



USAID
FROM THE AMERICAN PEOPLE

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

Executive Summary

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

Executive Summary

December 2009

The views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

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	<i>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.</i>
Activities	<i>Actions taken or work performed through which inputs are mobilized to produce outputs.</i>
Performance Indicators	<i>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.</i>
Pathways	<i>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.</i>
Expected Outputs	<i>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.</i>
Expected Outcomes	<i>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.</i>
Expected Impact	<i>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.</i>
Assumptions	<i>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.</i>
External Factors	<i>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.</i>

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 **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. 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