the information. However, problems will still exist with regard to the reliability of the data, different definitions of trafficking, and duplication of data. Issues such as protection of victim identity will also have to be addressed.
2. Surveys – While an international database would improve collection of data on known incidences of trafficking, it will not allow us to extrapolate to measure levels of actual incidences of trafficking. The use of surveys in some countries has produced interesting results about incidences of trafficking (Rosenberg, 2008). Development of a standard survey tool would allow us to develop more information about the reliability and validity of the survey instruments. Standardization of the procedures in carrying out such a survey would help to address some of the ethical issues involved in undertaking such a survey, for example, by ensuring privacy and security protocols, and by linking victims identified in such surveys to assistance mechanisms in the country.

This report has discussed the evaluation framework and the strategies for making decisions in evaluation design. This section has specifically addresses many of the challenges in evaluating anti-TIP programs, and we have included recommendations to address these challenges. The next two sections are more illustrative and lay out sample plans for evaluating TIP prevention and victim protection programs. Both examples consider many of the challenges and recommendations mentioned here.

# 4. Sample Plan for Evaluating a TIP Prevention Program

USAID-funded programs sponsor many interventions to prevent trafficking among vulnerable populations. In this section we will build an evaluation framework for a model TIP prevention program, a typical program that USAID might fund. The evaluation framework will be based on the sample program described in Table 8.

**Table 8. Program: Economic Tools for Resistance and Resilience (ETRR)**

**Program Purpose:**
This program is designed to provide employment and income opportunities to individuals who are vulnerable to trafficking in order to prevent them from becoming victims.

Vulnerability to trafficking is assumed to stem from a combination of high debt or cost/income ratio, an interest in migration, and presence in the family of females aged 15 to 25.

The program intervention will take a two-pronged approach: the development of entrepreneurship programs and micro-credit for those interested in opening a business and employment programs for those interested in finding jobs. The program has funding to help 300 individuals in the first year.

**Program Design:**
The *entrepreneurship component* will include the following activities:

1) Development of entrepreneurship training programs: The implementing organization will work with local organizations with experience to tailor an entrepreneurship training course to the needs of the vulnerable population.
2) 100 vulnerable individuals will attend the training.
3) Development of business plans: Following the training, beneficiaries will develop business plans. The organizations providing the entrepreneurship training will work with the individuals to improve and polish their business plans.
4) Initial financing of business: Beneficiaries will be assisted to apply for micro-credit loans through existing micro-credit programs.
5) Additionally, all program participants will attend anti-TIP workshops where they learn about TIP, what it is, how victims are deceived, and how victims can get help if needed.

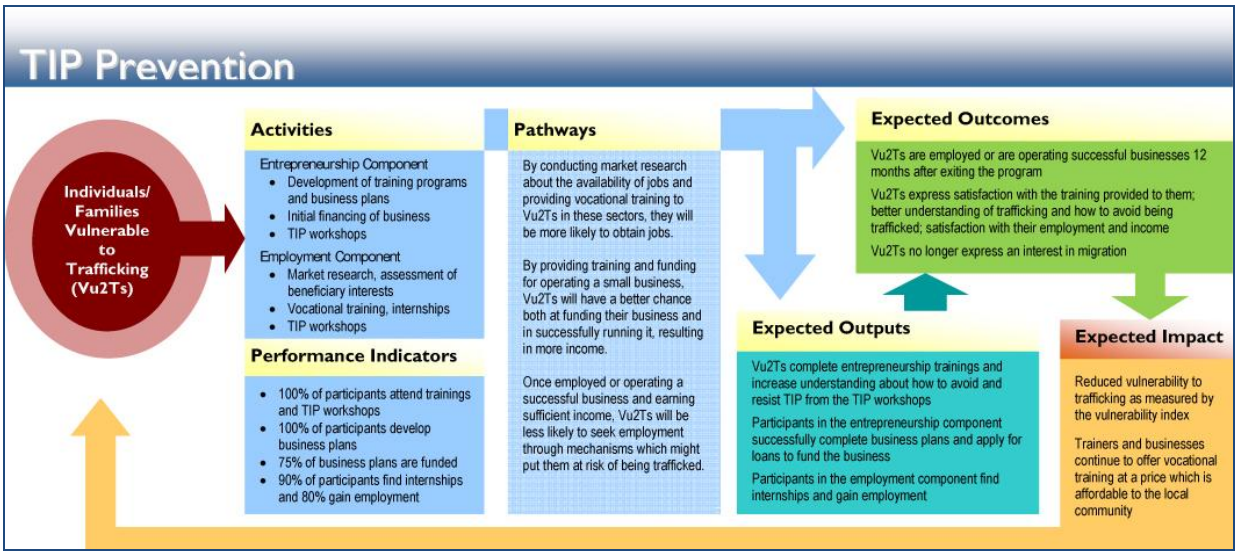
The *employment component* will include the following activities:

1) Market research: The implementing organization will conduct an assessment of the market in the target communities to identify areas where employment opportunities exist.
2) 200 vulnerable individuals will participate in the activity.
3) Assessment of beneficiary's interest: The implementing organization will work with beneficiaries to assess their employment interests.
4) Vocational training: 200 beneficiaries will be offered vocational training programs appropriate to their interests and in line with the market research. Use of existing vocational training programs will be prioritized. However, if there is a need for development of specialized training which does not currently exist, the implementing organization will work with local partners to develop such courses.
5) Internships: The implementing organization will identify employers willing to take beneficiaries as interns and provide them with on-the-job training. The training will be 3-6 months in length, depending on the field of employment and beneficiary experience. During the initial three months of the internship, the program will provide the intern with a stipend. Following the initial three months, employers will be expected to provide a stipend to the intern or offer them a regular employment contract.
6) Job placements: Following internships, additional job placement services will be offered to those beneficiaries who have not yet received regular employment contracts.
7) Additionally, all program participants will attend anti-TIP workshops where they learn about TIP, what it is, how victims are deceived, and how victims can get help if needed.

## A. Creating the Logic Model

Figure B depicts the logic model for the Economic Tools for Resistance and Resilience (ETRR) program. The model outlines the activities, results, outcomes, and impact of the intervention in preventing individuals from trafficking and leading to reduced trafficking overall.

**Figure B Depiction of the ETRR Program in a Logic Model**




The inputs are individuals vulnerable to trafficking (Vu2Ts); these are the targeted beneficiaries of the program. The Activities are the entrepreneurship and employment activities listed. Expected Results, Outcomes, and Impact are listed and labeled as such. Table 9 represents a detailed outline of the information in the graphic.

**Table 9. Components of the ETRR Logic Model**

| | |
|---|---|
| **Inputs** → | *Individuals assumed to be vulnerable to trafficking (Vu2Ts) as defined by a vulnerability index (see below)* |
| **Activities** → | *Entrepreneurship training, business plan development, micro-credit referrals, vocational training, internships, job placements, anti-TIP workshops* |
| **Performance Indicators** → | *100% of participants attend entrepreneurship trainings and anti-TIP workshops*<br>*100% of participants develop business plans, 75% of business plans are funded*<br>*90% of participants in the employment program find internships*<br>*80% of participants in employment program obtain employment* |
| **Pathways** → | *By conducting market research about the availability of jobs and providing vocational training to Vu2Ts, they will be more likely to obtain jobs*<br>*By providing training and funding for operating a business, Vu2Ts will have a chance at starting and successfully running their business, resulting in more income*<br>*Once employed or operating a successful business and earning sufficient income, Vu2Ts will be less likely to seek* |

| Table 9. Components of the ETRR Logic Model | | |
|---|---|---|
| | | *employment through mechanisms which might put them at risk of being trafficked* |
| **Expected Outputs** | → | *80% of Vu2Ts in the employment component complete vocational trainings and are placed in jobs*<br>*75% increase in Vu2Ts' understanding about how to avoid and resist TIP*<br>*75% of Vu2Ts in the entrepreneurship component successfully complete business plans and apply for loans to fund their business* |
| **Expected Outcomes** | → | *60% of Vu2Ts are employed and 50% operating successful businesses 12 months after exiting the program*<br>*75% of Vu2Ts express satisfaction with the training provided to them*<br>*70% have a better understanding of trafficking and how to avoid being trafficked*<br>*60% express satisfaction with their employment and income, 60% no longer express an interest in migration* |
| **Expected Impact** | → | *75% of Vu2Ts demonstrate reduced vulnerability to trafficking as measured by the vulnerability index*<br>*50% of trainers and businesses continue to offer vocational training at a price which is affordable to the local community* |
| **Assumptions** | → | *Vulnerable individuals are interested in the training and activities offered*<br>*The community can support new businesses and can employ those trained*<br>*Increased income will reduce vulnerability to trafficking* |
| **External factors** | → | *Micro-credit programs exist which will fund the business plans of program participants*<br>*The economy could change reducing employment opportunities and success of new businesses* |

### *1. Understanding the Timeline*

The evaluator should understand the potential timeline along which expected results, outcomes, and impact will happen. Part of this is recognizing that while results may happen at the end of the intervention, it may take some time for expected outcomes, and especially impact, to be realized. The outcomes should be captured when the results have had a chance to mature, which may take time, and the impact should be captured when the significance of the outcomes has had time to become clear. Table 10 predicts a potential timeline for the ETRR program.

| Table 10. Expected Timeline for ETRR Logic Model | | |
|---|---|---|
| **Activities** | Month 1 – 3 | *Market research conducted; Vocational and entrepreneurial modules developed* |
| | Month 4 – 6 | *First round vocational and entrepreneurial training taught to N individual participants* |
| | Month 6 – 12 | *Business plans developed; students placed in internships* |
| | Month 12 – 13 | *Businesses funded; job placements confirmed* |

| Table 10. Expected Timeline for ETRR Logic Model | | |
|---|---|---|
| **Expected Outputs** | Month 6 | *w% complete the trainings, meeting requirements, with different configurations of competencies* |
| | Month 13 | *x% have found jobs; y% have opened a business* |
| **Expected Outcomes** | Month 24 | *y% have been able to retain their job or are in a similar job for at least nine months to a year and are satisfied with their job and income* |
| | | *y% of businesses are still in operation and providing a satisfactory income* |
| **Expected Impact** | Month 24 – 30 | *z% have improved vulnerability scores on the vulnerability index* |

## B. Evaluation Questions

The overall question for the ETRR program seems obvious: *Did the program reduce vulnerability to trafficking?* Additionally, there are subsidiary questions that could be important for stakeholders in order to understand the impact leveraged and to focus additional programming. These additional questions are included in Table 11.

| Table 11. Specified Evaluation Questions for ETRR | |
|---|---|
| **Overall Question** | ***Did the ETRR program reduce vulnerability to trafficking?*** |
| **Expanded Overall Question** | *How did the ETRR program reduce vulnerability of Vu2Ts to trafficking?* *(For how many, to what extent, for how long…)* |
| **Program Importance** | *To what extent did beneficiaries require services of the ETRR program to prevent being trafficked?* |
| **Program Contribution** | *What are the strengths and weaknesses of different service elements or configurations; and how does this intervention compare with others like it?* |
| **Program Effects on Others** | *What are the effects of active participation in the program for other vulnerable individuals, members of participants' family, friends, and community?* |
| **Program Effects on Resources** | *How much did the program cost for the impact achieved? How does this compare to other interventions that might have been undertaken?* |
| **Appropriate analysis** | *To what extent are the initial program assumptions about what makes a person vulnerable to TIP correct?* |

Testing the impact of the prevention intervention has to go beyond the immediate program to incorporate knowledge of what has happened in other programs and what one would like to see happen. In this context the above questions are not only relevant, they may provide an important clue to how the intervention can be stronger, impacts broadened or made more sustainable.

## C. Measuring Vulnerability

Measuring vulnerability to trafficking is almost always based on untested assumptions. As an example, and for this particular sample program evaluation, the following factors may determine vulnerability to trafficking:

- Knowledge of what constitutes trafficking and how to avoid it;
- Self-confidence and life skills that assert control over one's life and reduce vulnerability;
- Job skills that can provide for economic sustenance (short-term) and sustainability (long-term);

- Knowledge of and access to resources for defeating or resisting advances of traffickers; and
- Strong family or community ties.

Due to the complexity of measuring vulnerability, for the ETRR program we have selected a *vulnerability index* as an indicator for success.[12] Such a vulnerability index would need to be constructed based on the main indicators of vulnerability. Each indicator is scored per beneficiary and combined into a composite index of vulnerability. In the vulnerability index, a high score indicates lower vulnerability (preferred), and a low score indicates the reverse (not preferred).

Once the appropriate measures of the above predictors have been decided upon, and seem from the literature and previous work to have validity, the statistical integrity of the index has to be assessed. If the index does not show statistical reliability, then it has to be modified. Development of such an index requires research and testing (see Section 3, Challenges and Recommendations, for more information).

## D. ETRR Evaluation Design

The evaluation design proposed for the ETRR program is a longitudinal analysis with two parts: (1) a longitudinal analysis of participants of the ETRR program only, and (2) a cohort longitudinal analysis where the participants are evaluated against a comparison group of individuals from the same vulnerable communities.

### *1. Longitudinal Analysis of the Treatment Only*

Given the vulnerability index discussed above, reduced vulnerability is determined by comparing scores on the index at intake into the program with scores at the conclusion of the intervention, creating an estimate of the extent and direction of change (higher or lower vulnerability). Table 12 outlines a detailed description of this first method of the evaluation design.

**Table 12. ETRR Evaluation Plan 1**

| |
|---|
| **Evaluation Question:**<br>Does the ETRR program reduce beneficiaries' vulnerability to becoming victims of trafficking? |
| **Evaluation Design:**<br>• Evaluation Method: Longitudinal analysis of the treatment only, which includes:<br>  ○ Treatment group: Beneficiaries of the ETRR program.<br>• Sample Selection: Census (100 percent) of beneficiaries. The unit of analysis is the individual.<br>• Sample Size: It is possible to get a census of the population, especially since the program universe (300) is known. |

For the longitudinal analysis of the treatment group only, it is important to get a stratified sample of beneficiaries, including those who:

- Registered as interested but did not enroll;
- Enrolled as a participant but did not attend trainings;

[12] A vulnerability index is a social science tool used to capture complex phenomena that may be hard to measure and may have underlying trends that are masked or cancel each other out. Kaly et al. (1998) constructed a composite index to measure environmental vulnerability using a categorical scale. There is a level of subjectivity in assigning scores to the index and weighting the variables included, and this should be discussed and operationalized in the evaluation methodology.

- Enrolled as a participant, attended some configuration of services, but did not complete any activities up to the standards;
- Enrolled as a participant, completed some services to required standards, but not all; and
- Enrolled as a participant and completed all the services to the required standards.

The reason for this is to be able to depict the diversity of beneficiary characteristics and service experiences to examine how the intervention worked, and how vulnerability may differ among the groups.

Since the universe population of the program is known to be 300, the sample size for the evaluation should be a census, meaning 100 percent of beneficiaries interviewed. As stated in the performance criteria, one of the indicators of the ETRR program is that "80 percent [of beneficiaries] of the employment program find jobs." This is a very high goal, and in fact is a measure of program effect. Since we know the entire beneficiary population, it is reasonable to assume that this large effect could indeed be true. However, this inference cannot be made unless a full census of the program population is taken. If this is not possible, the sample effect size should be reduced for caution. A key characteristic of this longitudinal analysis is sustainability of the changes found, so one would want to measure at least one year after the intervention (or some period during which it has been documented that a "relapse" often occurs).

Extent of change and whether the change could have happened by chance are both still of concern, however, as the treatment group is a special group. Would others, equally vulnerable in the same community but not participating in the program, have experienced the same kinds of transformations? For this we need the longitudinal cohort analysis.

### 2. *Longitudinal Cohort Analysis*

A comparison group is needed for the longitudinal cohort analysis. As stated in section 2, the most significant comparison is a randomly selected group from the same communities as the beneficiaries in the ETRR program. Baseline, intervention, and follow-up post intervention data are required, independent of program components, for both those who participated in the program and others who did not from the same community, using the vulnerability index. Details of the second method of the ETRR evaluation design is in Table 13.

**Table 13. ETRR Evaluation Plan 2**

**Evaluation Question:**

Does the ETRR program reduce vulnerability to trafficking?

**Evaluation Design:**

- Evaluation Method: Longitudinal cohort analysis of treatment and comparison groups which includes:
  - Treatment group: Beneficiaries of the ETRR program.
  - Comparison group: Individuals who did not participate in the ETRR program, but are from the same vulnerable communities and share similar characteristics.
- Sample Selection: Random sample of vulnerable individuals, both beneficiaries and comparison. The unit of analysis is the individual.
- Sample Size: Based on the cost of the evaluation, the ideal sample size will be based on the desired program effect to be studied. The best case scenario would be a census of the program group, with an equal comparison group. If this is not feasible, sampling should be left to sampling experts, since effect size of the program will need to be determined for the entire vulnerable communities.

As a reminder from section 2, the comparison group is *not* a control group, in that it is not identical to the treatment group. It is similar in certain characteristics, but is followed over time independently from the treatment group. With comprehensive baseline, mid- and post-intervention data, it is possible to analyze the difference within and between the two groups. For this analysis it is also extremely important to gather data about the social and economic characteristics of the treatment and comparison groups that may affect program outcomes. These will need to be included in the analysis of the vulnerability index. The next section details some of the necessary data for collection and analysis.

## E. Data Collection

The extent of data collection and analysis of the ETRR program depends on the size of the evaluation. The scenario presented here is of an extensive, two-pronged longitudinal analysis, which would require a considerable commitment of data. Though this may require more funds than the Mission has available to commit to evaluation, it is provided here to demonstrate a full-scale evaluation, from which different components can be selected as funds allow.

Both methods described in the previous sub-sections require baseline, mid- and post-intervention data. In addition, the evaluator should collect key information about both the treatment and comparison groups. A potential list of the data to be collected includes:

1. Treatment group data:
   - Vulnerability index: perception of vulnerability, tests of skills and knowledge gained, network resources utilized, participation in community counter-trafficking groups, extent of retention, and satisfaction ratings on program completion;
   - Internship and job placement and retention rate, business plan performance;
   - Income and expenses;
   - Demographic information; and
   - Social and economic background.

2. Comparison group data:
   - Vulnerability index: perception of vulnerability, tests of skills and knowledge gained, network resources utilized, participation in community counter-trafficking groups;
   - Employment;
   - Income and expenses;
   - Interest in migration;
   - Demographic information; and
   - Social and economic background.

3. Third party data (this is used to triangulate data from the treatment and comparison groups):
   - Known incidences of trafficking from community (before, during and after program);
   - Presence and involvement of counter-trafficking groups in the community; and
   - Social and demographic characteristics of the community.

This is but a sample of the data necessary for a comprehensive evaluation. Further data could also be gathered about the businesses and counter-trafficking networks in the community, and how participation in the ETRR program made a difference on their operations. Depending on what is known about trafficking in the community, and the level of investment in the ETRR program, it may be worthwhile to interview a small subset of potential "anti-trafficking network" members to understand more deeply what it takes for a prevention program such as ETRR to impact community support and active engagement against trafficking.

# 5. Plan for Evaluating a VoT Protection Program

A sample evaluation plan developed to assess a program focused on direct assistance to victims is presented in Table 14. This evaluation plan is developed based on the evaluation framework described in Sections 1 and 2.

**Table 14. Program: Wide-Range Assistance to Victims of Trafficking (WRAVoT)**

**Program Purpose:**
This program is designed to provide a wide range of assistance to victims of human trafficking. The program will include shelter and non-shelter based services and a hotline to aid in the identification of victims of trafficking.

The intervention population for this program is extremely hard to capture. The WRAVoT program is being implemented in a medium-size country (about twenty million people) with a significant low-income population. To date, approximately 200 victims of trafficking per year have been identified. The potential impact of the program could reach several hundred people.

**Program Design:**

The program includes two main components, assistance services and a hotline to increase reporting of trafficking cases.

The *assistance services* component will include the following activities:

1) Identification, selection, and funding of NGOs and other local providers to provide services to VoTs. Such services include counseling, medical and psychiatric care as needed, legal aid, accommodation, continuing education, vocational training, job placements, and family mediation.
2) Training for government-run and private service providers on the special needs of trafficked persons, including psycho-social counseling, shelter management, and income generation programming.
3) Workshop to alert stakeholders in a position to identify VoTs about the existence of services.
4) Referral of VoTs to government-run services, if available, such as those provided by employment bureaus and social welfare centers; or private services.
5) Funding for the provision of shelters for victims of trafficking in two locations in the country to be run by local NGOs.

The *hotline component* will include the following activities:

1) Training for hotline operators and development of a hotline operators' manual.
2) Funding for operation of the hotline.
3) Media campaign to alert the population to the existence and purpose of the hotline.

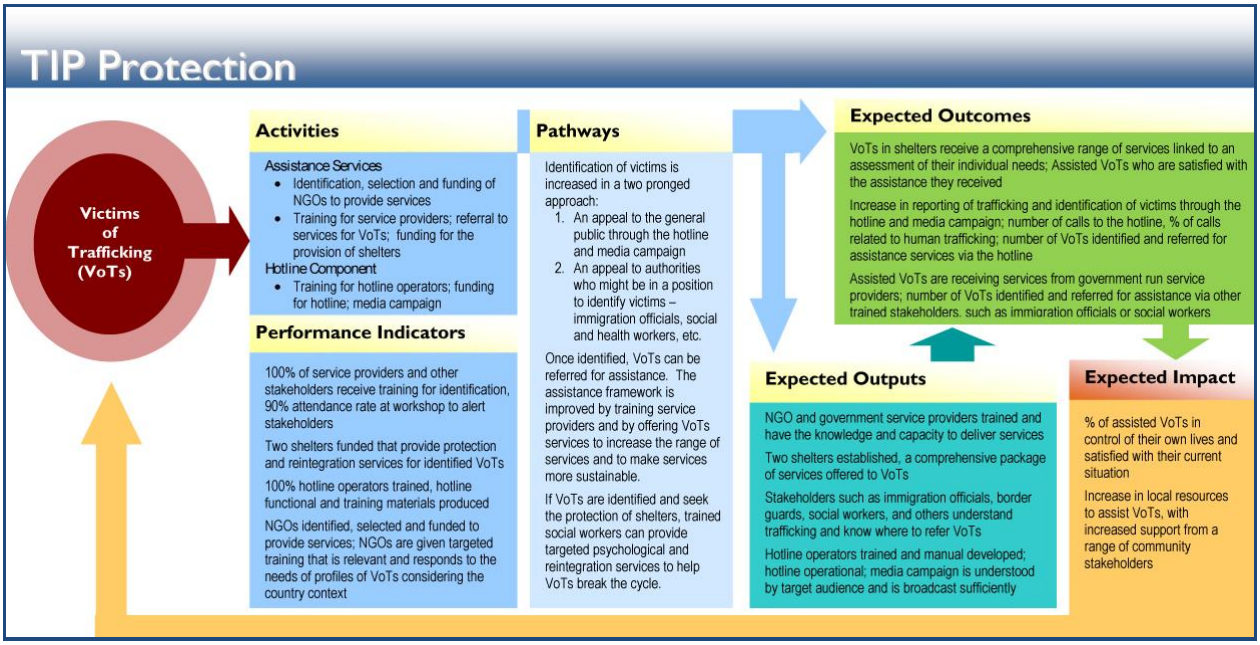
## A. Creating the Logic Model

Figure C depicts the logical reasoning to measure how the WRAVoT program will protect victims of trafficking and increase reporting of trafficking cases and identification of victims. While with prevention programs the inputs focused on individuals assumed to be vulnerable to being trafficked, here the inputs are exclusively placed on the individuals who have already been victimized as well as those in a position to identify or assist them.

The main issue highlighted throughout the logic model is the way in which resources for protecting victims will continue beyond shelter and hotline assistance, to protecting victims in their home communities or elsewhere, thus enabling assistance to continue beyond the program by using locally available resources. It is indicative of the activities of the WRAVoT program that efforts will be made to raise the awareness of local stakeholders and enforcement networks of the availability of services for

victims of trafficking, increasing the protection of VoTs through common recognition and responsiveness to trafficking. Also the program includes continued outreach to community leaders and authorities to reinforce messaging about service availability, recognition, and protection of VoTs.

**Figure C Depiction of the WRAVoT Program in a Logic Model**




The measures of expected results, outcomes and impact are detailed in Table 15. As with prevention programs, it is very difficult to specify measures that are valid, reliable and attainable with the resources available. For this reason, it is extremely important to keep the evaluation questions specific and to understand and state clearly the challenges and shortcomings of the evaluation methodology. A big issue for protection programs, more so than for prevention, may be how long program evaluations have to wait to ensure that the results of the interventions have had time to materialize. It will be important to understand also what it takes to make stakeholders rely on the services and want to invest further in them. So both the messaging strategy and response to the media campaign about the services will be important to review in terms of differential contribution to effectiveness of the intervention strategies.

**Table 15. Components of the WRAVoT Logic Model**

| | |
|---|---|
| **Inputs** | *Victims of trafficking (VoTs) and service providers* |
| **Activities** | *Training of NGO and government service providers, assistance and shelter services to VoTs, training of hotline operators, development of hotline manual, operation of the hotline* |
| **Performance Criteria** | *100% of private and government-run service providers receive trainings on how to provide assistance to VoTs*<br>*Two shelters established to deliver assistance services*<br>*90% attendance rate at workshop to alert stakeholders (such as immigration officials, border guards, police, social workers, hospital and clinic staff, teachers and NGOs) in a position to identify VoTs*<br>*100% of hotline operators receive training, hotline manual* |

| Table 15. Components of the WRAVoT Logic Model | | |
|---|---|---|
| | | *developed* |
| **Pathways** | → | *Identification of victims is increased in a two pronged approach – 1) an appeal to the general public through the hotline and media campaign associated with it; and 2) an appeal to authorities who might be in a position to identify victims*<br>*The assistance framework is improved by training service providers to increase the range of services available to VoTs and to make services more sustainable by integrating government-funded services* |
| **Expected Outputs** | → | *75% of service providers trained and workshop attendants understand trafficking and know where to refer victims for assistance*<br>*2 hotlines are operational 24 hours per day, seven days a week for at least two years, 90% of hotline operators refer to trainings and manual when assisting VoTs, 70% of calls relate to human trafficking*<br>*Media campaign is understood by target audience and is broadcast sufficiently on appropriate channels to be widely received* |
| **Expected Outcomes** | → | *60% of assisted VoTs are satisfied with the assistance they receive*<br>*70% increase in reporting of trafficking and identification of victims through the hotline and media campaign*<br>*75% of VoTs identified and referred for assistance services via the hotline* |
| **Expected Impact** | → | *50% of assisted VoTs in control of their own lives and satisfied with their current situation*<br>*Increase in local resources to assist VoTs, with increased support from a range of community stakeholders* |
| **Assumptions** | → | *Victims need and desire the offered assistance, victims are being identified and referred for assistance*<br>*The community has sufficient resources to facilitate victim reintegration*<br>*The economy is strong enough for victims to be able to find jobs that provide a living wage* |
| **External Factors** | → | *Victims of trafficking are eligible for government run services* |

## B. Evaluation Questions

The overall evaluation question for the WRAVoT program is: *Did the WRAVoT program increase identification of victims of trafficking and establish a system of support to ensure their economic, physical and psycho-social well being?*

It is extremely important to specify the results, outcomes and impact of the program in operational ways, meaning in ways that can be realistically measured. It is an overstatement to imply that the WRAVoT program will lead to a reduction in trafficking overall, at least as a direct impact. Therefore, though the overall question may not state that trafficking will be decreased, it is reasonable to imply that the WRAVoT program will increase protection of VoTs that use the program, whether through direct assistance or an increase in resources available to VoTs and local stakeholders.

Table 16 includes the specified evaluation questions for the WRAVoT program. Additional data may need to be collected to look for the differential impacts of services for different kinds of victims, and potential leveraged resources, such as data from service providers related to documented use and outcomes from individual services, and the configuration of services used.

**Table 16. Specified Evaluation Questions for WRAVoT**

| **Overall Question** | ***Did the WRAVoT program increase protection for VoTs from trafficking?*** |
|---|---|
| **Expanded Overall Question** | *How did the WRAVoT program increase identification of and protection for VoTs? (For how many, to what extent, for how long…)* |
| **Program Importance** | *To what extent did VoTs require services of the WRAVoT program to get help in breaking out of trafficking and reintegrating into independent self-directed lives? What services were most needed?* |
| **Program Contribution** | *What strengths and weaknesses did the WRAVoT program have compared to other programs of the same type, for victims referred by different sources?* |
| **Program Effects on Others** | *What are the effects on participation in countering TIP of other actors such as service providers, community members, families, and social networks around victims?* |
| **Program Effects on Resources** | *To what extent and how did the interventions affect the resources available for identification and service provision to victims of trafficking?* |
| **Appropriate analysis** | *To what extent are the initial program assumptions about the demographics of victims of trafficking and the ways of reaching out to increase identification of trafficked persons and the service needs of VoTs correct?* |

Testing the impact of the WRAVoT program has to go beyond the immediate activities to incorporate knowledge of what has happened in other programs and what one would like to see happen. In this context, the above questions are not only relevant, they may provide an important clue to how the intervention can be stronger, impacts broadened or made more sustainable.

## C. Measuring Protection

Protection and assistance that is effective for one individual in one context does not work necessarily for all victims of trafficking. One challenge of the measurement of effective protection is to be able to map the range of predatory practices and demographics of potential victims and how protection and assistance could occur, both individually and systemically.

Prevention and vulnerability in the ETRR program was measured with a composite index of variables, the vulnerability index. A similar index could also be developed for this program – one which would attempt to measure indicators of successful reintegration. However, in this example, in order to present a different and a less costly approach, the evaluator will use a mixed-methods approach, in which three quantitative measures of program success will be supported by qualitative indicators of program performance. Once again, these indicators are but a few of many that can be developed.

### *1. VoTs Identified and Receiving Services*

The first measure of the WRAVoT program is the quantitative change over time of VoTs that contact the WRAVoT hotline and that use the assistance services provided. A numeric assessment could be made of the change in activity of the following two types of victims:

- VoTs who were identified via the hotline and successfully referred to assistance services; and

- VoTs who were identified via the hotline and used assistance services successfully.

Here the measure of "success" is subject to the standards prescribed by the evaluator, and this should be noted in the evaluation methodology (as noted above, an index of successful reintegration would be a useful tool here). These indicators will not necessarily require a comparison group, though the evaluation could be strengthened if one could find VoTs who were referred to assistance services, but chose not to use them.

This rough measure of an increase or decrease in victim identification and assistance should be supported by qualitative descriptions of the types of assistance given. Some qualitative questions could include:

- How does the hotline system help VoTs? How are hotline operators updated and informed about the changing needs of VoTs or changes in available services?
- Do the shelters, hotline, stakeholder training and re-integration program have features that build sustainability for resources beyond the life of the program? Has the integration of locally available government funded services improved sustainability of services?

To ensure that the services resulted in lasting success for the victims assisted, one would want to measure impact at least one or two years later. Understanding the patterns of "re-trafficking" is an important factor here. Similarly, defining and measuring success is critical, as the victims being assisted may have a different definition of success than those assisting them. For example, victims may wish to migrate again, but this time successfully securing a satisfactory job, while service providers may measure success as the individual remaining in the country and undertaking employment in a locally available job (Rosenberg, 2008).

#### *2. Resources for VoTs Strengthened*

In order to evaluate VoT protection activities, one must be able to infer that the services offered enabled VoTs to become more economically and socially independent and able to exert control over their wellbeing. An indicative measure of this is to show that the service providers increased the resources available to VoTs, and that VoTs have continued access to and ability to make use of these services.

For this second predictor, statistical significance would be assessed by a concrete standard envisioned by the statistical experts for a quantitative threshold whereby service providers are assessed on their ability to successfully provide assistance services to build life and job skills.

Qualitative questions that could further inform the use of assistance services include:

- How do the assistance services build in understanding of the kinds of skills VoTs need to reestablish normal life and to avoid the advances of traffickers?
- What kinds of services or service packages had lasting effects for VoTs, once they returned home (or established a new one), and what configurations of services or support worked best and fastest?

#### *3. Resources in Community Strengthened*

Another measure of the WRAVoT program could be a significant change in community resources accessed for understanding trafficking and identifying/helping victims of trafficking. This analysis would require a comparison community where the WRAVoT program has not been implemented.

Effects of media campaigns and trainings to inform the public and service providers about identification and services for VoTs are extremely hard to measure, particularly because it is unknown what the universe audience could be and how many people could potentially hear the announcements. Therefore, the most that the evaluator can measure is the change in the rate of referral to hotlines and assistance websites based on the media campaign and trainings (from quantitative surveys of a sample of the population).

Several qualitative questions that could support this indicator could be:

- What services were the most effective for the community and the victim's overall support system? Were there effects on the community response and support to victims?
- How does the program work with communities, social networks, and authorities referring victims about how they could be more efficient and provide more of an integrated network, and engage the victims themselves in being part of the network?

## D. WRAVoT Evaluation Design

The evaluation design proposed for the WRAVoT program uses a mixed-methods approach. The methods used include:

1. Longitudinal analysis of the treatment only for VoTs who called the hotline; and
2. Cross-sectional analysis of the treatment group community with a comparison group community to assess impact of the media campaign and trainings.

### *1. Longitudinal Analysis of the Treatment Group Only*

A longitudinal analysis of the treatment group will give the evaluator an understanding of the changes among VoTs who participated in the WRAVoT program. As already stated, in protection programs it is extremely difficult to get a comparison group that has not been involved with the program *at all*. Unless the country government has a detailed tracking system for all VoTs, comparing assistance services between those that used services and those that potentially could have used them is virtually impossible. Comparison groups, therefore, cannot be applied here. [Note: If it is possible to contact known victims who declined assistance, this method could be complimented by a cross-sectional analysis of this group with those who declined assistance.] Table 17 details the first evaluation plan.

**Table 17. WRAVoT Evaluation Plan 1**

**Evaluation Question:**

*Does the WRAVoT program increase identification of and services for VoTs?*

**Evaluation Design:**

- Evaluation Method: Longitudinal analysis of the treatment only, which includes:
  - Treatment group: Beneficiaries of the program (VoTs and services providers).
- Sample Selection: Random sample of VoT beneficiaries. The unit of analysis is the individual.
- Sample Size: It will not be possible to get a census of the population, so a random sample should be chosen based on specific characteristics of the treatment group. Attrition will be a serious problem and needs to be addressed when tracking victims over time.[13]

[13] Attrition is a problem that affects most longitudinal studies. Participants of a program drop out over time for various reasons, including relocation or unwillingness to continue with the program. Since statistical tests are conducted based on the baseline sample, this presents a problem. Statistical techniques can be used to overcome

For the longitudinal analysis of the treatment group only, it is important to get a representative sample of beneficiaries. The reason for this is to be able to depict the diversity of beneficiary characteristics and service experiences. With the assumption that the program would have a tracking system from hotline to service referral to provision of services, it is possible to track victims over time. However, as is common with protection interventions, attrition will be a serious problem. VoTs may not wish to participate in the evaluation, may be hard to find after reintegration, or may leave the program on their own accord for several reasons: they are afraid of being found, they find another job, they are back with their families and don't feel they need more assistance, and many others.

For evaluation purposes, issues of attrition will be unavoidable, and should be mentioned clearly in the evaluation methodology. The best recommendation to confront this issue is to keep a detailed database of victims in the program, with bi-annual or annual check-ins to the victim's current location and occupation. In some instances, evaluation techniques may be used to replicate individuals who have dropped out of the sample, but this should be left to technical experts.

Finally, one of the pluses of longitudinal analysis is to be able to track and measure change over time. Sustainability is an important component of all protection programs. If VoTs can positively use assistance services to break out of the cycle and move forward to a life of non-trafficking with support from local community and family networks, this can be a best practice for any program. Due, however, to high dropout and attrition rates, measuring sustainability and impact over time will be a particular challenge. If funding permits, ideally former victims could be tracked for years, presenting the potential to successfully show the impact of the program. Realistically, programs probably will not have funding to show impact longer than one or two years after the life of the program. Either way, sustainability and impact should be considered when planning data collection.

### 2. *Cross-Sectional Data Analysis*

The evaluation design chosen for the WRAVoT evaluation is a mixed-methods approach for two reasons: to show an example of a cross-sectional data analysis and to give an example of a cost-strained evaluation. Cross-sectional data analysis can be used in combination with longitudinal analysis to find possible changes in the community without the ability to study these changes over time. In this sample, it is understood that the WRAVoT program does not have the funds to track individuals in a comparison community over the program cycle, nor to have baseline data from either communities. The evaluators must understand the challenges of employing this method, but it should not be understated that this approach is often used and not completely uncommon.

Table 18 outlines the second method used in the WRAVoT evaluation. The community where the program is being implemented is very large, and so effect size should be relatively small. Also, it may be difficult to construct a random comparison sample in a similar community (unless there is a nearby city of the same size and characteristics). With these constraints in mind, a random treatment sample is constructed of various individuals in the community. These can include community members, stakeholders, enforcement and government personnel.

**Table 18. WRAVoT Evaluation Plan 2**

| **Evaluation Question:** |
| --- |
| Does the WRAVoT program increase resources for the community to understand trafficking and assist VoTs? |

issues of attrition, but one needs to be cautious when attributing common characteristics to new subjects added to the sample.

**Table 18. WRAVoT Evaluation Plan 2**

**Evaluation Design:**

- Evaluation Method: Cross-sectional data analysis which includes:
  - Treatment group: Sample of individuals in the WRAVoT program community.
  - Comparison group: Sample of individuals in a comparison community where the WRAVoT program was not implemented.
- Sample Selection: Random sample of the community, treatment and comparison. The unit of analysis is the individual.
- Sample Size: Gathering a random sample for this evaluation method is difficult. Due to the unknown size of the community (we can consider this to be the whole city), it is doubtful that the program would be able to have a large effect on the entire population. The sample will need to be constructed based on knowledge of the desired impact of the media campaign and hotline.

A comparison group is constructed to share similar characteristics with the treatment group, randomly choosing individuals from a similar community. Throughout this portion of the analysis, one must be overly cautious about making overarching inferences or conclusions. Though the evaluator can control for known differences between the two communities, there may still be characteristics that underlie the outcomes of the analysis, and this may lead to false positive or false negative conclusions.

Considering all of the challenges, should this method be used? For various reasons, the answer points to "yes". If the evaluators can collect data about programs implemented in both communities, legal frameworks for trafficking policies, and demographic and economic statistics about the communities, perhaps this analysis will be able to show some of the differences as a result of the intervention. Finally, the bottom line may be that program funding allows for this method, if nothing else. As has been mentioned several times in this report, evaluators often are constrained by the funding allotted to evaluation design, and this sample presents a possibility when such constraints are the reality.

## E. Data Collection

For the longitudinal analysis, one will require baseline, mid- and post-intervention data from VoTs and service providers. The evaluator should collect key information about the treatment group, including:

- Use of program services (hotline, shelter, job assistance) and satisfaction with services;
- Measure of knowledge about protection services, network resources utilized,
- Measure of personal and job skills, change in economic status;
- Rate of referral (hotline to service) and rate of attrition among service providers;
- Rate of referrals from professionals trained by the program;
- Demographic and service provider information; and
- Social and economic background.

For the cross-sectional data analysis, treatment and comparison data are collected from two communities. Since this collection is on a one-time basis, the evaluators must be specific about the information they need to obtain. Third party data is very important here as well. The underlying differences between the communities can be controlled if they are apparent to the evaluators. The following key information should be gathered:

1. Treatment group data:
   - Knowledge about trafficking;
   - Awareness of the dangers of trafficking and tips to help VoTs;
   - Knowledge about identifying and referring VoTs to appropriate resources;

- If resources are available, use of any resources and satisfaction;
- Demographic information; and
- Social and economic background.

2. Comparison group data:
   - Knowledge about trafficking;
   - Awareness of the dangers of trafficking and tips to help VoTs;
   - Knowledge about identifying and referring VoTs to appropriate resources;
   - If resources are available, use of any resources and satisfaction;
   - Demographic information; and
   - Social and economic background.

3. Third party data:
   - Incidence of trafficking from community;
   - Presence and involvement of counter-trafficking groups in the community;
   - Legal frameworks, if any, for government prevention and protection initiatives;
   - Perceptions of community about trafficking; and
   - Social and demographic characteristics of the community.

Third party data may also be critical in obtaining information about the two communities at the time the WRAVoT program began. For example, incidence rates of trafficking may be available from both communities before and after the program intervention. The presence of various prevention and protection programs in each community may also be documented, as well as legal frameworks for assisting victims of trafficking. The argument can still be made that economic, political, and social effects on the treatment and comparison communities during the time of the program affected the outcomes of the study; however, as long as the evaluators are upfront about these limitations and do not make generalizations about the community as a whole, the method can be used to compare differences between the two sample groups.

# References


Asher, H. B. (1976). *Causal modeling.* Beverly Hills: Sage Publications.

Bamberger, M. J., Rugh, J., & Mabry, L. (2006). *Realworld evaluation: Working under budget, time, data, and political constraints.* Beverly Hills: Sage Publications.

Blalock, H. M. (1964). *Casual inferences in nonexperimental research.* Chapel Hill: University of North Carolina Press.

Briguglio, L. (1995). Small Island States and their Economic Vulnerabilities, *World Development, 23*, 1615-1632.

Briguglio, L. *Measuring Vulnerability.* Retrieved from http://www.unep.org/OurPlanet/imgversn/103/17_mea2.htm

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Newbury Park: Sage Publications.

Clawson, H. J., Williamson, E., & Garrett, A. (2008). *Improving data to combat human trafficking.* ICF International. Retrieved from http://thehill.com/resources/white-papers/human-trafficking/48567-improving-data-to-combat-human-trafficking/.

Cohen, B.H., & Lea, R.B. (2004). *Essential statistics for the social and behavioral sciences.* New Jersey: John Wiley and Sons.

Cooley, W. W., & Lohnes, P. R. (1978). *Evaluation research in education*. New York: Irvington Publishers.

Crawford, M., & Kaufman, M. R. (2008). Sex trafficking in Nepal: Survivor characteristics and long-term outcomes. *Violence against Women, 14:8,* 903–916.

Donaldson, S.I., & Christie, C.A. (2005). The 2004 Claremont Debate: Lipsey versus Scriven. Determining causality in program evaluation and applied research: Should experimental evidence be the gold standard? *Journal of Multidisciplinary Evaluation, 3*, 60-77

Dottridge, M. (2007a). *A handbook on planning projects to prevent child trafficking.* Lausanne: Terre des Hommes Foundation.

Dottridge, M. (2007b). *Measuring responses to trafficking in human beings in the European Union: an assessment manual*. Brussels: European Union.

Government Accountability Office. (2005). *Performance measurement and evaluation: definitions and relationships.*

Government Accountability Office. (2007). *Report to Congressional requesters: Human trafficking. Monitoring and evaluation of international projects are limited, but experts suggest improvements.*

Henkel, R. E. (1976). *Tests of significance.* Beverly Hills: Sage Publications.

Human Security Centre. (2005). *Human security report 2005: War and peace in the 21st century*. New York: Oxford University Press.

International Labour Organization. (2007). *Decent work indicators: Towards better statistics on forced labour and human trafficking.*

International Organization on Migration. (2008). *Handbook on performance indicators for counter-trafficking projects*. Washington, D.C.: International Organization for Migration.

Kaly, U., Briguglio, L., McLeod, H., Schmall, A., Pratt C. and Pal, R. (1998). *Environmental Vulnerability Index (EVI) to summarize national environmental vulnerability profiles*. Fiji: SOPAC.

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on impact evaluation: quantitative methods and practices.* Washington, DC: World Bank.

Kish, L. (2004). *Statistical design for research*. New Jersey: John Wiley and Sons.

Knoke, D., & Kuklinski, J. H. (1982). *Network analysis*. Beverly Hills: Sage Publications.

Kumar, R. (1999). *Research methodology: A step-by-step guide for beginners*. (2nd ed.). London: Sage Publications.

Kusek, J. Z., & Rist, R. C. (2004). *Ten steps to a results-based monitoring and evaluation system.* Washington, DC: World Bank.

Kutnick, B., Belser, P., & Danailova-Trainor, G. (2007). *Methodologies for global and national estimation of human trafficking victims: Current and future approaches.* Geneva: International Labour Office.

Laczko, F., & Gramegna, M. A. (2003). Developing better indicators for human trafficking. *Journal of World Affairs, X:1,* 179-194.

Langbein, L. I., & Lichtman, A. J. (1978). *Ecological inference*. Newbury Park: Sage Publications.

Louviere, J. J. (1988). *Analyzing decision making; Metric conjoint analysis.* Beverly Hills: Sage Publications.

Organization for Security and Co-operation in Europe. (2005). Proceedings from *Combating Trafficking in Children*. Vienna, Austria: OSCE.

Rosenberg, R. (2008). *Best practices for programming to protect and assist victims of trafficking in Europe and Eurasia*. Washington, D.C.: USAID/E&E/DGST.

Sanders, J.R. (Chair) (1994). The joint committee on standards for educational evaluation. *The program evaluation standards* (2nd ed.). Beverly Hills: Sage Publications.

U.S. Agency for International Development. (2003). *Trafficking in persons: The USAID strategy for response.* Washington, D.C.: USAID.

United Nations Statistics Division. (2006). *The world's women 2005*. New York: United Nations.

Warnath, S. (2009). *Best practices in trafficking prevention in Europe and Eurasia*. Washington, D.C.: USAID/E&E/DGST.

Weiss, C. H. (1998). *Evaluation*. (2nd ed.). New Jersey: Prentice Hall.

# Appendix A

# Evaluation Design in Depth

| Evaluation Issues | | |
|---|---|---|
| **Component** | **Definition** | **Example** |
| Identifying Respondents: *Where would information most knowledgeably and reliably come from to address evaluation issues?* | Participant | Self-report – perspectives, attitudes, knowledge, behavior pattern |
| | Non-participant | Self-report – perspectives, attitudes, knowledge, behavior pattern evolution |
| | Managers | Report: perspectives, attitudes, barriers, perceived changes |
| | Document | 3rd party information on use of services or resources |
| Sample relative to a standard for determination of impact *What kind of data sample is required for the evaluation purposes?* | Project | Documentation of services, attitudes, and/or skills gained |
| | Non-random, matched Comparison | Assessment of project effect in comparison with similar TIP-relevant individuals, but not experiencing the project intervention – very difficult to document "match" |
| | Longitudinal or cohort analysis | Assessment of project effect through follow-up with the same clients, or a cohort, sometimes beyond the end of the project, regarding response to project intervention, its outcomes, & sustainability of results |
| | External Standard | Assessment of project effect through comparison of outcomes with a standard or statistics, accepted in the field in the region – e.g., change in length of victim shelter stay, over time. |
| Sample *How should the respondents be selected to meet the requirements of the evaluation?* | Purposive | Measurement derived from respondents not representative of the TIP population – e.g., they may represent a specific or non-specific sub-group of the population, for example service users. |
| | Random | Measurement derived from respondents selected to reflect the range of respondent characteristics in the population. This sample strategy can produce findings that may be able to extend beyond the project, if the measures are reliable and valid. |
| | Stratified Random | Measurement derived from respondents selected to reflect the population, with attention to representation of specific characteristics that might not be captured by chance and are important for understanding the impact on TIP. |
| Size of the sample *How many people have to provide data to accurately represent the findings?* | Adequacy | Size should be adequate to represent the diversity of the TIP population your intervention addresses – e.g., the more diverse the population, the larger the sample will need to be in order to have the power to identify stable significant findings |
| | | Size should be adequate to accommodate the complexity of the evaluation questions and the underlying logic model and measures – e.g., the more granular the questions and complex the model and measures, the larger the sample will need to be. |
| Data Collection methods: *How should data be collected to provide the level of detail needed, in the most reliable way?* | Check lists | Independently conducted inventory of resources or deployed skills |
| | Inventory | Samples of things - laws, materials offered, service records , receipts |
| | Focus Group | Generated discussion, following an agenda, among small groups of targeted informants relevant to the beneficiary community(ies) |
| | Observations or 3rd party assessments | Direct viewing of application of skills or behavior in context, over a set period of time, using either a structured format or an agenda. |
| | Survey/ Interview | Independently or self-administered questionnaire designed to gather opinions, attitudes, or informant's self-report of behavior, |

| **Evaluation Issues** | | |
|---|---|---|
| **Component** | **Definition** | **Example** |
| | Tests | Extent of mastery of targeted skills or information either face-to-face or via telephone or computer, independently or self-administered |
| Frequency of data collection<br>*How often should data be collected to provide sound findings related to the evaluation question?* | 2x –<br>Baseline & post | Change in prevalence of attitudes or acquisition of knowledge and skills from the start of the program, e.g., intake and post |
| | 1x –<br>Post only | Data collected once, at completion of activity - e.g., victim shelter resident records |
| | Multiple times | Data collected multiple times, e.g., time series, surveillance, changes in economic status |
| | Infrequently | Data collected infrequently, e.g., Every x years, represent change in incidence |
| Analysis<br>*How shall I report the findings to be clearest about the level of impact?* | Description | Findings– e.g., documentation of project targeted beneficiaries and comparison sample characteristics, frequency participated in project activities, percent of targeted beneficiaries who met intended benchmarks, percent of intended beneficiaries that were re-trafficked or who avoided re-trafficking. |
| | Patterns of Association | Findings– e.g., differences in findings among sub-groups of targeted beneficiaries, or by demographic factors like age or marital status, or by geographic variables; variables that seem to modulate the findings, e.g., previous experience with trafficking, length of time in community, intensity with which experienced intervention. |
| | Estimates of causation | Findings- e.g., strength and direction of relation between intervention and measures of outcomes related to re-trafficking, for different subgroups within the population; |
| | Certainty | Probability level – e.g., rigorousness of statistical tests |
| | Error rates | Extent to which take into account design and other error rates |

# Appendix B

## Net Impact Assessment – an Evaluation Initiative

A net impact assessment is more than the sum of the individual evaluations. This initiative refocuses attention on the goals of trafficking overall, what is known about what has been achieved, the progress made, and the holes in the fabric of progress – what would have been achieved without or with different funding. The net impact assessment adds some requirements for evaluation design and coordination; however, it may be little different in cost from individual assessments, and may actually save money based on easier access to respondents and more controlled evaluation focus. A proposal for managing this type of assessment is presented as part of the specific evaluation model recommended here: a USAID Anti-TIP Evaluation Initiative.

An anti-TIP Evaluation Initiative would develop a common evaluation strategy and data stream for benchmarking progress and impact, and institute standard and sound protocols for collecting and managing data. This Initiative is directly responsive to GAO (2007) and USAID (2009) concerns (as part of the commission for this report) inquiring about:

1. How to make evaluation cost-effective, with a model that can encompass different project designs and variation in the nature and definitions of trafficking across borders, and can assure the integrity and qualify of the data;
2. Requirements for selecting appropriate evaluation method(s) rigorous enough to support decision making and generalizing findings to national level initiatives about the value of anti-TIP projects, including whether a control group, or even the more stringent randomized control group trial (RCT) is necessary in these contexts; and
3. Requirements for managing and implementing evaluation operations.

The next section presents how the anti-TIP Evaluation Initiative could address these issues. From there we present a potential framework for assessing impact prevention and protection programs, within the anti-TIP Evaluation Initiative.

### Thinking Beyond the Immediate Project Evaluation

A USAID anti-TIP Evaluation Initiative could be funded out of some percent of the monies allocated to projects for their separate evaluations, while leaving some monies to support project participation and cooperation with the Initiative, and to do their own monitoring. However there are probably a variety of ways to fund this, and program monies might be independently allocated to it.

Whatever the most viable and effective funding organization, the benefits yielded would be to improve the standards of evaluation while reducing the burden, duplication of effort, and costs. Specifically an Initiative would, or could:

- Remove from each program a majority of the need for undertaking individual impact evaluations, although they would be involved in helping set the standards, strategy for evaluation, building on what is a known, setting evaluation priority, and spreading the burden of evaluation across projects.
- Set and promote common definitions of VoTs, TIP, traffickers, and individuals and families at risk of being either or both.
- Develop an agreed upon evaluation agenda in order to systematically develop knowledge about progress and impact while cutting costs and burden of evaluation.

- Develop agreed upon and technically sound evaluation methodology, maintaining baseline data annually – or at some agreed period – regarding the incidence and patterns of TIP and the characteristics of its victims, while maintaining the integrity and needed confidentiality.
- Assist in training, implementing, and managing data collection, analysis, and reporting, through developing data collection plans with participating projects and agencies, training data collection agents, setting up protocols for entering and managing the data, setting up and managing quality control protocols, implementing analysis and some part of the reporting.
- Collect and disseminate lessons learned about effective practices and their relative and absolute and relative impacts for different groups and different ways of being vulnerable to or emerging from trafficking.

Orchestrating the evaluation of initiatives in a more comprehensive fashion would enable setting priorities for evaluation. Not everything has to be evaluated all the time; some things do not need to be evaluated right now, as they are pretty well understood or less central to the anti-TIP efforts and the directions it is moving; other issues require immediate attention. One of the real benefits (though somewhat painful for evaluators working in it) is the ability to come to agreement on a plan which not only reduces duplication of effort, but also enhances the evaluation tools available and the ability to make sense of evaluation findings. In order to make these kinds of decisions, there needs to be a central body that can foster agreement on evaluation priorities and develop a plan to stage attention on them. This is, in part, the function of Initiative management. There are a variety of initiatives currently in place that provide potential models for such cooperation, including the World Bank EFA/FTI and UIS for household surveys.[14]

## Conceptual Framework for a USAID Anti-TIP Evaluation Initiative

An Evaluation Initiative, as envisioned here (see Table below), would negotiate priorities among evaluation issues, and, if relevant, with funding groups about which projects, and which evaluation issues, require priority attention. Criteria for priority attention would be negotiated each year.

**Conceptual Framework for a USAID anti-TIP Evaluation Initiative**

1. Project monitoring carries the weight for management; evaluation is subsumed in the annual review for a subset of priority issues.
2. Annual content priorities are combined with attention to needs for immediate and continuing information
   a. To benchmark incidence, prevalence, and changes in core characteristics.
   b. To identify ways to enhance impact and sustainability of outcomes from project work to prevent trafficking and protect potential and actual victims from it.
4. Annual evaluation targets selected from a pool.
3. Shared data collection responsibilities, for 'experimental group,' within certain parameters; Initiative staff would collect, enter, and maintain comparison group data.
4. Shared data entry responsibilities, within certain parameters for 'experimental group,' within certain parameters.
5. Shared analysis, under certain conditions. Initiative staff would undertake basic analyses as agreed; Final data set would be shared, under certain conditions developed to protect

[14] There are also existing initiatives within other US Departments. Census' SIPP project (http://www.census.gov/sipp has a similar evolutionary motivation to that focusing this report – namely that the data were precious, there were concerns about having good and standardized data nationally, and having relevant data for project evaluation purposes.

**Conceptual Framework for a USAID anti-TIP Evaluation Initiative**

privacy and integrity.

6. Annual meetings feedback findings to stakeholders, distill implications for project work, and develop priorities for the next year's content.
7. Evaluation is funded by 8% of funded project value contributed from each funder (or project or program) directly to the Evaluation Management Consortium, with project participation support funded by 2% of project value.
8. Project monitoring funded through 5% of project value.

An example here of something that might benefit as a focus across projects is the design of a good household survey. One of the key problems with assessing anti-TIP interventions – and especially with changes in the incidence or prevalence of TIP– is the lack of standard methods for measuring these indicators. One approach which has shown some promise is the use of household surveys. For some issues, there are standard questions in place which have been tested over time and which make the results of the surveys more reliable and interpretable across projects. For other issues there are no standard questions; and even where there are they need to be further tested and the utility discussed.

There are other similar issues which could augment the example above. However, the overall objective should be clear: to cultivate effective evaluation strategies and techniques and to build a knowledge bank about what is known about effective interventions and how to enhance them.

The USAID Anti-TIP Evaluation Initiative is essentially operated by an Evaluation Consortium, composed of key funding, technical, and implementing stakeholders, as well as representation of community members, and managed by a USAID contractor. An evaluation strategy, crafted through an Initiative structure, would have four over-arching activity focus areas – namely to:

1. Represent concerns of decision-makers at all levels. Activities related to this would be to develop, negotiate, and develop buy-in for a long-term anti-TIP evaluation strategy, appropriate to all interventions and all funding participants, with annual (or some period) evaluation objectives,
2. Develop and implement sound evaluation procedures. Activities related to this would be the development of an operational plan requiring the negotiating the detailed logistics of its deployment across funding agencies participating in the Initiative. However, activities also would include establishing and maintaining evaluation design, implementation, data management and reporting protocols which maintain the rigor of the research and its appropriateness for interpretation; and carrying out training of stakeholder groups related to these standards.
3. Assess status of TIP and changes in its manifestations. Activities related to this would include designing, carrying out, and maintaining an annual data collection effort; appropriately publishing documentation and conducting analyses on the parameters of and changes in the nature of the TIP problem, in a way that represents the variety of involved groups (victim, potential, and trafficker, as well as protector), to allow these data to be used as baseline benchmarks for assessing program impact.
4. Conduct annual feedback presentations and training regarding findings and evaluation techniques. Activities related to this would be to track relative impact of different intervention elements for different beneficiary groups, and the documentation for how that impact was assessed; to create awareness of lessons learned related to these intervention elements, or specific evaluation techniques found relevant; and about changes overall in level of incidence etc., both related to and also independent of Interventions.

Through a strategy such as this, one group could be responsible for maintaining data to the required quality standards, and would get specific funding to support that. The function of the Anti-TIP Evaluation Consortium Initiative would be to help work through the strategic information needs and to reduce the proliferation of data sets that absorb valuable time and money and effort that could be better dedicated to either service or serviceable evaluation.

### Specific Benefits from an Anti-TIP Evaluation Consortium Initiative

A primary benefit of such an orchestrated evaluation strategy is the ability to balance the level of effort invested in a single project component with the burden of providing the information, and the ability to ensure that the results for the project are provided in a way that contributes to learning about how to impact trafficking as a whole. Not everything has to be evaluated all the time for each project. Where there are known effective practices and where the benchmarks for the practices are previously established, documentation of this performance may be used as proxies for effectiveness.

A second but equally important benefit is indeed the need for sophisticated technical evaluation and statistical expertise for measuring changes in dynamic contexts like trafficking in persons. Because of the interaction effects between attitudes, aspirations, behavior, economic need, social status, and potential social rigidity impeding ability to take advantage of economic opportunities, advanced methods and statistical techniques are required to understand unique contributions of specific interventions. A more centralized approach to evaluation design, where the same statistical and evaluation expertise oversees the work for a period of time, will assist interpretation and fine tuning the work to better understand intervention impact.

# Appendix C

## Index of Performance Indicators Referenced in Reviewed USAID Programs

| Index of Performance Indicators Used by Counter-Trafficking Programs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Prevention | | | Victim Protection | | | |
| *Indicators* | *Page in Doc* | *Income* | *Empower-ment* | *Violence prevention* | *Awareness Raising* | *Identification*[15] | *Direct assistance* | *Legal proceedings* |
| Source: *Best practices for programming to protect and assist victims of trafficking in Europe and Eurasia* (Rosenberg, 2008) | | | | | | | | |
| 1. Number of victims identified in a specific area or border | 8 | | | | | x | | |
| 2. Time lapse between identification and repatriation | 9 | | | | | | x | |
| 3. Type and number of key actors involved (e.g., law enforcement, consular officials, medical personnel). | 9 | | | | | x | | |
| 4. Number of victims identified by consular officials and others, such as social workers, doctors, family and friends, after training and outreach | 9 | | | | | x | | |
| 5. # Victims referred for assistance | 10 | | | x | | x | x | |
| 6. # of calls to hotlines | 10 | | | x | | x | | |
| 7. # of calls resulting in identification of victims | 10 | | | | | x | | |
| 8. # of calls resulting in assistance to victims | 10 | | | | | x | x | |
| 9. Incidence or threats of trauma, suicide or murder of victims, (or their families or assisters) inflicted as a result of assistance or attempt at assistance. | 13 | | | | | | x | x |
| 10. Program identifies and addresses the needs of victims of all ages, gender, and ethnicities. | 14 | | | | | x | x | x |
| 11. Geographic distribution of services | 15 | | | | | | x | |
| 12. Victims understand what services are being offered. | 17 | | | | | | x | |
| 13. # of calls by victims | 17 | | | x | | x | | |
| 14. Shelters are available and located where victims need them | 19 | | | | | | x | |
| 15. Victims describe shelter staff as empowering and caring, rather than insensitive or neglectful | 20 | | | | | | x | |
| 16. Shelters are staffed with full-time, properly qualified personnel | 21 | | | | | | x | |

[15] Awareness programs also fall under the "prevention" heading.

## Index of Performance Indicators Used by Counter-Trafficking Programs

| | | Prevention | | | Victim Protection | | | |
|---|---|---|---|---|---|---|---|---|
| ***Indicators*** | ***Page in Doc*** | ***Income*** | ***Empower-ment*** | ***Violence prevention*** | ***Awareness Raising*** | ***Identification***[15] | ***Direct assistance*** | ***Legal proceedings*** |
| 17. Services have sustainable funding sources | 21 | | | | | x | x | x |
| 18. Victims receive the assistance they need (e.g., are able to receive follow-up, social worker time, family counseling, etc. even if they're not in a shelter.) | 22 | | | | | | x | |
| 19. Reporting and Conviction rates | 29 | | | | | | | x |
| 20. Impact of testifying is minimized | 29 | | | | | | x | x |
| 21. Occurrence of threats of, and actual retribution against victims who testify | 31 | | | | | | x | x |
| 22. Criminal cases successfully prosecuted | 32 | | | | | | | x |
| 23. Civil cases successfully prosecuted | 33 | | | | | | | x |
| 24. Victim compensation for damages and lost wages[16] | 33 | | | | | | x | x |
| 25. Victim attempts to migrate again, OR victims who migrate and find a job that allows them to send wages home[17] | 36 | | | | | | x | |
| 26. Number of people trafficked[18] | 37 | | | | | | | |
| 27. Number of victims served[19] | 38 | | | | | | x | |
| 28. Number of victims re-trafficked | 41 | | | | x | x | x | |
| 29. Accuracy of public perceptions regarding the circumstances and victims of trafficking | 42 | | | | x | x | | |
| 30. Victim awareness of available services | 43 | | | | | x | x | |
| 31. # of victims identified belonging to neglected target groups | 56 | | | | | x | | |
| 32. # victim complaints about their treatment by law enforcement and border officials and reports of positive treatment by law enforcement border officials. | 56 | | | | | x | x | |
| 33. Official instructions for officials address all forms of trafficking and demographics | 58 | | | | | x | x | x |

[16] This is broken down on Page 71 in Rosenberg (2008).

[17] While programs may define the former as an indicator, some victims consider the latter a better indicator. Either indicator is best tracked over a long period of time, e.g., years rather than months (36). The GAO report (2007) also notes the importance of longitudinal studies to track impact—page 24 and elsewhere.

[18] This is extremely difficult to estimate, since a decrease in victims could just result in less effective identification. Also, data is often not consolidated across countries (38).

[19] Sometimes difficult to quantify accurately, since many countries have multiple service agencies and some victims receive services from more than one.

## Index of Performance Indicators Used by Counter-Trafficking Programs

| | | Prevention | | | Victim Protection | | | |
|---|---|---|---|---|---|---|---|---|
| ***Indicators*** | ***Page in Doc*** | ***Income*** | ***Empower-ment*** | ***Violence prevention*** | ***Awareness Raising*** | ***Identification[15]*** | ***Direct assistance*** | ***Legal proceedings*** |
| of victims | | | | | | | | |
| 34. Repatriated victims are able to earn income (short- and longer-term) | 23, 36[20] | | | | | | x | |
| 35. Number of victim complaints about their treatment by consular officials and reports of positive treatment by consular officials. | 59 | | | | | x | x | x |
| 36. Existence of a formalized and implemented and effective national referral mechanism. | 61 | | | | | x | x | x |
| 37. Victims use services in remote locations | 64 | | | | | | x | |
| 38. Victim eligibility for and utilization of state-funded social services | 64 | | | | | | x | |
| 39. Handbooks and reference guides for social workers are developed; social workers and other service professionals are aware of VoT-specific issues. | 64 | | | | x | x | x | |
| 40. Victim access to health care services | 66 | | | | x | | x | |
| 41. Amount of time victims have to spend in closed shelters awaiting trials | 68 | | | | | | x | x |
| 42. Existence and enforcement of laws that protect victim identities. | 69 | | | | | | x | x |
| 43. Number of cases in which a TIP victim's name, address, photo, or other identifying information is revealed publicly. | 69 | | | | | | x | x |
| 44. Victims receive amnesty for crimes committed while they were trafficked | 70 | | | | | | | x |
| 45. Victims are not arrested or threatened with arrest for refusing to testify against traffickers | 70 | | | | | | | x |
| 46. Assistance provided addresses the different needs of male, female and child-victims. | 14, 61 | | | | | | x | |
| Source: *Best practices in trafficking prevention in Europe and Eurasia* (Warnath, 2009) | | | | | | | | |
| 47. Existence of empowerment programs, conferences, dialogues, lectures, exchanges, town meetings, discussion groups, or | 5 | | | | x | | x | |

[20] It can be difficult to track wages over a long period of time following repatriation and retraining.

## Index of Performance Indicators Used by Counter-Trafficking Programs

| | | Prevention | | | Victim Protection | | | |
|---|---|---|---|---|---|---|---|---|
| ***Indicators*** | ***Page in Doc*** | ***Income*** | ***Empower-ment*** | ***Violence prevention*** | ***Awareness Raising*** | ***Identification[15]*** | ***Direct assistance*** | ***Legal proceedings*** |
| roundtables | | | | | | | | |
| 48. Attendance at empowerment programs, conferences, dialogues, lectures, exchanges, town meetings, discussion groups, or roundtables | 5 | | | | x | | x | |
| 49. Saturation of message to children through schools and other programs targeting children and families. | 6 | | | | x | | | |
| 50. Stickers, leaflets, brochures, and other materials distributed | 7 | | | | x | | | |
| 51. Posters and billboards erected | 7 | | | | x | | | |
| 52. Public awareness of anti-trafficking message as reported in surveys | 8 | | | | x | | | |
| 53. Increased actual wage-earning for participants[21] | 10 | x | x | | | | | |
| 54. Increased confidence and self-esteem of participants | 10 | | x | | | | | |
| 55. Number of participants trained in entrepreneurship who actually start (and then continue) businesses | 11 | x | x | | | | | |
| 56. Reduction of children and youth | 11 | | x | x | x | | | |
| 57. Leaving school, family problems, abuse and violence, stigma, and exclusion | 11 | | x | x | x | | | |
| 58. Presence of and attendance in after-school programs and summer camps[22] | 12 | | x | x | x | | | |
| 59. Number of people using "safe migration" services and following their guidelines.[23] | 14 | | | | x | | | |
| 60. Utilization of job-vetting by trustworthy NGO | 15 | | | | x | | | |
| *Source: "Sex Trafficking in Nepal"* (Crawford et al., 2008) | | | | | | | | |
| 61. Proficiency in a marketable skill following rehabilitation | 911 | x | x | | x | | x | |
| 62. Visits from family members in transitional residence or shelter | 911 | | x | | | | x | |
| 63. Reintegration into home community following repatriation | 912 | x | x | | | | x | |

[21] Difficult to measure, especially as needed over long-term; most reported results thus far have been anecdotal.
[22] Rosenberg noted in 2004 that nobody had documented the prevention impact of such programs; follow-up studies with participants could be conducted (12). The author notes that data revealing long-term outcomes among participants will not likely be available for analysis (12).
[23] Author notes that data on these programs hasn't been gathered.

| Index of Performance Indicators Used by Counter-Trafficking Programs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | **Prevention** | | | **Victim Protection** | | | |
| ***Indicators*** | ***Page in Doc*** | ***Income*** | ***Empower-ment*** | ***Violence prevention*** | ***Awareness Raising*** | ***Identification***[15] | ***Direct assistance*** | ***Legal proceedings*** |
| 64. Reintegration into family of origin following repatriation | 913 | | x | x | | | x | |
| 65. Victim reports satisfactory readjustment | 912 | x | x | | | | x | |
| 66. Victim trafficked for prostitution gets married following repatriation | 912 | | | | | | x | |
| 67. Incidence of STDs among victims trafficked for prostitution | 912 | | | | | | x | |
| 68. Prevalence of mental illness among repatriated victims | 913 | | x | x | | | x | |
| Source: *Human Trafficking*, (GAO, 2007) | | | | | | | | |
| 69. Number of victims assisted | 19 | x | x | | | | x | x |